IA Mathematics Revision

Abigail Tan
25th May 2021

We will give brief overviews of some important things in the courses, and outline how the proofs work (but not actually include the proofs here - refer to lecture notes for them).
These notes contain a mixture of content/examples and cover different courses in different levels of depth.
I realise it is frustrating when you go somewhere to look at a proof and it's omitted or "to follow". Sorry about this.

Contents

1 Numbers and Sets

1.1 Number Theory

Fermat's Little Theorem. Let p be a prime. Then in $\mathbb{Z}_p$,

$$a^{p-1} \equiv 1 \pmod{p}$$

for all $a \neq 0$.
To prove, we consider $a \times 1, a \times 2, \ldots, a \times (p-1)$ in $\mathbb{Z}_p$, which are all distinct and nonzero since a is invertible, so they must be $1, 2, \ldots, p-1$ in some order. Multiplying, we obtain

$$a^{p-1}(p-1)! \equiv (p-1)! \pmod{p} \implies a^{p-1} \equiv 1 \pmod{p}.$$

For general n, we have the Fermat-Euler theorem

$$a^{\phi(n)} \equiv 1 \pmod{n}$$

for every invertible a. We prove this exactly the same way, but only work with the invertibles.

Wilson's Theorem. Let p be a prime. Then

$$(p-1)! \equiv -1 \pmod{p}.$$

To prove, suppose $p > 2$ and consider $1, 2, 3, \ldots, p-1$. We can pair up each a with its inverse a^{-1} for $a \neq a^{-1}$. But $a = a^{-1} \implies a^2 = 1 \iff a = 1$ or $a = -1$. Multiplying, we obtain in $\mathbb{Z}_p$ that

$$(p-1)! = 1^{p-3/2} \times 1 \times -1 = -1.$$

Chinese Remainder Theorem. Let $u_1, u_2, \ldots, u_k$ be pairwise coprime. Then for all $a_1, \ldots, a_k$, there exists x such that $x \equiv a_i \pmod{u_i}$ for all i. The solution is unique mod $u_1 u_2 \ldots u_k$.
This can be proved by induction, and we will start with the case where $k = 2$. Let $u_1 = u$ and $u_2 = v$, and $a_1 = a$ and $a_2 = b$.
To show existence, we notice that since u, v are coprime, we have

$$su + tv = 1$$

for some $s, t \in \mathbb{Z}$. Then observe that

$$su \equiv 0 \pmod{u}, \quad su \equiv 1 \pmod{v}, \quad tv \equiv 1 \pmod{u}, \quad tv \equiv 0 \pmod{v}.$$

Then $x = atv + bsu$ has $x \equiv a \pmod{u}$ and $x \equiv b \pmod{v}$ as required.
Now we show uniqueness. Suppose x' is also a mod u and b mod v. Then both u and v divide $x' - x$ but are coprime, so uv divides $x' - x$ so $x' \equiv x \pmod{uv}$.

1.2 The Reals

Proof that e is irrational. Suppose that e is rational, so $e = p/q$ for $p, q \in \mathbb{Z}$, $q > 1$, which implies that

$$q!e = \sum_{n=0}^{\infty} \frac{q!}{n!} \in \mathbb{Z}.$$

Then we have

$$\underbrace{q! + \frac{q!}{1!} + \frac{q!}{2!} + \cdots + \frac{q!}{q!}}_{\text{an integer}} + \frac{q!}{(q+1)!} + \frac{q!}{(q+2)!} + \ldots$$

but by considering the terms after this, we observe that in general

$$\frac{q!}{(q+n)!} \leq \frac{1}{(q+1)^n}$$

so

$$\sum_{n=q+1}^{\infty} \frac{q!}{n!} \leq \frac{1}{q+1} + \frac{1}{(q+1)^2} + \cdots = \frac{1}{q} < 1$$

so this is not an integer, and we have a contradiction.

2 Groups

This revision section is not exhaustive, and will mainly consist of stating important theorems and outlining the ideas behind the proofs.

2.1 Important/General Properties

Direct product theorem. Let H and K be subgroups of G where

- $H \cap K = \{e\}$
- for all $h \in H, k \in K$, we have $hk = kh$
- for all $g \in G$, there exist $h \in H, k \in K$ such that $g = hk$.

Then $G \cong H \times K$.
We prove this by considering the homomorphism $f : H \times K \to G$ where $(h, k) \mapsto hk$. Just check that it's a homomorphism, and show that it is injective (by checking trivial kernel) and surjective.

Lagrange's Theorem. Let $H \leq G$ for a finite group G. Then

$$|G| = |G : H||H|.$$

We prove this by showing that

- $|H| = |gH|$ for all $g \in G$ (prove: bijection)
- for $g_1, g_2 \in G$, either $g_1H = g_2H$ or g_1H and g_2H are disjoint (g in both implies $g = g_1h_1 = g_2h_2 \implies g_1h = g_2h_2h_1^{-1}h \in g_2H$ and vice versa)
- $G = \bigcup_{g \in G} gH$ (show each is a subset of the other).

This way, we show that the cosets partition the group into equally sized subsets, and the result follows immediately.

2.2 Symmetric Groups

Disjoint cycles theorem. Any permutation σ in S_n can be written as a composition of disjoint cycles, which is unique up to reordering.
To prove this, we consider the elements

$$1, \sigma(1), \sigma^2(1), \sigma^3(1), \dots$$

and since S_n is finite, we eventually get a repeat, so there is some smallest k with $\sigma^k(1) = 1$. This gives us a cycle of distinct elements. We can repeat this starting with the next number that hasn't been used yet, and since σ is bijective, there can be no repeats. Eventually we will get the entire permutation in disjoint cycles. It is fairly straightforward to check uniqueness by considering two disjoint representations and looking at the unique determination of elements in the cycles.

Product of transpositions. Let $\sigma \in S_n$. Then σ is a product of transpositions.
To prove this, doing it for a cycle is sufficient (because of DCD). We simply observe that

$$(a_1\, a_2 \dots a_k) = (a_1\, a_2)(a_2\, a_3) \dots (a_{k-1}\, a_k).$$

Sign of permutation is well-defined. A permutation σ is always either a product of an even number of transpositions, or a product of an odd number of transpositions.

To prove, we write $\#(\sigma)$ for the number of cycles in the DCD of σ, including 1-cycles. We then show that multiplying σ by a transposition $\tau = (c\,d)$ corresponds to

$$\#(\sigma) = \#(\sigma\tau) + 1 \pmod 2.$$

If c and d are in the same cycle then multiplying by $(c\,d)$ gives

$$(c\,a_2\,a_3 \ldots a_{k-1}\,d\,a_{k+1} \ldots a_l)(c\,d) = (c\,a_{k+1} \ldots a_l)(d\,a_2\,a_3 \ldots a_{k-1}) \implies \#(\sigma\tau) = \#(\sigma) + 1.$$

If c and d are in different cycles, then we have

$$(c\,a_2 \ldots a_k)(d\,b_2 \ldots b_l)(c\,d) = (c\,b_2\,b_3 \ldots b_l\,d\,a_2 \ldots a_k) \implies \#(\sigma\tau) = \#(\sigma) - 1.$$

In either case we have

$$\#(\sigma) = \#(\sigma\tau) + 1 \pmod 2.$$

But $\#(\sigma)$ is uniquely determined by σ, so if we have two ways of writing σ in m transpositions or in n transpositions, then constructing σ one transposition at a time, we must have $m \equiv n \pmod 2$ as required.

2.3 Möbius Groups

Definition. A Möbius map is a function $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ of the form

$$f(z) = \frac{az + b}{cz + d}$$

with $a, b, c, d \in \mathbb{C}$, $ad - bc \neq 0$, $f(-d/c) = \infty$, $f(\infty) = a/c$ if $c \neq 0$ and $f(\infty) = \infty$ if $c = 0$.
The set of Möbius maps forms a group.

Composition of maps. Every Möbius map can be written as a composition of maps of the form

- $f(z) = az, a \neq 0$ (dilation/rotation)
- $f(z) = z + b$ (translation)
- $(f(z) = \frac{1}{z}$ (inversion).

We can check this:

$$z \mapsto z + \frac{d}{c} \mapsto \frac{1}{z + \frac{d}{c}} \mapsto \frac{(-ad + bc)c^{-2}}{z + \frac{d}{c}} \mapsto \frac{a}{c} + \frac{(-ad + bc)c^{-2}}{z + \frac{d}{c}} = \frac{az + b}{cz + d}.$$

In fact, the set of all such maps generates the Möbius group M.

Fixed points. A fixed point of a Möbius map $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ is a point $z \in \hat{\mathbb{C}}$ with $f(z) = z$.

Theorem. A Möbius map with three or more fixed points is the identity.
To prove this, we consider the case where ∞ is not a fixed point; then the quadratic obtained by solving

$$\frac{az + b}{cz + d} = z$$

has three complex roots, which is impossible. Then considering the case where ∞ is fixed, we have

$$\frac{a\infty + b}{c\infty + d} \coloneqq \frac{a}{c} = \infty$$

so $c = 0$ and then we have two solutions to a linear equation, which is also impossible.

Corollary. If two Möbius maps coincide on three distinct points in $\hat{\mathbb{C}}$, then they are equal.
This follows immediately if we consider two maps f and g that coincide on three distinct points - then $g^{-1}f$ fixes all three points so is the identity.

Theorem. There is a unique Möbius map sending any 3 distinct points z_1, z_2, z_3 of $\hat{\mathbb{C}}$ to any 3 distinct points w_1, w_2, w_3 of $\hat{\mathbb{C}}$.
To prove this, we can check that

$$f(z) = \frac{(z_2 - z_3)(z - z_1)}{(z_2 - z_1)(z - z_3)}$$

satisfies $(z_1, z_2, z_3) \mapsto (0, 1, \infty)$.
Then we can find f_1 that sends (z_1, z_2, z_3) to $(0, 1, \infty)$ and f_2 that sends (w_1, w_2, w_3) to $(0, 1, \infty)$. Then $f := f_2^{-1} f_1$ works as required. Uniqueness follows immediately from the above corollary.

Theorem. Every non-identity element $f \in M$ has either one or two fixed points. If f has one fixed point, then it is conjugate to $z \mapsto z + 1$. If f has two fixed points, then it is conjugate to a map of the form $z \mapsto az, a \in \mathbb{C} \setminus \{0\}$.
To prove this, we observe that a non-identity element has less than 3 fixed points. Then the quadratic equation obtained from $f(z) = z$ has at least one distinct complex root.
If f has exactly one fixed point z_0, then we choose $z_1 \in \mathbb{C}$ not fixed by f. Then $(z_1, f(z_1), z_0)$ are all distinct. So there is some $g \in M$ sending this triple to $(0, 1, \infty)$ and we can check that gfg^{-1} must take the form $z \mapsto az + 1$ ($a \in \mathbb{C}$). If $a \neq 1$ then this fixes $1/(1 - a) \neq \infty$ so f would have more than one fixed point, a contradiction. Hence $z \mapsto z + 1$.
If f has exactly two fixed points z_0 and z_1, then let g be any Möbius map sending $(z_0, z_1) \mapsto (0, \infty)$ then we can check that gfg^{-1} fixes 0 and ∞, so gfg^{-1} has the form $z \mapsto (gfg^{-1}(1))z$.

Cross-ratios. Let $z_1, z_2, z_3, z_4 \in \hat{\mathbb{C}}$ be distinct. Then their cross-ratio $[z_1, z_2, z_3, z_4]$ is defined to be $f(z_4)$ where $f \in M$ is the unique map with $f(z_1) = 0, f(z_2) = 1, f(z_3) = \infty$. We have

$$[z_1, z_2, z_3, z_4] = \frac{(z_4 - z_1)(z_2 - z_3)}{(z_2 - z_1)(z_4 - z_3)}.$$

These four points lie on a circle in $\hat{\mathbb{C}}$ if and only if their cross ratio is real.
We can check also that Möbius maps preserve the cross-ratio.

2.4 Isomorphism Theorems

First isomorphism theorem. Let $\phi : G \to H$ be a homomorphism. Then

$$G/\ker\phi \cong \operatorname{im}\phi.$$

We prove this by considering the (fairly natural) map $\psi : G/\ker\phi \to \operatorname{im}\phi$ given by

$$g\ker\phi \mapsto \phi(g).$$

Then we check that this is well-defined, that it is a homomorphism, and that it is bijective.

Second isomorphism theorem. Let $H \leq G$ and $N \trianglelefteq G$. Then $H \cap N \trianglelefteq H$ and

$$H/H \cap N \cong HN/N.$$

To prove this we consider the well-defined surjective homomorphism

$$\phi : H \to HN/N, \quad \phi(h) = hN$$

which has kernel $N \cap H$. Then apply the first isomorphism theorem.

Third isomorphism theorem. To be done, this never comes up.

2.5 Group Actions

Definition. Let G be a group and X be a set. An action of G on X is a function $\alpha : G \times X \to X$ with

- $\alpha_g(x) = g(x) \in X$ for all $g \in G, x \in X$
- $\alpha_e(x) = e(x) = x$ for all $x \in X$
- $\alpha_g \alpha_h(x) = g(x)h(x) = \alpha_{gh}(x) = gh(x)$ for all $g, h \in G, x \in X$.

A group action can also be thought of as a homomorphism ρ from the group G to the symmetry group of the set X.
[Add proof for this]

A group action is faithful if $\ker \rho = \{e\}$, that is "only e fixes everything".

Orbit-stabiliser theorem.

$$|G| = |\operatorname{Orb} x||\operatorname{Stab} x|$$

Proof follows from Lagrange's theorem: $g(x) = h(x) \iff h^{-1}g(x) = x \iff h^{-1}g \in \operatorname{Stab} x \iff g \operatorname{Stab} x = h \operatorname{Stab} x$ so distinct points in $\operatorname{Orb} x$ are in bijection with distinct cosets of $\operatorname{Stab} x$. So $|\operatorname{Orb} x| = |G : \operatorname{Stab} x|$ and the result follows from Lagrange.

3 Differential Equations

3.1 Perturbation Analysis

A fixed point or equilibrium point is where

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f(y, x) = 0$$

for all x. The fixed point is stable if the solution curves in a small neighbourhood of the fixed point converge towards the fixed point. Otherwise, it is unstable.

Let $y = a$ be a fixed point, so $f(a, x) = 0$ for all x. Then let $y = a + \varepsilon(x)$. Then using the original differential equation, we get

$$\frac{\mathrm{d}\varepsilon}{\mathrm{d}x} = f(a + \varepsilon(x), x) = f(a, x) + \varepsilon \frac{\partial f}{\partial y}(a, x) + O(\varepsilon^2).$$

Since a is a fixed point, $f(a, x) = 0$, and $O(\varepsilon^2)$ terms can be ignored. We then get the approximation

$$\frac{\mathrm{d}\varepsilon}{\mathrm{d}x} = \varepsilon \frac{\partial f}{\partial y}(a, x).$$

Then if ε tends to 0 as x tends to infinity, then we have a stable fixed point, and otherwise it is an unstable fixed point.

Note in the autonomous case where $f = f(y)$ independent of x, the differential equation can be solved to give

$$\varepsilon = \varepsilon_0 e^{kx}$$

with

$$k = \frac{\mathrm{d}f}{\mathrm{d}y}(a)$$

and then we can straight away see that $f'(a) < 0$ gives a stable fixed point and $f'(a) > 0$ gives an unstable fixed point.

4 Analysis I

Like Groups, this section is not at all exhaustive.

4.1 Sequences and Series

Cauchy sequences. A sequence is Cauchy if for all $\varepsilon > 0$, there exists $N > 0$ such that $|a_n - a_m| < \varepsilon$ for all $n, m \geq N$. That is, a sequence whose terms eventually become arbitrarily close to each other.

Claim. A sequence is convergent if and only if it is a Cauchy sequence.
To prove that a convergent sequence is Cauchy, we just apply the triangle inequality to a_n, a_m and the limit of the sequence. To prove that every Cauchy sequence is convergent, we show that every Cauchy sequence is bounded by writing

$$\begin{aligned} &|a_n - a_m| < 1 \ \forall n, m \geq N(1) \\ \implies &|a_m| \leq |a_m - a_N| + |a_N| < 1 + |a_N| \end{aligned}$$

by the triangle inequality, then we have

$$|a_n| \leq \max\{1 + |a_N|, |a_n| \ (n = 1, 2, \ldots, N-1)\}$$

as required. By Bolzano-Weierstrass, the sequence has a convergent subsequence $(a_{n_j}) \to a$. We then show that the whole sequence tends to the same limit. To show this, we notice that given $\varepsilon > 0$, there exists j_0 such that for all $j \geq j_0$, $|a_{n_j} - a| < \varepsilon$. Also by the Cauchy property there exists $N(\varepsilon)$ such that $|a_m - a_n| < \varepsilon$ for all $m, n \geq N(\varepsilon)$. If we choose j such that

$$n_j \geq \max\{N(\varepsilon), n_{j_0}\}$$

then if $n \geq N(\varepsilon)$ then

$$|a_n - a| \leq |a_n - a_{n_j}| + |a_{n_j} - a| < 2\varepsilon$$

as required. □

Geometric series. We claim that the series

$$\sum_{j=0}^{n} x^j$$

converges if and only if $|x| < 1$. This follows from writing

$$\sum_{j=1}^{n} x^{j-1} = \frac{1 - x^n}{1 - x}$$

and observing that $x^n \to 0$ as $n \to \infty$.

Harmonic series. The harmonic series is divergent, and we prove this by writing S_n for the nth partial sum, and then showing that $S_{2n} > S_n + \frac{1}{2}$ by using that

$$\frac{1}{n+k} \geq \frac{1}{2^n}$$

for $k = 1, 2, \ldots, n$.

4.2 Convergence Tests

Comparison test. Suppose $0 \leq b_n \leq a_n$ for all n. Then

$$\sum_{n=1}^{\infty} a_n \text{ converges } \implies \sum_{n=1}^{\infty} b_n \text{ converges.}$$

We prove this simply by considering the Nth partial sum for each series and using monotone convergence.

Root test. Suppose $a_n \geq 0$ and $a_n^{1/n} \to a$ as $n \to \infty$. Then if $a > 1$, then $\sum a_n$ diverges, and if $a < 1$ then $\sum a_n$ converges. (Nothing can be said if $a = 1$).
To prove, we first consider $a < 1$. We choose $a < r < 1$, and then by the definition of a limit, there exists N such that for all $n \geq N$, $a_n^{1/n} < r \implies a_n < r^n$. Then by comparison with the geometric series in r, the required series converges. It diverges for $a > 1$ since we can quite easily show that $a_n \not\to 0$.

Ratio test. Suppose $a_n > 0$ and $a_{n+1}/a_n \to L$. Then if $L > 1$, then $\sum a_n$ diverges, and if $L < 1$ then $\sum a_n$ converges. (Nothing can be said if $L = 1$).
Similarly to the root test, we first consider $L < 1$. We choose $L < r < 1$, and then by limit definition, there exists N such that for all $n \geq N$, $a_{n+1}/a_n < r \implies a_n < r^n$. Then

$$a_n = \frac{a_n}{a_{n-1}} \frac{a_{n-1}}{a_{n-2}} \dots \frac{a_{N+1}}{a_N} a_N < a_N r^{n-N}, \ n > N$$

(obtained by applying $a_{n+1}/a_n < r$ several times). Then $a_n < Kr^n$ for some constant K and then by comparison with the geometric series, $\sum a_n$ converges.
We can prove that it diverges for $L > 1$ in a similar way, reversing the directions of inequalities and showing that the series is bounded below by a divergent geometric series.

Cauchy's condensation test. Let a_n be a decreasing sequence of positive terms. Then

$$\sum_{n=1}^{\infty} a_n \text{ converges } \iff \sum_{n=1}^{\infty} 2^n a_{2^n} \text{ converges.}$$

[Proof to follow.]
We can use this to show that the Riemann zeta function $\zeta(s)$ converges if and only if $s > 1$.

Intermediate value theorem. Let $f : [a, b] \to \mathbb{R}$ be a continuous function.
Then for all $d \in [\min\{f(a), f(b)\}, \max\{f(a), f(b)\}]$, there exists $c \in [a, b]$ such that $f(c) = d$.

To prove this, we suppose WLOG that $f(a) \leq f(b)$. We consider the supremum of the set

$$S = \{x \in [a, b] : f(x) < d\}$$

and then show that $f(\sup S) \leq d$ and that $f(\sup S) \geq d$, so $f(\sup S) = d$.

Boundedness of a continuous function. A continuous function $f : [a, b] \to \mathbb{R}$ is bounded and attains its bounds.
To show that it is bounded, suppose for contradiction that it is not. Then there exists a sequence $(x_n) \in [a, b]$ with $f(x_n) > n$ for all n. The sequence has a convergent subsequence $(x_{n_j}) \to x$ by Bolzano-Weierstrass, but $f(x_{n_j}) \to \infty$ whilst $x_{n_j} \to x$, contradicting continuity.
To show that it attains its bounds, we consider the supremum and infimum of the set of values attained by the function on the interval

$$S = \{f(x) : x \in [a, b]\}.$$

We can choose a sequence (x_n) where $f(x_n) \to \sup S$ and then apply Bolzano-Weierstrass to show that the limit x of a convergent subsequence of (x_n) satisfies $f(x) = \sup S$.

5 Vector Calculus

5.1 Line, Surface and Volume Integrals

Length of curve. We can parametrise a curve $\mathbf{x}(t)$ using the parameter t. To find the length of the part of the curve parametrised by $[a,b] \ni t \mapsto \mathbf{x}(t)$, we use

$$l(C) = \int_a^b |\mathbf{x}'(t)|\,\mathrm{d}t = \int_C \mathrm{d}s$$

where $\mathrm{d}s$ is the arc-length element $|\mathbf{x}'(t)|\,\mathrm{d}t$.

We can also integrate functions over a region of a curve with

$$\int_C f(\mathbf{x})\,\mathrm{d}s = \int_a^b f(\mathbf{x}(t))|\mathbf{x}'(t)|\,\mathrm{d}t.$$

Line integrals. For a vector field $\mathbf{F}(x)$ and a piecewise smooth parametrised curve $[a,b] \ni t \mapsto \mathbf{x}(t)$, we define the line integral of the field over the curve by

$$\int_C \mathbf{F}\cdot\mathrm{d}\mathbf{x} = \int_a^b \mathbf{F}(\mathbf{x}(t))\cdot\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}\,\mathrm{d}t.$$

If $\mathbf{x}(a) = \mathbf{x}(b)$ then we have a closed curve and write

$$\oint_C \mathbf{F}\cdot\mathrm{d}\mathbf{x}$$

which is sometimes called the circulation of $\mathbf{F}$ about C.

Area integrals. To integrate over a region of $\mathbb{R}^2$, we use

$$\int_D f(x,y)\,\mathrm{d}A = \iint_D f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \iint_{D'} f(x(u,v),y(u,v))|J|\,\mathrm{d}u\,\mathrm{d}v$$

where J is the Jacobian given by

$$J = \det\begin{pmatrix}\partial x/\partial u & \partial x/\partial v\\ \partial y/\partial u & \partial y/\partial v\end{pmatrix}.$$

Volume integrals. We use almost exactly the same ideas as we use for area integrals, where $\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = |J|\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w$ and

$$J = \det\begin{pmatrix}\partial x/\partial u & \partial x/\partial v & \partial x/\partial w\\ \partial y/\partial u & \partial y/\partial v & \partial y/\partial w\\ \partial z/\partial u & \partial z/\partial v & \partial z/\partial w\end{pmatrix}.$$

Surface integrals. We define the scalar area element $\mathrm{d}S$ and the vector area element $\mathrm{d}\mathbf{S}$ by

$$\mathrm{d}S = \left|\frac{\partial\mathbf{x}}{\partial u}\times\frac{\partial\mathbf{x}}{\partial v}\right|\mathrm{d}u\,\mathrm{d}v$$

$$\mathrm{d}\mathbf{S} = \left(\frac{\partial\mathbf{x}}{\partial u}\times\frac{\partial\mathbf{x}}{\partial v}\right)\mathrm{d}u\,\mathrm{d}v = \mathbf{n}\,\mathrm{d}S$$

where $\mathbf{n}$ is a unit normal vector which points in an outwards direction relative to the surface. This way, we can compute integrals of the form

$$\int_S f\,\mathrm{d}S \text{ and } \int_{\partial S}\mathbf{F}\cdot\mathrm{d}\mathbf{S}.$$

5.2 Tangent, Normal and Binormal Vectors

We can parametrise curves by arc length:

$$s(t) = \int_a^t |\mathbf{x}'(\tau)| \, \mathrm{d}\tau.$$

We can write $\mathbf{r}(s) = \mathbf{x}(t(s))$, then by the chain rule we obtain the tangent vector

$$\mathbf{t}(s) \coloneqq \mathbf{r}'(s) = \frac{\mathbf{x}'(t(s))}{|\mathbf{x}'(t(s))|}$$

and since this is a unit vector, its derivative measures only the change in direction. We define the curvature $\kappa(s)$ by

$$\kappa(s) = |\mathbf{r}''(s)| = |\mathbf{t}'(s)|.$$

The greater the curvature, the greater the extent to which the curve changes direction at that point.

The principal normal vector $\mathbf{n}$ is defined by

$$\mathbf{t}' = \kappa \mathbf{n}.$$

The binormal vector $\mathbf{b}$ is defined by

$$\mathbf{b} = \mathbf{t} \times \mathbf{n}.$$

The torsion τ of a curve is defined by

$$\mathbf{b}' = -\tau \mathbf{n}.$$

In fact, the curvature and torsion define a curve up to translation or orientation.

Radius of curvature. The radius of curvature

$$R(s) = \frac{1}{\kappa(s)}$$

is the radius of the circle that best fits the curve when the circle goes through the curve tangentially at the given point.

5.3 Gradient

Definition. For a scalar field $f : \mathbb{R}^3 \to \mathbb{R}$, we define the gradient of f, ∇f, by

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot \mathbf{h} + o(\mathbf{h}).$$

Example. For $f(x) = \frac{1}{2}|\mathbf{x}|^2$, we have

$$f(\mathbf{x} + \mathbf{h}) = \frac{1}{2}(\mathbf{x} + \mathbf{h}) \cdot (\mathbf{x} + \mathbf{h}) = \frac{1}{2}|\mathbf{x}|^2 + \frac{1}{2}(2\mathbf{x} \cdot \mathbf{h}) + \frac{1}{2}|\mathbf{h}|^2 = f(\mathbf{x}) + \mathbf{x} \cdot \mathbf{h} + o(\mathbf{h})$$

and hence $\nabla f(\mathbf{x}) = \mathbf{x}$.
The gradient vector points in the direction of greatest increase of f.

Calculating the gradient. See the Vector Calculus formula sheet for the appropriate scale factors in each coordinate system. In a system (u, v, w) of orthogonal curvilinear coordinates, the gradient can be computed using

$$\nabla f = \frac{1}{h_u}\frac{\partial f}{\partial u}\mathbf{e}_u + \frac{1}{h_v}\frac{\partial f}{\partial v}\mathbf{e}_v + \frac{1}{h_w}\frac{\partial f}{\partial w}\mathbf{e}_w.$$

Conservative fields. We say that a vector field $\mathbf{F}$ is conservative if $\mathbf{F} = \nabla f$ for some scalar field f. We write $\mathrm{d}f = \nabla f \cdot \mathrm{d}\mathbf{x}$, so $\mathbf{F} \cdot \mathrm{d}\mathbf{x}$ is said to be exact if and only if $\mathbf{F}$ is conservative.

If $\mathbf{F}$ is conservative, then

$$\oint_C \mathbf{F} \cdot \mathrm{d}\mathbf{x} = 0$$

for any closed curve C.

5.4 Integral Theorems

Green's theorem. Let $P = P(x, y)$ and $Q = Q(x, y)$ be continuously differentiable functions on $A \cup \partial A \in \mathbb{R}^2$ where ∂A is the piecewise smooth boundary of A. Then

$$\oint_{\partial A} P \,\mathrm{d}x + Q \,\mathrm{d}y = \iint_A \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \mathrm{d}x \,\mathrm{d}y.$$

Stokes' theorem. Let $\mathbf{F} = \mathbf{F}(\mathbf{x})$ be a continuously differentiable vector field and let S be an orientable piecewise regular surface with piecewise smooth boundary ∂S. Then

$$\int_S (\nabla \times \mathbf{F}) \cdot \mathrm{d}\mathbf{S} = \oint_{\partial S} \mathbf{F} \cdot \mathrm{d}\mathbf{x}.$$

Divergence theorem. Let $\mathbf{F} = \mathbf{F}(\mathbf{x})$ be a continuously differentiable vector field and let V be a volume with piecewise regular boundary ∂V. Then

$$\int_V \nabla \cdot \mathbf{F} \,\mathrm{d}V = \int_{\partial V} \mathbf{F} \cdot \mathrm{d}\mathbf{S}.$$

5.5 Maxwell's Equations

Here, $\mathbf{B} = \mathbf{B}(\mathbf{x}, t)$ is magnetic field and $\mathbf{E} = \mathbf{E}(\mathbf{x}, t)$ is electric field. These fields depend on charge density $\rho = \rho(\mathbf{x}, t)$ and current density $\mathbf{J} = \mathbf{J}(\mathbf{x}, t)$. The four equations are

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0}$$

$$\nabla \cdot \mathbf{B} = 0$$

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0$$

$$\nabla \times \mathbf{B} - \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} = \mu_0 \mathbf{J}.$$

Note that $1/\mu_0\varepsilon_0 = c^2$.

5.6 Poisson's Equation

Many problems can be reduced to Poisson's equation

$$\nabla^2 \phi = F$$

or Laplace's equation

$$\nabla^2 \phi = 0.$$

The problem $\nabla^2\phi = F$ in Ω with $\phi = f$ on $\partial\Omega$ is called the Dirichlet problem.
The problem $\nabla^2\phi = F$ in Ω with $\frac{\partial\phi}{\partial\mathbf{n}} = \mathbf{n}\cdot\nabla\phi = g$ on $\partial\Omega$ is called the Neumann problem.

Note: when we obtain a solution ϕ, it must be well-defined everywhere in Ω. If we obtain a term of the solution which is not defined everywhere, then the multiplicative constant must be 0.

Spherical symmetry. If the problem takes the form

$$\nabla^2\phi = f(r)$$

then it is reasonable to assume that $\phi = \phi(r)$. We then have

$$\nabla^2\phi = \frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}\phi}{\mathrm{d}r}\right)$$

and then we can substitute this into the original equation and solve an ODE in r.

Uniqueness. The solution to the Dirichlet problem is unique, and the solution to the Neumann problem is unique up to a a constant. The way this is proved is to consider the difference of two solutions, $\psi = \phi_1 - \phi_2$, and then show that ψ must be zero or a constant (respectively). Using the parameters of the problem, it follows that $\nabla^2\psi = 0$ in Ω and either $\psi = 0$ (Dirichlet) or $\frac{\partial\psi}{\partial\mathbf{n}} = 0$ (Neumann) on the boundary.
We consider the non-negative functional

$$I(\psi) = \int_\Omega |\nabla\psi|^2\,\mathrm{d}V = \int_\Omega \nabla\psi\cdot\nabla\psi\,\mathrm{d}V \geq 0.$$

Using a standard identity and the divergence theorem, we expand out

$$I(\psi) = \int_\Omega (\nabla\cdot(\psi\nabla\psi) - \psi\nabla^2\psi)\,\mathrm{d}V = \int_{\partial\Omega}(\psi\nabla\psi)\cdot\mathrm{d}\mathbf{S} = \int_{\partial\Omega}\psi\frac{\partial\psi}{\partial\mathbf{n}}\,\mathrm{d}S = 0.$$

It follows that $\nabla\psi = 0$ throughout Ω so ψ is constant. In the Dirichlet problem, ψ is 0 on the boundary, so by continuity it is zero everywhere. For the Neumann problem, we just have uniqueness up to a constant.

Note. Many questions that ask you to show uniqueness involve considering $I(\psi)$ or some variant of it.

Newton's shell theorem. Suppose for $r < a$ we have charge density $\rho(\mathbf{x}) = 0$ and for $r \geq a$ we have $\rho(\mathbf{x}) = F(r)$. We know that the electric potential ϕ satisfies

$$\nabla^2\phi = -\frac{\rho(\mathbf{x})}{\varepsilon_0} = 0$$

for $r < a$ (this follows from Maxwell's equations and $\mathbf{E} = -\nabla\phi$).
By spherical symmetry $\phi = \phi(r)$ so $\phi = \phi(a)$ which is constant on $r = a$. The unique solution to $\nabla^2\phi = 0$ on $r < a$ and ϕ constant on $r = a$ is ϕ constant throughout $r \leq a$.
It follows that $\mathbf{E} = -\nabla\phi = \mathbf{0}$ throughout $r < a$ so there is no electric field inside a ball of radius a with no charge density.

The superposition principle. We are working with linear problems here, so any linear combination of solutions is also a solution.

Example sheet question. Vector Calculus Example Sheet 3 Q4(b)(iii) is an example where this principle is applied. Since we are working in the region $a < r < b$, and are given $\frac{\partial\phi}{\partial\mathbf{n}}(a,\theta) = 0$, we have

$$\frac{\partial\phi}{\partial\mathbf{n}}(a,\theta) = \mathbf{n}\cdot\nabla\phi = -\mathbf{e}_r\cdot\nabla\phi = -\frac{\partial\phi}{\partial r}\Big|_{r=a} = 0.$$

The condition that $\phi(b,\theta) = \lambda\cos 2\theta$ and the constraint that $|\alpha| = |\beta|$ for $\phi(r,\theta) = Ar^\alpha \cos\beta\theta$ must mean that $|\alpha| = |\beta| = 2$ to match the $\cos 2\theta$ part. Hence the general solution form must be the superposition

$$Br^2\cos 2\theta + \frac{C}{r^2}\cos 2\theta$$

and then we use this together with $\frac{\partial\phi}{\partial r}|_{r=a} = 0$ to find the appropriate particular solution.

5.7 Tensors (to be done)

6 Probability

6.1 Probability Spaces

Let Ω be a set, and let $\mathcal{F}$ be a collection of subsets of Ω. Then $\mathcal{F}$ is a σ-algebra if

- $\Omega \in \mathcal{F}$
- if $A \in \mathcal{F}$ then $A^{\mathsf{c}} \in \mathcal{F}$
- for any countable collection $(A_n)_{n\geq 1}$ with $A_n \in \mathcal{F}$ for all n, we also have $\bigcup_n A_n \in \mathcal{F}$.

(When Ω is countable, we usually take $\mathcal{F}$ to be all subsets of Ω, i.e. its power set - then certainly above conditions hold).

Let $\mathbb{P}$ be a function $\mathbb{P} : \mathcal{F} \to [0, 1]$. Then $\mathbb{P}$ is a probability measure if

- $\mathbb{P}(\Omega) = 1$
- For any countable disjoint collection $(A_n)_{n\geq 1}$ in $\mathcal{F}$, we have $\mathbb{P}(\bigcup_{n\geq 1} A_n) = \sum_{n\geq 1} \mathbb{P}(A_n)$.

When these properties are satisfied, we call $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. From these axioms, we can deduce basic properties of probability.

6.2 Stirling's Formula

Stirling's formula states that as $n \to \infty$, we have

$$n! \sim n^n \sqrt{2\pi n}\, e^{-n}.$$

The proof is non-examinable, but we may have to prove a weaker statement in the exam. This states that

$$\log n! \sim n \log n$$

as $n \to \infty$. To do this we bound using

$$\log\lfloor x \rfloor \leq \log x \leq \log\lfloor x + 1 \rfloor$$

then integrate from 1 to n to get bounds for $\log(n!)$, then divide through by $n \log n$ to get the required result.

6.3 Properties of Probability Measures

Countable subadditivity. Given a countable sequence A_n of events in $\mathcal{F}$, we have

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

To prove this, we construct a particular disjoint sequence B_n of events whose union is the same as the union of A_n, satisfying also that $B_n \subseteq A_n$ for all n. Then it follows that $\mathbb{P}(\cup A_n) = \sum \mathbb{P}(B_n) \leq \sum \mathbb{P}(A_n)$.

Continuity. Let (A_n) be an increasing sequence in $\mathcal{F}$, i.e. $A_n \subseteq A_{n+1}$ for all n. Then

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\cup A_n).$$

This is proved similarly to the countable subadditivity property: we construct a particular disjoint sequence B_n, which satisfies $\bigcup_{k=1}^{n} B_k = A_n$, then we can rewrite expressions in (A_n) using unions in (B_n) and deduce

the required result.

Inclusion-exclusion property. Let $A_1, ..., A_n$ be in $\mathcal{F}$. Then

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} \left((-1)^{k+1} \sum \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k})\right).$$

We prove this by induction, and essentially do a number of manipulations that involve considering $B_i = A_i \cap A_n$. This is essentially bookwork.

The Bonferroni inequalities. Truncating the sum in the inclusion-exclusion formula after the rth term gives an overestimate if r is odd and an underestimate if r is even. These are proved by induction.

Independence. Let (A_n) be a sequence of events. Then (A_n) are mutually independent if

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \ldots \mathbb{P}(A_{i_k})$$

for all subsets $A_{i_1}, \ldots, A_{i_k}$ of (A_n). Note that pairwise independence does not imply mutual independence.

6.4 Conditional Probability

We define the probability of A given B by

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Countable additivity (conditional). Suppose (A_n) is a disjoint sequence in $\mathcal{F}$. Then

$$\mathbb{P}((\cup A_n)|B) = \sum_n \mathbb{P}(A_n|B).$$

The proof follows from using the conditional probability definition, rewriting the union as $(\cup A_n) \cap B = \cup(A_n \cap B)$, and using the usual countable additivity.

Law of total probability. Consider $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose (B_n) is a disjoint collection of events in $\mathcal{F}$, with $\cup B_n = \Omega$ and $\mathbb{P}(B_n) > 0$ for all n. Then

$$\mathbb{P}(A) = \sum_n \mathbb{P}(A|B_n)\,\mathbb{P}(B_n).$$

To prove this, we write $\mathbb{P}(A) = \mathbb{P}(A \cap \Omega) = \mathbb{P}(A \cap (\cup B_n))$ and then exploit the disjoint property of (B_n) and countable additivity to get the required result.

Bayes' Formula. Suppose that (B_n) are disjoint events with $\cup B_n = \Omega$ and $\mathbb{P}(B_n) > 0$ for all n. Then

$$\mathbb{P}(B_n|A) = \frac{\mathbb{P}(A|B_n)\,\mathbb{P}(B_n)}{\sum_k \mathbb{P}(A|B_k)\,\mathbb{P}(B_k)}.$$

This follows straight away from expanding out $\mathbb{P}(B_n|A)$ and then rewriting $\mathbb{P}(A)$ using the law of total probability.

6.5 Probability Distributions

Bernoulli distribution. Distribution $\text{Ber}(p)$ takes parameter p: here $\Omega = \{0, 1\}$ with $\mathbb{P}(1) = p$ and $\mathbb{P}(0) = 1 - p$.

Binomial distribution. Distribution is $\text{Bin}(n, p)$ with n the number of trials (all independent) and p the probability of success. Then $\mathbb{P}(\text{k successes}) = \binom{n}{k} p^k (1-p)^{n-k}$.

Multinomial distribution. Distribution $M(n, p_1, \ldots, p_k)$ where probabilities sum to 1. Models throwing n balls into k boxes, where $\mathbb{P}(\text{pick box } i) = p_i$, independently. Then

$$\mathbb{P}(n_1 \text{ in box } 1, \ldots, n_k \text{ in box } k) = p_1^{n_1} \ldots p_k^{n_k} \frac{n!}{n_1! \ldots n_k!}$$

Geometric distribution. Parameter p of getting a head: toss a p-biased coin until the first head appears. Then $\mathbb{P}(\text{k tosses}) = (1-p)^{k-1} p$.

Poisson distribution. Parameter λ: models number of occurrences of an event in a given time interval. For k occurrences:

$$\mathbb{P}(\text{k occurrences}) = \frac{e^{-\lambda} \lambda^k}{k!}$$

6.6 Expectation and Variance

Independence of random variables. Let $X_1, \ldots, X_n$ be discrete random variables. Then $X_1, \ldots, X_n$ are independent if

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \ldots \mathbb{P}(X_n = x_n).$$

Expectation. Let X be a non-negative discrete random variable. Then the expectation of X is given by

$$\mathbb{E}(X) = \sum_{x \in \Omega} x \, \mathbb{P}(X = x).$$

From this we can quickly show that if $\mathbb{E}(X) = np$ if $X \sim \text{Bin}(n, p)$ and $\mathbb{E}(X) = \lambda$ if $X \sim \text{Poi}(\lambda)$. From the definition we can prove the standard properties of expectation.

Countable additivity for expectation. Let $X_1, X_2, \ldots$ be non-negative random variables. Then the following can be proved from a few quick rearrangements:

$$\mathbb{E}\left(\sum_n X_n \right) = \sum_n \mathbb{E}(X_n).$$

Functions of random variables. Let $g : \mathbb{R} \to \mathbb{R}$. Define $g(X)$ to be the random variable corresponding to applying g to X. Then

$$\mathbb{E}(g(X)) = \sum_{x \in \Omega} g(x) \, \mathbb{P}(X = x).$$

We can prove this by letting $Y = g(X)$, rewriting the event $\{Y = y\}$ in the form $\{X \in g^{-1}(y)\}$ and expanding out $\mathbb{E}(Y)$.

Variance. We define

$$\text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

From this we can prove standard properties fairly quickly.

Covariance. Let X and Y be random variables. Then their covariance is given by

$$\mathrm{Cov}(X,Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y).$$

This measures how dependent on each other X and Y are. There are a couple of useful identities, which can mostly just be proved by expanding expectation or variance:

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$$

$$\mathrm{Cov}(X+Y,Z) = \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z).$$

There is a more general version of the second identity - see notes.

Covariance and Independence. If X and Y are independent random variables and $f, g : \mathbb{R} \to \mathbb{R}^+$ then

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\,\mathbb{E}(g(Y)).$$

Prove this by expanding the standard expectation expression. Also note that if X and Y are independent, then their covariance is zero, but zero covariance does not necessarily imply independence.

6.7 Inequalities

Markov's inequality. Let X be a non-negative random variable. Then for all $a > 0$:

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

This is proved by writing $X \geq a \cdot 1(X \geq a)$ and taking expectations.

Chebyshev's inequality. Let X be a random variable with finite expectation. Then for all $a > 0$:

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}$$

To prove, square both sides of the inequality and apply Markov.

Cauchy-Schwarz inequality. Let X and Y be random variables. Then

$$|\,\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)\,\mathbb{E}(Y^2)}.$$

To prove, we may assume that $X, Y \geq 0$ with finite expectations. Assume that $\mathbb{E}(X^2), \mathbb{E}(Y^2) > 0$. Let t be real, and consider $\mathbb{E}((X - tY)^2) \geq 0$, then expand and minimise the expectation by choosing $t = \mathbb{E}(XY)/\mathbb{E}(Y^2)$. The inequality follows from the non-negativity of the expectation.

Jensen's inequality. Let f be a convex function and X a random variable. Then

$$\mathbb{E}(f(X)) \geq f(\mathbb{E}(X)).$$

First show that if f is convex, then f is the supremum of all the lines lying below it. Then to prove the inequality we consider $f(m) = am + b$ and set $m = \mathbb{E}(X)$. Note this can be applied to prove the AM-GM inequality.

6.8 Conditional Expectation

Let $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$ and let X be a random variable. Then we define the conditional expectation by

$$\mathbb{E}(X|B) = \frac{\mathbb{E}(X \cdot 1(B))}{\mathbb{P}(B)}$$

Law of total expectation. Suppose $X \geq 0$ and let (Ω_n) be a partition of Ω (into disjoint events). Then

$$\mathbb{E}(X) = \sum_n \mathbb{E}(X|\Omega_n)\,\mathbb{P}(\Omega_n).$$

(Note this has essentially the same form as the law of total probability).
To prove this, we write $X = X \cdot 1(\Omega) = \sum_n X \cdot 1(\Omega_n)$ then take expectations and use countable additivity.

Convolution and conditional expectation. Let X and Y be two discrete independent random variables. Then

$$\mathbb{P}(X+Y=z) = \sum_y \mathbb{P}(X+Y=z, Y=y) = \sum_y \mathbb{P}(X=z-y, Y=y) = \sum_y \mathbb{P}(X=z-y)\,\mathbb{P}(Y=y).$$

Recall the conditional expectation formula implies that

$$\mathbb{E}(X|Y=y) = \sum_x x\,\mathbb{P}(X=x|Y=y).$$

Note that $\mathbb{E}(X|Y)$ is a random variable that depends only on Y.

There are some properties of conditional expectation that we can check from the definition. In particular we have

$$\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X).$$

To prove this, we write

$$\mathbb{E}(X|Y) = \sum_y 1(Y=y)\,\mathbb{E}(X|Y=y)$$

and take expectations of both sides. Alternatively after doing this, we can write

$$\sum_y \mathbb{E}(X|Y=y)\,\mathbb{P}(Y=y) = \sum_x \sum_y x\,\mathbb{P}(X=x|Y=y)\,\mathbb{P}(Y=y) = \mathbb{E}(X).$$

We also have

$$\mathbb{E}(\mathbb{E}(X|Y)|Z) = \mathbb{E}(X).$$

Since it is a function of Y, we can write $\mathbb{E}(X|Y) = g(Y)$ which is independent of Z, and hence $\mathbb{E}(g(Y)|Z) = \mathbb{E}(g(Y)) = \mathbb{E}(X)$ using the above relationship.
We can also show similarly that

$$\mathbb{E}(h(Y)X|Y) = h(Y)\,\mathbb{E}(X|Y).$$

6.9 Random Walks

Let $a \in \mathbb{Z}^+$ and choose $x \in \mathbb{Z}$ with $0 < x < a$. We start at x and at every step, either increase by 1 with probability p or decrease by 1 with probability $q = 1-p$ (representing this random walk by (X_n)). What is the probability that we reach a before reaching 0?
We define $h(x) = \mathbb{P}(\text{hit } a \text{ before } 0 | X_0 = x)$ and then apply law of total probability to solve a recurrence for $h(x)$ and hence derive

$$h(x) = \frac{\left(\frac{q}{p}\right)^x - 1}{\left(\frac{q}{p}\right)^a - 1}$$

We can similarly find the "expected time to absorption" τ_x (the expected number of steps to hit either 0 or a) and use law of total expectation conditioning on the first step to show that for $p = 1/2$

$$\tau_x = x(a - x)$$

and for $p \neq 1/2$

$$\tau_x = \frac{x}{q-p} - \frac{q}{q-p} \frac{\left(\frac{q}{p}\right)^x - 1}{\left(\frac{q}{p}\right)^a - 1}$$

6.10 Probability Generating Functions

Let X be a random variable taking positive integer values. Define the probability generating function $p(z)$ (for $|z| \leq 1$) by

$$p(z) = \sum_{r=0}^{\infty} p_r z^r = \mathbb{E}(z^X).$$

This uniquely determines the distribution of X - check essentially by induction on nth partial sums.

Expectation and Variance. We can show that

$$\lim_{z \to 1^-} p'(z) = p'(1) = \mathbb{E}(X)$$

and that

$$\mathrm{Var}(X) = p''(1) + p'(1) - (p'(1))^2.$$

Sum of a Random Number of Variables. To be added.

Another Conditional Expectation Proof. To be added.

6.11 Branching Processes

Consider a random process where every individual produces a random number of offspring. Let X_n be the number of individuals in generation n, where we start with $X_0 = 1$.

Iteratively, we can show that

$$\mathbb{E}(X_n) = (\mathbb{E}(X_1))^n$$

and also that if $G(z) = \mathbb{E}(z^{X_1})$ and $G_n(z) = \mathbb{E}(z^{X_n})$ (generating functions) then

$$G_{n+1}(z) = G(G_n(z)) = G_n(G(z))$$

for all n.

Extinction probability. Assume that $\mathbb{P}(X_1 = 1) < 1$. Then the extinction probability is the minimal non-negative solution to $t = G(t)$.

6.12 Continuous Random Variables

For a random variable X, the probability distribution function $F(x)$ is $F : \mathbb{R} \to [0, 1]$ with

$$F(x) = \mathbb{P}(X \leq x).$$

The probability density function f is given by

$$f(x) = F'(x).$$

Expectation. Let $X \geq 0$ with density f. Then

$$\mathbb{E}(X) = \int_0^\infty x f(x)\,\mathrm{d}x = \int_0^\infty \mathbb{P}(X \geq x)\,\mathrm{d}x$$

and

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

Memoryless property. This is the property that for a random variable T, given $s, t > 0$ we have

$$\mathbb{P}(T > t + s | T > s) = \mathbb{P}(T > t).$$

This essentially means that any time may be marked off as zero and the probability distribution is independent of its history.

We can show that if T is a continuous positive random variable, then T has the memoryless property if and only if it is exponential (pdf $\lambda e^{-\lambda x}$). The idea is essentially solving a functional equation.

Density of function of a random variable. Let X be a continuous random variable with density f and let g be a strictly monotone function with g^{-1} differentiable. Then $g(X)$ is a continuous random variable with density

$$f(g^{-1}(x)) \left| \frac{\mathrm{d}}{\mathrm{d}x} g^{-1}(x) \right|.$$

The proof follows from differentiating the probability distribution function and doing some rearrangements.

6.13 Multivariate Density Functions

We can generalise this concept to several variables. Let $X = (X_1, \ldots, X_n)$ be a multivariate random variable. We say X has density f if

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(y_1, \ldots, y_n)\,\mathrm{d}y_1 \ldots \mathrm{d}y_n.$$

Then we have

$$f(x_1, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \ldots \partial x_n} F(x_1, \ldots, x_n).$$

Independence. We say $X_1, \ldots, X_n$ are independent if for all real $x_1, \ldots, x_n$, we have

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \mathbb{P}(X_1 \leq x_1) \ldots \mathbb{P}(X_n \leq x_n).$$

If the X_i are independent, then the probability densities factorise:

$$f(x_1, \ldots, x_n) = f_1(x_1) \ldots f_n(x_n).$$

We can check this by writing out the probability distribution functions as integrals of the density functions.

Marginal density. Suppose $(X_1, \ldots, X_n)$ has density f. Then

$$\begin{aligned}\mathbb{P}(X_1 \leq x) &= \mathbb{P}(X_1 \leq x, X_2 \in \mathbb{R}, \ldots, X_n \in \mathbb{R}) \\ &= \int_{-\infty}^{x} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n) \, \mathrm{d}x_1 \ldots \mathrm{d}x_n \\ &= \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n) \, \mathrm{d}x_2 \ldots \mathrm{d}x_n \right) \mathrm{d}x_1.\end{aligned}$$

The integrand is the marginal density of X_1.

Convolution. Let X and Y be independent random variables with densities f_X and f_Y. Then

$$\mathbb{P}(X + Y \leq z) = \int_{-\infty}^{z} \int_{-\infty}^{\infty} f_Y(y - x) f_X(x) \, \mathrm{d}x.$$

The convolution of X and Y is the density of $X + Y$, which is

$$\int_{-\infty}^{\infty} f_Y(y - x) f_X(x) \, \mathrm{d}x.$$

We can check this by writing out the integrals and factorising the density function (as X and Y are independent).

6.14 Conditional Probability and Transformations

Law of total probability. Let X and Y be continuous variables with joint density $f_{X,Y}$ and marginal densities f_X and f_Y. Then the conditional density of X given $Y = y$ is

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

If we define g by

$$g(y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) \, \mathrm{d}x$$

then the conditional expectation can be found by setting

$$\mathbb{E}(X|Y) = g(Y).$$

Transformation of a multidimensional random variable. Let X be a random variable with values in $D \subset \mathbb{R}^d$ and density f_X. Let g be a bijection D to $g(D)$ with continuous derivative and non-vanishing determinant. Then the random variable $Y = g(X)$ has density

$$f_Y(y) = f_X(x)|J|$$

where $x = g^{-1}(y)$ and

$$J = \det \left(\left(\frac{\partial x_i}{\partial y_j} \right)_{i,j=1}^{d} \right).$$

This is not proved in this course.

Order statistics for a random sample. Let $X_1, \ldots, X_n$ be iid with distribution F and density f. We can put them in increasing order

$$Y_1 \leq Y_2 \leq \cdots \leq Y_n$$

and here (Y_i) are called the order statistics. We have

$$\mathbb{P}(Y_1 \leq x) = \mathbb{P}(\min(X_1, \ldots, X_n) \leq x) = 1 - \mathbb{P}(\min(X_1, \ldots, X_n) > x) = 1 - (1 - F(x))^n.$$

In general we have that

$$\mathbb{P}(Y_n \leq x) = (F(x))^n$$

and

$$f_{Y_n}(x) = n(F(x))^{n-1} f(x).$$

The density of $Y_1, \ldots, Y_n$ is

$$f_{Y_1,\ldots,Y_n}(x_1, \ldots, x_n) = n! f(x_1) \ldots f(x_n)$$

when $x_1 < \cdots < x_n$ and zero otherwise. Check this by writing the probability distribution out in integral form again.

6.15 Moment Generating Functions

Let X be a random variable with density f. Then the moment generating function of X is

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

whenever this is finite; we set $m(0) = 1$. This uniquely determines the distribution of the random variable provided that it is defined for an open interval of values of θ. If this is the case, then we also have

$$m^{(r)}(0) = \frac{\mathrm{d}^r}{\mathrm{d}\theta^r} m(\theta)|_{\theta=0} = \mathbb{E}(X^r).$$

6.16 Limit Theorems

Weak law of large numbers. Let $(X_n : n \in \mathbb{N})$ be a sequence of iid random variables with $\mu = \mathbb{E}(X_1) < \infty$. Set $S_n = X_1 + \cdots + X_n$. Then for all $\varepsilon > 0$:

$$\mathbb{P}\left(\left| \frac{S_n}{n} - \mu \right| > \varepsilon \right) \to 0 \text{ as } n \to \infty.$$

We prove this by applying Chebyshev's inequality.

Convergence in probability and almost surely. A sequence X_n converges to X in probability (write $X_n \xrightarrow{\mathbb{P}} X$ as $n \to \infty$) if for all $\varepsilon > 0$, as $n \to \infty$ we have

$$\mathbb{P}(|X_n - X| > \varepsilon) \to 0.$$

A sequence X_n converges to X almost surely if

$$\mathbb{P}\left(\lim_{n \to \infty} X_n = X \right) = 1.$$

If $X_n \to 0$ almost surely as $n \to \infty$, then $X_n \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$. To prove this we show that $\mathbb{P}(|X_n| \leq \varepsilon) \to 1$ as $n \to \infty$.

Strong law of large numbers. Let $(X_n)_{n \in \mathbb{N}}$ be an iid sequence of random variables with finite expectation μ. Let $S_n = X_1 + \cdots + X_n$. Then

$$\frac{S_n}{n} \to \mu$$

as $n \to \infty$ almost surely.
The proof is non-examinable - refer to notes.

Central limit theorem. Let $(X_n)_{n\in\mathbb{N}}$ be iid random variables with expectation μ and variance σ^2. Let $S_n = X_1 + \cdots + X_n$. Then for all $x \in \mathbb{R}$, as $n \to \infty$,

$$\mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) \to \Phi(x) = \int_{-\infty}^{x} \frac{e^{-y^2/2}}{\sqrt{2\pi}}\,\mathrm{d}y.$$

This says that for large enough n,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \approx Z \implies S_n \approx n\mu + \sigma\sqrt{n}Z \sim N(n\mu, \sigma^2 n).$$

Summing up the proof briefly isn't the easiest thing, so see notes for the proof.

Using the central limit theorem, we can find the required population size for a sample such that the sampling error is as small as required.

Buffon's needle. Suppose we have many parallel lines, each a distance L apart. Suppose we have a needle of length $l \leq L$. If we throw the needle at random, the probability that it intersects a line is $2l/\pi L$ (taking angle and relative position to be independent; $\theta \sim U[0, \pi]$ and $X \sim U[0, L]$). A line is intersected if and only if $X \leq l \sin\theta$.
Hence

$$\pi = \frac{2l}{pL}$$

and we can approximate π by repeating this experiment, and get to an arbitrary degree of accuracy by making the sample size large enough.

6.17 Multidimensional Random Variables

A random variable X with real values is Gaussian or normal if $X = \mu + \sigma Z$ where $Z \sim N(0, 1)$. Then X has density

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

Gaussian vectors. Let $X = (X_1, \ldots, X_n)^{\mathrm{T}} \in \mathbb{R}^n$. Then X is a Gaussian vector if for all $u = (u_1, \ldots, u_n)^{\mathrm{T}}$ in $\mathbb{R}^n$, we have that

$$u^{\mathrm{T}}X = \sum_{i=1}^{n} u_i X_i$$

is a Gaussian random variable in $\mathbb{R}$.

Mean and variance of Gaussian vectors. We have

$$\mu = \mathbb{E}(X) = \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_n) \end{pmatrix}$$

and $V = \mathrm{Var}(X) = \mathbb{E}((X - \mu)(X - \mu)^{\mathrm{T}})$ so V is a symmetric matrix with

$$V_{ij} = \mathbb{E}((X_i - \mu_i)(X_j - \mu_j)) = \mathrm{Cov}(X_i, X_j).$$

We can check from the definitions and it follows that

$$u^{\mathrm{T}}X \sim N(u^{\mathrm{T}}\mu, u^{\mathrm{T}}Vu).$$

The moment generating function of X is

$$m(\theta) = \mathbb{E}(e^{\theta^{\mathrm{T}}X}) = \exp\left(\theta^{\mathrm{T}}\mu + \frac{\theta^{\mathrm{T}}V\theta}{2}\right)$$

for all $\theta \in \mathbb{R}^n$.

Construction of Gaussian vectors. Let $Z_1, \ldots, Z_n$ be iid $N(0,1)$ random variables. Then

$$Z = (Z_1, \ldots, X_n)^{\mathrm{T}}$$

is a Gaussian vector. We can check this by considering the generating function.

Constructing a Gaussian vector with given mean and variance. To be added.

Density of Gaussian vectors. Let $X \sim N(\mu, V)$, then

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det V}} \exp\left(-\frac{(x-\mu)^{\mathrm{T}}V^{-1}(x-\mu)}{2}\right).$$

6.18 More Properties of Gaussian Vectors

If the X_i components of X are independent, then V must be a diagonal matrix (zero covariance for distinct components). In fact, for a Gaussian vector $X_1, \ldots, X_n$ are each independent if and only if $\mathrm{Cov}(X_i, X_j) = 0$ whenever $i \neq j$.

Bivariate Gaussian vectors. To be added.

6.19 Rejection Sampling and Simulation

Let X be a continuous random variable with distribution function F. Then if U is uniform on $[0,1]$, then $F^{-1}(U) \sim F$. (Just prove by setting $Y = F^{-1}(U)$ and considering $\mathbb{P}(Y \leq X)$).

Rejection sampling. To be added.

7 Vectors and Matrices

7.1 Complex Numbers

Logarithms. We define

$$w = \log z$$

for complex nonzero z by

$$e^w = e^{\log z} = z$$

which is a multivalued function since exp is many-to-one. We can write $z = re^{i\theta} = e^{\log r}e^{i\theta} = e^{\log r + i\theta}$ and hence

$$\log z = \log|z| + i \arg z.$$

We also define complex powers by

$$z^\alpha = e^{\alpha \log z}.$$

7.2 Vectors in General

Linear independence. The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r \in V$ are linearly independent if

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_r\mathbf{v}_r = \mathbf{0} \implies \lambda_1, \lambda_2, \ldots, \lambda_r = 0.$$

Note that this implies that if $\mathbf{0}$ is any of the vectors, then the set must be linearly dependent.

Dimension theorem. This states that any basis for a given vector space always has the same number of basis vectors (called the dimension of the vector space). See other V&M notes for proof.

7.3 Determinants

Definition. Let M be an $n \times n$ matrix with columns $\mathbf{C}_a = M\mathbf{e}_a$. Then the determinant $\det M$ is defined by

$$\det M = \sum_\sigma \varepsilon(\sigma) M_{\sigma(1)1} M_{\sigma(2)2} \ldots M_{\sigma(n)n}$$

where the sum is taken over all permutations σ of $(1, \ldots, n)$. This is the *same* as the alternating form applied to the columns:

$$\det M = [\mathbf{C}_1, \ldots, \mathbf{C}_n].$$

We could equally well have applied it to the rows instead.
Note: the alternating form is multilinear, totally antisymmetric, and satisfies $[\mathbf{e}_1, \ldots, \mathbf{e}_n] = 1$.

7.4 Diagonalisability

If an $n \times n$ matrix A has n distinct eigenvalues, then it is diagonalisable.
(If there are n distinct eigenvalues, then there are n linearly independent eigenvectors, so the eigenvectors form a basis for $\mathbb{R}^n$ and certainly the algebraic and geometric multiplicities coincide).

An $n \times n$ matrix A is diagonalisable *if and only if* the algebraic multiplicity M_λ (the multiplicity of λ as a root of the characteristic polynomial) and the geometric multiplicity m_λ (the dimension of the eigenspace corresponding to λ) *coincide* for every eigenvalue λ:

$$M_\lambda = m_\lambda.$$

7.5 Quadratic Forms

We can diagonalise quadratic forms. Given a vector

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

we have the quadratic form

$$F(\mathbf{x}) = 2x_1^2 - 4x_1x_2 + 5x_2^2 = \mathbf{x}^{\mathrm{T}} A \mathbf{x}$$

where

$$A = \begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix}.$$

The symmetric matrix A has eigenvalues 1 and 6, and eigenvectors

$$\frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ 1 \end{pmatrix} \text{ and } \frac{1}{\sqrt{5}}\begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

so we compute $P^{\mathrm{T}}\mathbf{x}$ where P is the diagonal matrix with the above two eigenvectors as columns, then we set

$$\mathbf{x}' = P^{\mathrm{T}}\mathbf{x}$$

and write

$$F = \lambda_1 x_1'^2 + \lambda_2 x_2'^2 = x_1'^2 + 6x_2'^2.$$

7.6 Change of Basis

Suppose we have a linear map $T : V \to W$ with V and W vector spaces of dimension n and m respectively, and T represented by a matrix A. Then the change of basis formula is

$$A' = Q^{-1}AP$$

where Q and P are defined as shown in the example below.

Example. Let $n = 2$ and $m = 3$ and suppose that

$$T(\mathbf{e}_1) = \mathbf{f}_1 + 2\mathbf{f}_2 - \mathbf{f}_3$$

$$T(\mathbf{e}_2) = -\mathbf{f}_1 + 2\mathbf{f}_2 + \mathbf{f}_3$$

where $\{\mathbf{e}_i\}$ and $\{\mathbf{f}_a\}$ are the original bases for V and W respectively. Then we have

$$A = \begin{pmatrix} 1 & -1 \\ 2 & 2 \\ -1 & 1 \end{pmatrix}.$$

Suppose the new bases $\{\mathbf{e}_i'\}$ and $\{\mathbf{f}_a'\}$ are defined by

$$\mathbf{e}_1' = \mathbf{e}_1 - \mathbf{e}_2$$

$$\mathbf{e}_2' = \mathbf{e}_1 + \mathbf{e}_2$$

and

$$\mathbf{f}_1' = \mathbf{f}_1 - \mathbf{f}_3$$

$$\mathbf{f}_2' = \mathbf{f}_2$$

$$\mathbf{f}_3' = \mathbf{f}_1 + \mathbf{f}_3.$$

Then we obtain the *columns* of the 2×2 matrix P and the 3×3 matrix Q by reading across:

$$P = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \text{ and } Q = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}.$$

Then we compute

$$A' = Q^{-1}AP.$$

8 Dynamics and Relativity

8.1 Energy Conservation

From the energy equation, we can obtain the following equation, which is often useful in solving problems such as "how long does it take for a satellite a great distance away to fall to earth".

$$T = \int_{x_0}^{x} \frac{1}{\sqrt{\frac{2}{m}(E - V(x))}} \, \mathrm{d}x$$

8.2 Gravity, Orbits and Forces

Gravitational forces. The following gives the gravitational force on mass m due to mass M, where $\mathbf{r}$ is the position vector of the mass m relative to M.

$$\mathbf{F}(\mathbf{r}) = -\frac{GMm}{|\mathbf{r}|^3}\mathbf{r} = -\frac{GMm}{r^2}\hat{\mathbf{r}}$$

Angular momentum. This is given by

$$\mathbf{L} = \mathbf{r} \times \mathbf{p} = \mathbf{r} \times m\dot{\mathbf{r}}.$$

The torque or moment of the force $\mathbf{F}$ is given by

$$\mathbf{G} = \frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = m\mathbf{r} \times \ddot{\mathbf{r}} = \mathbf{r} \times \mathbf{F}.$$

Central force fields. A central force field $\mathbf{F}(\mathbf{r})$ has a potential which is a function only of the distance from the origin; that is $V = V(r)$. Then we obtain

$$\mathbf{F}(\mathbf{r}) = -\nabla V(r) = -\frac{\mathrm{d}V}{\mathrm{d}r}\hat{\mathbf{r}}$$

which tells us that the force is either *attractive* or *repulsive* with respect to the origin.

Observe also that $\mathbf{r} \times \mathbf{F} = \mathbf{0}$ for a central force, so angular momentum is **conserved** for a central force. From the definition of angular momentum, we have also that $\mathbf{L} \cdot \mathbf{r} = 0$, so a particle in a central force field must move in a plane normal to the constant angular momentum vector.

The orbit equation. Defining $u = 1/r$, we have the orbit equation

$$\frac{\mathrm{d}^2 u}{\mathrm{d}\theta^2} + u = -\frac{1}{mh^2u^2}F(u^{-1}.$$

We can solve this for $u(\theta)$ and then use $\dot{\theta} = hu^2$ (which comes from the expression for h in the polar coordinates section).

8.3 Kepler's Laws

Kepler's First Law. This states that the orbit of a planet is an ellipse with the sun at one focus. This is consistent with the solution for bound orbits that we obtain from solving the orbit equation.

Kepler's Second Law. A line joining a planet and the sun sweeps out an equal area in equal time intervals.

To (sort of) demonstrate this, we denote the distance between the sun and the planet by r and suppose that a small change $\delta\theta$ in angle occurs in a small change in time δt. Then the area swept out is approximately

$$\delta A \approx \frac{1}{2} r^2 \delta\theta$$

and hence the rate of change of area with respect to time is

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \frac{1}{2} r^2 \dot{\theta} = \frac{h}{2} = \text{constant}.$$

Kepler's Third Law. The square of the orbital period T is proportional to the cube of the semi-major axis a - that is $T^2 \propto a^3$.
The area of an ellipse with semi-major and semi-minor axes a and b is

$$\pi ab = \frac{h}{2} T$$

(from Kepler's 2nd Law).

8.4 Polar Coordinates

As seen earlier, central forces cause a particle to move in a plane, so we can work in plane polar coordinates. We define

$$\mathbf{e}_r = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \text{ and } \mathbf{e}_\theta = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$$

and by differentiating, we find that

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbf{e}_r = \mathbf{e}_\theta \text{ and } \frac{\mathrm{d}}{\mathrm{d}\theta}\mathbf{e}_\theta = -\mathbf{e}_r.$$

We have also

$$\mathbf{r} = r\mathbf{e}_r$$

$$\dot{\mathbf{r}} = \dot{r}\mathbf{e}_r + r\dot{\theta}\mathbf{e}_\theta$$

$$\ddot{\mathbf{r}} = (\ddot{r} - r\dot{\theta}^2)\mathbf{e}_r + (2\dot{r}\dot{\theta} + r\ddot{\theta})\mathbf{e}_\theta$$

and using these to compute the angular momentum, we find that $mr^2\theta$ is constant. We then write that

$$h \coloneqq r^2\dot{\theta}$$

is a conserved quantity, because angular momentum is conserved. (Note that h itself is not the actual angular momentum, but we can essentially use it as such).

8.5 Electromagnetic Forces

The Lorentz force law states that the force $\mathbf{F}$ on a particle with charge q due to an electric field $\mathbf{E}$ and a magnetic field $\mathbf{B}$ is given by

$$\mathbf{F} = q\mathbf{E} + q\dot{\mathbf{r}} \times \mathbf{B}.$$

8.6 Special Relativity

$$\frac{E^2}{c^2} = p^2 + m^2c^2$$