

A first year mathematics degree

Yu Coughlin

# Contents

	<b>1</b>	<b>Analysis</b>	<b>1</b>
Lecture 1	1	Number systems . . . . .	2
	1.1	Naturals, integers and rationals . . . . .	2
	1.2	Decimal expansions . . . . .	2
	1.3	Countability . . . . .	3
	2	Bounded sets . . . . .	3
	2.1	Supremums and infimums . . . . .	3
	2.2	Completeness . . . . .	3
	2.3	Dedekind cuts . . . . .	3
	3	Sequences . . . . .	3
	3.1	Convergence . . . . .	3
	3.2	Divergence . . . . .	3
	3.3	Limits . . . . .	4
	3.4	Monotone sequences . . . . .	4
	3.5	Cauchy sequences . . . . .	4
	3.6	Subsequences . . . . .	4
	4	Series . . . . .	4
	4.1	Convergence . . . . .	5
	4.2	Absolute convergence . . . . .	5
	4.3	Rearrangement of series . . . . .	6
	4.4	Power series . . . . .	6
	4.5	Exponential series . . . . .	7
	5	Continuity . . . . .	7
	5.1	Continuous functions . . . . .	7
	5.2	Properties of continuity . . . . .	8
	6	Properties of subsets . . . . .	8
	6.1	Open sets . . . . .	8
	6.2	Closed and compact sets . . . . .	9
	7	Uniform continuity and convergence . . . . .	9
	7.1	Uniform continuity . . . . .	9
	7.2	Convergence of sequences of functions . . . . .	9
	7.3	Convergence of series of functions . . . . .	9
	8	Differentiation . . . . .	10
	8.1	Differentiability . . . . .	10
	8.2	Local extrema and mean values . . . . .	10
	8.3	L'Hôpital's rule . . . . .	10
	8.4	Taylor's theorem . . . . .	11
	8.5	Convexity . . . . .	11
	8.6	Exchange of limits and derivatives . . . . .	11
	8.7	Trigonometric properties . . . . .	12
	9	Integration . . . . .	12
	9.1	Partitions . . . . .	12
	9.2	Darboux sums . . . . .	12
	9.3	Darboux integral . . . . .	12
	9.4	Properties of integration . . . . .	13
	9.5	Fundamental theorems of calculus . . . . .	13
	9.6	Methods of integration . . . . .	13

9.7	Limits and integrals . . . . .	14
9.8	Improper integrals . . . . .	14

## 2 Linear Algebra 15

Lecture 1	1	Linear Systems and matrices . . . . .	16
	1.1	Linear systems . . . . .	16
	1.2	Matrix algebra . . . . .	16
	1.3	EROs . . . . .	16
	1.4	Matrices of note . . . . .	17
	2	Vector Spaces . . . . .	17
	2.1	Fields . . . . .	17
	2.2	Vector spaces . . . . .	18
	2.3	Subspaces . . . . .	18
	3	Spanning and Linear Independence . . . . .	18
	3.1	Spanning . . . . .	18
	3.2	Linear independence . . . . .	18
	4	Bases . . . . .	19
	4.1	Definition . . . . .	19
	4.2	Dimension . . . . .	19
	5	Matrix rank . . . . .	19
	6	Linear transformations . . . . .	20
	6.1	Definition . . . . .	20
	6.2	Image and kernel . . . . .	20
	6.3	Rank nullity . . . . .	20
	7	Representations . . . . .	20
	7.1	Matrices of transformations . . . . .	20
	7.2	Matrices of transformations . . . . .	21
	8	Determinants . . . . .	21
	8.1	Definition . . . . .	21
	8.2	Properties . . . . .	22
	9	Eigen-things . . . . .	22
	9.1	Eigenvectors and eigenvalues . . . . .	22
	9.2	Characteristic polynomial . . . . .	22
	9.3	Diagonalisation . . . . .	23
	10	Orthogonality . . . . .	23
	10.1	Inner product spaces* . . . . .	23
	10.2	Orthonormal sets . . . . .	23
	10.3	Gramm-Schmidt process . . . . .	23
	11	Real symmetric matrices . . . . .	24
	11.1	Introduction . . . . .	24
	11.2	Spectral theorem . . . . .	24

## 3 Groups 25

Lecture 1	1	Binary operations and groups . . . . .	26
	2	Subgroups . . . . .	26
	2.1	Subgroups . . . . .	26
	2.2	Cyclic groups and orders . . . . .	27
	2.3	Cosets . . . . .	27
	2.4	Lagrange's theorem . . . . .	27
	2.5	Generating groups . . . . .	27
	3	Group homomorphisms . . . . .	28
	4	Symmetric groups . . . . .	28
	4.1	Disjoint cycle decomposition . . . . .	28
	4.2	Alternating groups . . . . .	29
	4.3	Dihedral groups . . . . .	29

	<b>4</b>	<b>Calculus</b>	<b>30</b>
Lecture 1	1	Lengths, volumes and surfaces . . . . .	31
	1.1	Lengths . . . . .	31
	1.2	Volumnes and volumes of revolution . . . . .	31
	1.3	Surfaces . . . . .	31
	1.4	Centres of mass . . . . .	32
	1.5	Moments of inertia . . . . .	32
	1.6	Polar coordinates . . . . .	32
	2	Fourier series . . . . .	32
	2.1	Orthogonal and orthonormal function spaces . . . . .	32
	2.2	Periodic functions . . . . .	33
	2.3	Trigonometric polynomials . . . . .	33
	2.4	Fourier series . . . . .	33
	3	Plane curves . . . . .	34
	4	Laplace transform . . . . .	34
	<b>5</b>	<b>Differential Equations</b>	<b>35</b>
Lecture 1	1	Fourier transform . . . . .	36
	2	Ordinary differential equations . . . . .	36
	3	Qualitative analysis . . . . .	36
	4	Bifurcations . . . . .	36
	5	Multivariate calculus . . . . .	36
	6	Partial differential equations . . . . .	36
	<b>6</b>	<b>Probability</b>	<b>37</b>
Lecture 1	0.1	Sample spaces and set theory . . . . .	38
Lecture 2	0.2	Interpretation of probability . . . . .	38
Lecture 3	1	Counting . . . . .	39
	1.1	Multiplication principle . . . . .	39
	1.2	Power sets . . . . .	39
	1.3	Combinatorial coefficients . . . . .	39
Lecture 4	1.4	Sampling with and without replacement . . . . .	39
Lecture 5	2	Axiomatic probability . . . . .	40
	2.1	Event space . . . . .	40
Lecture 6	2.2	Probability measure . . . . .	40
	2.3	Probability space . . . . .	40
Lecture 7	3	Conditional probability . . . . .	40
Lecture 8	3.1	Bayes' rule and total probability . . . . .	40
Lecture 9	4	Independence . . . . .	41
	4.1	Event independence . . . . .	41
Lecture 10	4.2	Conditional independence . . . . .	41
	4.3	Product rule for general independence . . . . .	41
	5	Discrete random variables . . . . .	42
Lecture 11	5.1	Images and their properties . . . . .	42
	5.2	DRVs and their distributions . . . . .	42
	6	Common DRVs . . . . .	43
	6.1	Bernoulli distribution . . . . .	43
	6.2	Binomial distribution . . . . .	43
	6.3	Hypergeometric distribution . . . . .	43
	6.4	Discrete uniform distribution . . . . .	43
	6.5	Poisson distribution . . . . .	43
	6.6	Geometric distribution . . . . .	44
	6.7	Negative binomial distribution . . . . .	44
	7	Continuous random variables . . . . .	44
	7.1	General random variables and their distributions . . . . .	44
	7.2	CRVs and pdfs . . . . .	45
	8	Common CRVs . . . . .	45
	8.1	Uniform distribution . . . . .	45
	8.2	Exponential distribution . . . . .	45
	8.3	Gamma distribution . . . . .	45

	8.4	Chi-squared distribution . . . . .	46
	8.5	F-distribution . . . . .	46
	8.6	Beta distribution . . . . .	46
	8.7	Normal distribution . . . . .	46
	8.8	Cauchy distribution . . . . .	46
	8.9	Student t-distribution . . . . .	47
9		Transformations of random variables . . . . .	47
	9.1	DRVs . . . . .	47
	9.2	CRVs . . . . .	47
10		Expectation of random variables . . . . .	47
	10.1	Definition . . . . .	47
	10.2	LOTUS . . . . .	47
	10.3	Variance . . . . .	48
11		Multivariate random variables . . . . .	48
	11.1	Multivariate distributions . . . . .	48
	11.2	Independence . . . . .	48
	11.3	Multivariate DRVs . . . . .	48
	11.4	Multivariate CRVs . . . . .	49
	11.5	Transformations of random vector . . . . .	49
	11.6	Multivariate LOTUS . . . . .	49
	11.7	Covariance . . . . .	49
12		Generating functions . . . . .	50
	12.1	Probability generating functions . . . . .	50
	12.2	Common pgfs . . . . .	50
	12.3	Moment generating functions . . . . .	50
13		Conditional distribution and expectation . . . . .	51
	13.1	Discrete: Conditional expectation and total expectation . . . . .	51
	13.2	Conditioning on a DRV . . . . .	51
	13.3	Continuous: Conditional density, distribution and expectation . . . . .	51
<b>7</b>	<b>Statistics</b>		<b>52</b>
Lecture 1	1	Introduction . . . . .	52
	2	Central tendency and dispersion . . . . .	53
	2.1	Mean, variance and moments . . . . .	53
	2.2	Parameter estimation . . . . .	53
	2.3	Other measures of central tendency . . . . .	53
	2.4	Sampling from normal RVs . . . . .	53
	3	Hypothesis testing . . . . .	53
	3.1	Introduction . . . . .	53
	3.2	Single sample hypothesis testing . . . . .	53
	3.3	Distribution of p-values . . . . .	53
	3.4	Errors . . . . .	53
	3.5	Two sample hypothesis testing . . . . .	53
	3.6	Multiple hypothesis testing . . . . .	53
	4	Covariance and Correlations . . . . .	53
	4.1	Covariance . . . . .	53
	4.2	Correlation . . . . .	53
	5	Statistical models . . . . .	53
	5.1	Definitions . . . . .	53
	5.2	Likelihood . . . . .	53
	5.3	Linear regression . . . . .	53
	6	Bayesian inference . . . . .	53
	6.1	Definitions . . . . .	53
	6.2	Conjugate pair distributions . . . . .	53
	6.3	Intractable posteriors . . . . .	53
	6.4	Choosing a prior . . . . .	53
	7	Bootstrap . . . . .	53
	7.1	Empirical distribution . . . . .	53
	7.2	Bootstrap procedure . . . . .	53

	<b>8</b>	<b>Computation</b>	<b>54</b>
	1	Introduction . . . . .	54
Lecture 1	<b>9</b>	<b>Applied Mathematics</b>	<b>55</b>
	1	Introduction . . . . .	55

# Chapter 1

# Analysis

Lectured by Dr Ajay Chandra  
Typed by Yu Coughlin  
Autumn 2023 & Spring 2024

## Introduction

The following are core reading:

- M W Liebeck, A concise intro to pure mathematics, 2016
- F M Hart, Guide to analysis, 2001
- D A Brannan, A first course in mathematical analysis, 2006
- K G Binmore, Mathematical analysis: a straightforward approach, 1982

The following are supplementary reading:

- S R Lay, Analysis with an introduction to proof, 1994
- S Abbott, Understanding Analysis, 2015

The following are advanced reading:

- M Spivak, Calculus, 2010
- T Tao, Analysis 1, 2022

Lecture 1  
Thursday  
10/01/19

# 1 Number systems

## 1.1 Naturals, integers and rationals

**Definition 1.1.1** (Natural numbers). As in IUM, we define the **natural numbers**,  $\mathbb{N}$ , from the Peano axioms:

P1  $0$  is a natural number,

P6 if  $n$  is a natural number then  $S(n)$  is a natural number where  $S(n)$  is the successor of  $n$ ,

P9 the principle of mathematical induction.

Clearly, there are many Peano axioms not included, these are however not particularly relevant to this course. Addition and multiplication is defined as expected and will descend to our other number systems

**Definition 1.1.2** (Integers). The **integers** are defined as  $\mathbb{Z} := \mathbb{N} \times \mathbb{N} / \sim$  where  $\sim$  is the equivalence relation given by  $(a, b) \sim (c, d)$  iff  $a + d = b + c$ . Subtraction is defined as expected and will also descend to our other number systems.

**Definition 1.1.3** (Rationals). The **rationals** are defined as  $\mathbb{Q} := \mathbb{Z} \times \mathbb{N}^{>0} / \sim$  where  $\sim$  is the equivalence relation given by  $(a, b) \sim (c, d)$  iff  $ad = bc$ . The equivalence class  $(p, q)$  will be written as  $\frac{p}{q}$ . There is an element of each equivalence class  $\frac{p'}{q'}$  with  $\gcd(p', q') = 1$ , we say that  $\frac{p'}{q'}$  is in **lowest terms**.

**Theorem 1.1.4** (Axioms of the rationals). With the usual operations descended from  $\mathbb{N}$  and  $\mathbb{Z}$ ,  $\mathbb{Q}$  satisfies the following axioms with  $a, b, c \in \mathbb{Q}$  throughout:

Q1  $a + (b + c) = (a + b) + c$  ( $+$  is associative),

Q2  $\exists 0 \in \mathbb{Q}$  such that  $a + 0 = a$  ( $0$  is the additive identity of  $\mathbb{Q}$ ),

Q3  $\forall a \in \mathbb{Q}, \exists (-a) \in \mathbb{Q}$  such that  $a + (-a) = 0$  ( $\mathbb{Q}$  is closed under additive inverses),

Q4  $a + b = b + a$  ( $+$  is commutative),

Q5  $a \times (b \times c) = (a \times b) \times c$  ( $\times$  is associative),

Q6  $\exists 1 \in \mathbb{Q}$  such that  $a \times 1 = a$  ( $1$  is the multiplicative identity of  $\mathbb{Q}$ ),

Q7  $a \times (b + c) = (a \times b) + (a \times c)$  ( $\times$  is left distributive over  $+$ ),

Q8  $(a + b) \times c = (a \times c) + (b \times c)$  ( $\times$  is right distributive over  $+$ ),

Q9  $a \times b = b \times a$  ( $\times$  is commutative),

Q10  $\forall a \in \mathbb{Q}, \exists a^{-1} \in \mathbb{Q}$  such that  $a \times a^{-1} = 1$  ( $\mathbb{Q}$  is closed under multiplicative inverses),

Q11 for all  $a \in \mathbb{Q}$  either  $x < 0$ ,  $x = 0$  or  $x > 0$  (Trichotomy),

Q12 for all  $x, y \in \mathbb{Q}$  we have  $x > 0, y > 0 \implies x + y > 0$ ,

Q13 for all  $x \in \mathbb{Q}$  there exists a  $n \in \mathbb{N}$  such that  $x < n$  (Archimedean axiom).

1-4 says  $(\mathbb{Q}, +)$  is an abelian group, 1-9 says  $(\mathbb{Q}, +, \times)$  is a commutative ring, 1-10 says  $(\mathbb{Q}, +, \times)$  is a field.

## 1.2 Decimal expansions

**Definition 1.2.1.** For  $a_0 \in \mathbb{N}$  and  $a_i \in [1, 9]$  for  $i > 0 \in \mathbb{N}$ , define the **periodic decimal**

$$a_0.a_1a_2 \dots \overline{a_ia_{i+1} \dots a_j},$$

to be equal to the rational number

$$a_0 + \frac{a_1}{10} + \frac{a_2}{100} + \dots + \frac{a_i}{10^i} + \left( \frac{a_{i+1}a_{i+2} \dots a_j}{10^j} \right) \left( \frac{1}{1 - 10^{i-j}} \right).$$

**Theorem 1.2.2.** If  $x \in \mathbb{Q}$  has 2 decimal expansions, then they will be of the form

$$x = a_0.a_1a_2 \dots a_n\overline{9} = a_0.a_1a_2 \dots (a_n + 1), a_n \in [0, 8].$$

**Definition 1.2.3** (Real numbers). The **real numbers**,  $\mathbb{R}$ , can be defined as:

$$\mathbb{R} := \{a_0.a_1a_2 \dots : a_0 \in \mathbb{Z}, a_i \in [0, 9], \nexists N \in \mathbb{N} \text{ such that } a_i = 9 \forall i \geq N\}.$$



### 1.3 Countability

**Definition 1.3.1** (Countability). A set  $S$  is **countably infinite** iff there exists a bijection  $f : \mathbb{N} \rightarrow S$ , a set is **countable** if it is finite or countable infinite.

**Theorem 1.3.2.** All  $S \subseteq \mathbb{N}$  are countable,  $\mathbb{Z}$  and  $\mathbb{Q}$  are both countable,  $\mathbb{R}$  is uncountable.

## 2 Bounded sets

### 2.1 Supremums and infimums

**Definition 2.1.1** (Maximum and minimum).  $s \in \mathbb{R}$  is the **maximum** of a set  $S \subset \mathbb{R}$  iff  $\forall s' \in S, s \geq s'$ . **Minimums** are defined similarly. Maximums and minimums are unique.

**Definition 2.1.2** (Bounded). A non-empty set  $S \subset \mathbb{R}$  is **bounded above** if there exists some  $M \in \mathbb{R}$  such that  $\forall s \in S, s \leq M$  with **bounded below** defined similarly.  $S$  is **bounded** if it is both bounded above and bounded below.

**Theorem 2.1.3.** If  $S$  is bounded then  $\exists R > 0$  such that  $|s| < R$  for all  $s \in S$ .

**Definition 2.1.4** (Supremum and infimum). If  $S \subset \mathbb{R}$  is bounded above, we say  $x \in \mathbb{R}$  is the **least upper bound** or **supremum** iff  $x$  is an upper bound for  $S$  and for all  $y \in \mathbb{R}$  such that  $y$  is an upper bound of  $S$ ,  $x \leq y$ . The **infimum** is defined similarly.

### 2.2 Completeness

**Theorem 2.2.1** (Completeness axiom). For all non-empty  $S \subset \mathbb{R}$ , if  $S$  is bounded above then  $S$  has a supremum, and similarly for  $S$  bounded below.

### 2.3 Dedekind cuts

**Definition 2.3.1** (Dedekind cut). A non-empty set  $S \subset \mathbb{Q}$  is a **Dedekind cut** if it satisfies:

1.  $s \in S$  and  $s > t \in \mathbb{Q} \implies t \in S$  ( $S$  is a semi-infinite interval to the left),
2.  $S$  is bounded above with no maximum.

Dedekind cuts are in the form  $S_r := (-\infty, r) \cap \mathbb{Q}$ .

**Theorem 2.3.2** (Real numbers). We can redefine the reals as the set of Dedekind cuts,  $\mathbb{R} := \{S_r \subset \mathbb{Q}\}$ . All operations and orderings as well as the completeness axiom are held by this new Dedekind cut definition.

**Theorem 2.3.3** (Triangle inequality). For all  $a, b \in \mathbb{R}$  we have  $|a + b| \leq |a| + |b|$ .

## 3 Sequences

**Definition 3.0.1** (Real sequence). A **real sequence** is a function  $a : \mathbb{N} \rightarrow \mathbb{R}$  written  $(a_n)$ . Sequences of other number systems are defined similarly.

### 3.1 Convergence

**Definition 3.1.1** (Convergence of sequences). A real sequence  $(a_n)$  **converges** to some  $a \in \mathbb{R}$  as  $n \rightarrow \infty$  iff

$$\forall \epsilon > 0, \exists N_\epsilon \text{ such that } \forall n \geq N_\epsilon, |a_n - a| < \epsilon.$$

For complex series the definition is the same just with  $|\cdot|$  referring to the modulus instead of the absolute value. This is written  $a_n \rightarrow a$  (as  $n \rightarrow \infty$ ).

### 3.2 Divergence

**Definition 3.2.1** (Divergence). A sequence  $(a_n)$  **diverges** iff it doesn't converge.

**Definition 3.2.2** (Divergence to infinity). A sequence  $(a_n)$  **diverges to  $\infty$**  iff  $\forall R > 0, \exists N \in \mathbb{N}$ , such that  $\forall n \geq N, a_n > R$ . And similarly for a sequence diverging to  $-\infty$ .

### 3.3 Limits

**Theorem 3.3.1** (Uniqueness of limits). Given a sequence  $(a_n)$  if  $a_n \rightarrow a$  and  $a_n \rightarrow b$ ,  $a = b$ .

**Theorem 3.3.2.** If a sequence  $(a_n)$  is convergent then  $(a_n)$  is bounded.

**Theorem 3.3.3** (Algebra of limits). Given two sequences  $a_n \rightarrow a$  and  $b_n \rightarrow b$  the following hold:

- $a_n + b_n \rightarrow a + b$ ,
- $a_n b_n \rightarrow ab$  (a special case of this is  $ca_n \rightarrow ca$  for a constant  $c$ ),
- $\frac{a_n}{b_n} \rightarrow \frac{a}{b}$  given  $b \neq 0$ .

**Theorem 3.3.4.** If  $(a_n)$  is a positive sequence then  $a_n \rightarrow 0 \iff \frac{1}{a_n} \rightarrow +\infty$ , and similarly for negative sequences.

**Theorem 3.3.5** (Ratio test). If a sequence  $(a_n)$  satisfies  $\left| \frac{a_{n+1}}{a_n} \right| \rightarrow L < 1$  then  $a_n \rightarrow 0$ .

### 3.4 Monotone sequences

**Definition 3.4.1** (Monotonically increasing sequence). A sequence,  $(a_n)$ , is **monotonically increasing** iff  $\forall m, n \in \mathbb{N}$  with  $n > m$  we have  $a_n \geq a_m$ , and similarly for monotonically decreasing and their strict equivalents.

**Theorem 3.4.2** (Monotone convergence). If a sequence  $(a_n)$  is monotone increasing and bounded above then  $a_n \rightarrow a := \sup\{a_i : i \in \mathbb{N}\}$  written  $a_n \uparrow a$ . This holds similarly for monotone decreasing sequences.

### 3.5 Cauchy sequences

**Definition 3.5.1** (Cauchy sequence). A sequence  $(a_n)$  is a **Cauchy sequence** iff  $\forall \epsilon > 0 \in \mathbb{R}, \exists N \in \mathbb{N}$  such that  $\forall n, m < N, |a_n - a_m| < \epsilon$ .

**Theorem 3.5.2** (Cauchy convergence criterion). A sequence  $(a_n)$  converges iff it is a Cauchy sequence.

### 3.6 Subsequences

**Definition 3.6.1** (Subsequence). Given a strictly monotonically increasing function  $n : \mathbb{N} \rightarrow \mathbb{N}$  and a sequence  $(a_n)$ , the sequence  $(b_n)$  defined by  $b_i := a_{n(i)}$  is a **subsequence** of  $(a_n)$ .

**Theorem 3.6.2.** Given a subsequence of  $(a_n)$ ,  $(a_{n(i)})$ , if  $a_n \rightarrow a$  then  $a_{n(i)} \rightarrow a$  as  $i \rightarrow \infty$ .

**Theorem 3.6.3** (Bolzano-Weierstrass). If a sequence  $(a_n)$  is bounded then it has a convergent subsequence.

**Note 3.6.4** (Sketch of the Bolzano-Weierstrass theorem proof). The proof of the Bolzano-Weierstrass theorem is an equally valuable point as the statement of the theorem itself. The idea of the proof considers the “peak points” of the sequence: if there are infinitely many peak points, then the peak points themselves form a monotonically decreasing subsequence; if there are finitely many, then the points after the final peak must have a monotonically increasing subsequence bounded above by the final peak. By the monotone convergence theorem both of these subsequences must converge.

## 4 Series

**Definition 4.0.1** (Infinite series). An **(infinite) series** is an expression of the form  $\sum_{i=1}^{\infty} a_i$  of  $a_1 + a_2 + \dots$  for some sequence  $(a_n)$ . The sequence **partial sums** of the series  $(S_n)$  is given by

$$S_n := \sum_{i=1}^n a_i = a_1 + a_2 + \dots + a_n.$$

## 4.1 Convergence

**Definition 4.1.1** (Convergence of series). The series  $\sum_{i=1}^{\infty} a_i$  of  $(a_n)$  **converges** iff  $S_n \rightarrow A \in \mathbb{R}$ , written  $\sum_{n=1}^{\infty} a_n = A$ .

**Theorem 4.1.2.** For a sequence  $(a_n)$ ,  $\sum_{n=1}^{\infty} a_n$  converges if  $a_n \rightarrow 0$  (the converse is not true).

**Theorem 4.1.3.** Given a sequence non-negative sequence  $(a_n)$ , the convergence of the infinite series and the boundedness of  $(S_n)$  are equivalent.

**Theorem 4.1.4** (Algebra of limits for series). A similar algebra of limits for series can be established from the algebra of limits for sequences acting on the partial sums of the series.

**Theorem 4.1.5** (Comparison I test). Given sequences  $(a_n), (b_n)$  if  $0 \leq a_n \leq b_n$  then:

- If  $\sum_{n=1}^{\infty} a_n$  and  $\sum_{n=1}^{\infty} b_n$  converge,  $0 \leq \sum_{n=1}^{\infty} a_n \leq \sum_{n=1}^{\infty} b_n$ ,
- If  $\sum_{n=1}^{\infty} a_n$  diverges,  $\sum_{n=1}^{\infty} b_n$  also diverges.

**Theorem 4.1.6** (Comparison II test (Sandwich theorem)). Given sequences  $(a_n), (b_n), (c_n)$  with  $a_n \leq b_n \leq c_n$ , if  $\sum_{n=1}^{\infty} a_n$  and  $\sum_{n=1}^{\infty} c_n$  both converge,  $\sum_{n=1}^{\infty} b_n$  converges.

**Theorem 4.1.7.** If  $\alpha > 1 \in \mathbb{R}$ ,  $\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$  converges.

**Definition 4.1.8** (Alternating sequence). A sequence  $(a_n)$  is **alternating** iff  $a_{2n} \geq 0$  and  $a_{2n-1} \leq 0$  of vice versa for all  $n \in \mathbb{N}^{>0}$ .

**Theorem 4.1.9.** If  $(a_n)$  is alternating with  $|a_n| \downarrow 0$ ,  $a_n$  converges and  $\sum_{n=1}^{\infty} a_n$  converges.

## 4.2 Absolute convergence

**Definition 4.2.1** (Absolute convergence). Given a sequence  $(a_n)$  the series  $\sum_{n=1}^{\infty} a_n$  is **absolutely convergent** iff  $\sum_{n=1}^{\infty} |a_n|$  converges.

**Theorem 4.2.2.** Absolute convergence  $\implies$  convergence.

**Theorem 4.2.3** (Comparison III test). Given sequences  $(a_n), (b_n)$  with  $\frac{a_n}{b_n} \rightarrow L \in \mathbb{R}$  if  $\sum_{n=1}^{\infty} b_n$  is absolutely convergent then  $\sum_{n=1}^{\infty} a_n$  is also absolutely convergent.

**Theorem 4.2.4** (Ratio test). If the sequence  $(a_n)$  is such that  $\left| \frac{a_{n+1}}{a_n} \right| \rightarrow r < 1$  then  $\sum_{n=1}^{\infty} a_n$  is absolutely convergent or divergent if  $r > 1$ .

**Theorem 4.2.5** (Root test). If the sequence  $(a_n)$  is such that  $|a_n|^{\frac{1}{n}} \rightarrow r < 1$  then  $\sum_{n=1}^{\infty} a_n$  is absolutely convergent or divergent is  $r > 1$ .

**Remark 4.2.6.** Both the ratio test and the root test are inconclusive if  $r = 1$ .

### 4.3 Rearrangement of series

Sometimes, series are easier to deal with and have cancellations when their terms are rearranged. However, the rearrangement of terms will only preserve limits under certain conditions.

**Definition 4.3.1** (Reordering). Given a bijection  $n : \mathbb{N} \rightarrow \mathbb{N}$  and a sequence  $(a_n)$ , the sequence  $(b_n)$  with  $b_i := a_{n(i)}$  is a **rearrangement** or **reordering** of  $(a_n)$ .

**Theorem 4.3.2.** If  $(a_n)$  is a sequence satisfying  $a_n \rightarrow 0$ ,  $\sum_{n:a_n \geq 0} a_n = \infty$  and  $\sum_{n:a_n \leq 0} a_n = -\infty$  then  $\sum_{n=1}^{\infty} a_n$  can be rearranged to converge to any  $r \in \mathbb{R}$ .

**Theorem 4.3.3.** If  $(a_n)$  is a sequence with absolutely convergent series,  $\sum_{n:a_n \geq 0} a_n = A$  and  $\sum_{n:a_n \leq 0} a_n = B$  with all arrangements of  $(a_n)$  converging to  $A + B$ .

### 4.4 Power series

Throughout this subsection  $[0, \infty] := [0, \infty) \cup \{+\infty\}$ .

**Definition 4.4.1** (Power series). For  $z \in \mathbb{C}$  and a complex sequence  $(a_n)$ , a **power series** is an expression in the form  $\sum_{n=1}^{\infty} a_n z^n$ .

**Definition 4.4.2** (Radius of convergence). Given the power series  $\sum_{n=1}^{\infty} a_n z^n$ , there exists some  $R \in [0, \infty]$  such that:

- $|z| < R \implies \sum_{n=1}^{\infty} a_n z^n$  converges,
- $|z| > R \implies \sum_{n=1}^{\infty} a_n z^n$  diverges.

We cannot tell what happens when  $|z| = R$  so this has to be checked separately.  $R$  is the **radius of convergence** of the power series.

**Corollary 4.4.3.** Given the same power series  $\sum_{n=1}^{\infty} a_n z^n$ , have  $S := \{|z| \in \mathbb{R}^{\geq 0} : a_n z^n \rightarrow 0\}$  then

$$R := \begin{cases} \sup(S) & \text{if } S \text{ is bounded} \\ \infty & \text{otherwise} \end{cases}.$$

is the radius of convergence for the power series.

**Theorem 4.4.4** (Evaluating radius of convergence from tests). For the power series  $\sum_{n=1}^{\infty} a_n z^n$ :

- if  $\left| \frac{a_{n+1}}{a_n} \right| \rightarrow a \in [0, \infty]$  then  $R = \frac{1}{a}$  is the radius of convergence for the power series,
- if  $|a_n|^{\frac{1}{n}} \rightarrow a \in [0, \infty]$  then  $R = \frac{1}{a}$  is the radius of convergence for the power series,

**Definition 4.4.5** (Cauchy product). Given two series  $\sum_{n=1}^{\infty} a_n$ ,  $\sum_{n=1}^{\infty} b_n$ ; their **Cauchy product** is the series

$$\sum_{n=0}^{\infty} \sum_{i=0}^n a_i b_{n-i}.$$

**Remark 4.4.6.** If  $(a_n)$ ,  $(b_n)$  are the coefficients for a power series, then the Cauchy product of their series will be the coefficients of the product of the power series.

**Theorem 4.4.7.** If  $\sum_{n=1}^{\infty} a_n$ ,  $\sum_{n=1}^{\infty} b_n$  are absolutely convergent their Cauchy product converges absolutely to  $\left(\sum_{n=1}^{\infty} a_n\right) \left(\sum_{n=1}^{\infty} b_n\right)$ .

**Theorem 4.4.8.** If the power series  $\sum_{n=1}^{\infty} a_n z^n$ ,  $\sum_{n=1}^{\infty} b_n z^n$  have radii of convergence  $R_a, R_b$  respectively then their Cauchy product has radius of convergence  $R_c \geq \min(R_a, R_b)$ .

## 4.5 Exponential series

**Definition 4.5.1.** For  $z \in \mathbb{C}$ , its **exponential series** is

$$E(z) := \sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \dots,$$

with  $E(z)$  converging absolutely for all  $z \in \mathbb{C}$ .

**Theorem 4.5.2** (Properties of exponential series). For all  $z, w \in \mathbb{C}$ :

1.  $E(z)E(w) = E(z+w)$ , 2.  $\frac{1}{E(z)} = E(-z)$ , 3.  $E(z) \neq 0$ .

**Theorem 4.5.3.** For all  $x \in \mathbb{Q}$ ,  $E(x) = e^x$ , with  $e := E(1)$ .

## 5 Continuity

### 5.1 Continuous functions

**Definition 5.1.1** (Limit of real functions). For a function  $f : \mathbb{R} \rightarrow \mathbb{R}$  and some  $a, b \in \mathbb{R}$  we have  $f(x) \rightarrow b$  as  $x \rightarrow a$  if  $\lim_{x \rightarrow a} f(x) = b$  iff:

$$\forall \epsilon > 0, \exists \delta > 0 \text{ such that } |x - a| < \delta \implies |f(x) - b| < \epsilon.$$

**Definition 5.1.2** (Continuity of real functions). Given the function  $f : \mathbb{R} \rightarrow \mathbb{R}$

1.  $f$  is **continuous at a point**  $a \in \mathbb{R}$  iff  $\lim_{x \rightarrow a} f(x) = f(a)$ ,
2.  $f$  is **continuous (on  $\mathbb{R}$ )** iff  $f$  is continuous at all  $a \in \mathbb{R}$ .

**Definition 5.1.3** (Discontinuity of real functions). The function  $f : \mathbb{R} \rightarrow \mathbb{R}$  is **discontinuous** at a point if it is not continuous at that point.

**Definition 5.1.4** (Sequential continuity). The function  $f : \mathbb{R} \rightarrow \mathbb{R}$  is continuous at  $a \in \mathbb{R}$   $\iff f(a_n) \rightarrow f(a)$  as  $n \rightarrow \infty$  for all sequences  $(a_n)$  converging to  $a$ .

**Remark 5.1.5.** The definition for limits and continuity of complex functions is similar with  $|\cdot|$  being the modulus instead of the absolute values. The same definition also applies for functions that are continuous on certain subsets of  $\mathbb{R}$  or  $\mathbb{C}$ .

**Theorem 5.1.6.**  $E : \mathbb{C} \rightarrow \mathbb{C}$  given by  $E(z) := \sum_{n=0}^{\infty} \frac{z^n}{n!}$  is continuous on  $\mathbb{C}$ .

**Theorem 5.1.7** (Properties of the real exponential function). Given the exponential function  $E : \mathbb{R} \rightarrow (0, \infty)$ :

1. for all  $x \in \mathbb{R}$ ,  $E(x) > 0$ ,
2.  $x > 0 \implies E(x) > 1$ ,
3.  $E(x)$  is a strictly increasing function,

4. For  $|x| < 1$ ,  $|E(x) - 1| \leq \frac{|x|}{1 - |x|}$ ,
5.  $E$  is a continuous bijection.

**Theorem 5.1.8.** The inverse of  $E(x) = e^x$  is the **natural logarithm** function  $\ln : (0, \infty) \rightarrow \mathbb{R}$  satisfying  $y = \ln x \iff x = e^y$  for all  $x, y \in \mathbb{R}$ .

**Definition 5.1.9** (Exponentiation of positive bases). For  $a \in (0, \infty)$ , for all  $x \in \mathbb{R}$  define  $a^x := E(x \ln a)$ .

**Definition 5.1.10** (Trigonometric functions). The **sine** and **cosine** functions are defined as:

$$\sin(\theta) := \Im[E(i\theta)], \quad \cos(\theta) := \Re[E(i\theta)].$$

and are both continuous functions from  $\mathbb{R} \rightarrow [-1, 1]$ .

**Theorem 5.1.11** (Continuity of piecewise functions). For  $a, c \in \mathbb{R}$  with functions  $f_1 : (-\infty, a) \rightarrow \mathbb{R}$  and  $f_2 : (a, \infty) \rightarrow \mathbb{R}$ , the **piecewise function**  $f : \mathbb{R} \rightarrow \mathbb{R}$ , defined as,

$$f(x) := \begin{cases} f_1(x) & \text{if } x < a \\ c & \text{if } x = a \\ f_2(x) & \text{if } x > a \end{cases}$$

is continuous on  $\mathbb{R}$  iff both  $f_1$  and  $f_2$  are continuous on their respective domains and

$$\lim_{x \uparrow a} f_1(x) = \lim_{x \downarrow a} f_2(x) = c.$$

## 5.2 Properties of continuity

**Theorem 5.2.1.** For  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  continuous at  $a \in \mathbb{R}$  the following functions are also continuous at  $a$ :

1.  $\alpha f$  for all  $\alpha \in \mathbb{R}$ ;
2.  $f + g, f \cdot g$ ;
3.  $\frac{f}{g}$ , given  $g(a) \neq 0$ .

**Theorem 5.2.2.** The following functions (all by their well known definitions) are continuous:

1.  $f(x) = x^n$ , for  $n \in \mathbb{N}_0$  (**monomials**);
2.  $p(x) = \sum_{i=1}^n a_i x^i$ , given  $(a_n)$  is a real sequence (**polynomials**);
3.  $\frac{p(x)}{q(x)}$  at  $a \in \mathbb{R}$  given  $p, q$  are polynomials with  $q(a) \neq 0$  (**rational functions**);
4.  $\sin(x)$ ,  $\cos(x)$  on  $\mathbb{R}$  and  $\tan(x)$  whenever  $\cos(x) \neq 0$ , plus their reciprocals under similar conditions;
5.  $f \circ g$  at  $a \in \mathbb{R}$  when  $g$  is continuous at  $a$  and  $f$  is continuous at  $g(a)$ .

**Theorem 5.2.3** (Intermediate value theorem). Given  $a, b \in \mathbb{R}$  with  $a \leq b$ , if  $f : [a, b] \rightarrow \mathbb{R}$  is continuous, then for all  $c$  between  $f(a)$  and  $f(b)$  there exists some  $x \in [a, b]$  such that  $f(x) = c$ .

**Definition 5.2.4** (Boundedness of real functions). Given some  $S \subseteq \mathbb{R}$  a function  $f : S \rightarrow \mathbb{R}$  is **bounded above** iff  $\exists M \in \mathbb{R}$  such that  $f(x) \leq M$  for all  $x \in S$ . The definitions for **bounded below** and **bounded** extend naturally from this.

**Theorem 5.2.5** (Extreme value theorem). Given  $a, b \in \mathbb{R}$  with  $a \leq b$ , if  $f : [a, b] \rightarrow \mathbb{R}$  is continuous then  $f$  is bounded.

## 6 Properties of subsets

### 6.1 Open sets

**Definition 6.1.1** (Open sets). A set  $S \subset \mathbb{R}$  is **open** iff  $\forall x \in S, \exists \delta$  such that  $(x - \delta, x + \delta) \subset S$ .

**Theorem 6.1.2** (Union of open sets). For a collection of open sets in  $\mathbb{R}$ ,  $\{S_i\}$ , given the indexing set  $\mathcal{I}$  (could be countable or uncountable),  $\bigcup_{i \in \mathcal{I}} S_i$  is open in  $\mathbb{R}$ .

**Theorem 6.1.3** (Finite intersections of open sets). The intersection of finitely many open sets in  $\mathbb{R}$  is open in  $\mathbb{R}$ .

## 6.2 Closed and compact sets

**Definition 6.2.1** (Closed sets). A set  $S \subset \mathbb{R}$  is **closed** in  $\mathbb{R}$  if all convergent subsequences of  $S$  have a limit in  $S$ .

**Definition 6.2.2** (Compact sets). A set  $S \subset \mathbb{R}$  is **compact** in  $\mathbb{R}$  if it is closed and bounded in  $\mathbb{R}$ .

**Theorem 6.2.3.** The complement of an open set is closed.

**Remark 6.2.4.** Not every set in  $\mathbb{R}$  is either open or closed. Half-open intervals are neither open nor closed while  $\mathbb{R}$  and  $\emptyset$  are both open and closed.

**Theorem 6.2.5.** The finite union or any intersection of closed sets in  $\mathbb{R}$  is closed.

**Theorem 6.2.6.** A set  $S \subset \mathbb{R}$  is compact iff every subsequence of  $S$  has as convergent subsequence  $x_{n(i)} \rightarrow x \in S$ .

**Theorem 6.2.7** (Extreme value theorem for compact sets). If  $S \subset \mathbb{R}$  is compact with  $f : S \rightarrow \mathbb{R}$  continuous, there exists some  $c, d \in S$  with  $f(c) = \inf_{x \in S} f(x)$  and  $f(d) = \sup_{x \in S} f(x)$ .

## 7 Uniform continuity and convergence

### 7.1 Uniform continuity

**Definition 7.1.1** (Uniform continuity). A function  $f : S \rightarrow \mathbb{R}$  is **uniformly continuous** iff

$$\forall \epsilon > 0, \exists \delta > 0 \text{ such that } \forall x, y \in S, |x - y| < \delta \implies |f(x) - f(y)| < \epsilon.$$

Uniform continuity is a more powerful notion than continuity with  $f$  is uniformly continuous  $\implies f$  is continuous.

**Theorem 7.1.2.** If  $S \subset \mathbb{R}$  is compact and  $f : S \rightarrow \mathbb{R}$  continuous then  $f$  is uniformly continuous.

### 7.2 Convergence of sequences of functions

**Definition 7.2.1** (Pointwise convergence). For some  $S \subset \mathbb{R}$  with the sequence  $f_1, f_2, \dots : S \rightarrow \mathbb{R}$ ,  $f_n$  **converges pointwise** to some  $f : S \rightarrow \mathbb{R}$  if

$$\forall x \in S, \forall \epsilon > 0, \exists N \in \mathbb{N} \text{ such that } \forall n \geq N, |f(x) - f_n(x)| < \epsilon.$$

Written  $\forall x \in S, \lim_{n \rightarrow \infty} f_n(x) = f(x)$ .

**Definition 7.2.2** (Uniform convergence). For some  $S \subset \mathbb{R}$ , the sequence  $f_1, f_2, \dots : S \rightarrow \mathbb{R}$  **uniformly converges** to some  $f : S \rightarrow \mathbb{R}$  if

$$\forall \epsilon > 0, \exists N \in \mathbb{N} \text{ such that } \forall x \in S, \text{ and } \forall n > N, |f(x) - f_n(x)| < \epsilon.$$

**Theorem 7.2.3.** If a sequence of (uniformly) continuous functions converges uniformly to a function  $f$  then  $f$  is (uniformly) continuous.

**Theorem 7.2.4.** If, given  $S \subset \mathbb{R}$ ,  $(f_n) : S \rightarrow \mathbb{R}$  is a uniformly convergent sequence of continuous functions with  $a \in S$  open in  $S$ ,  $\lim_{n \rightarrow \infty} \lim_{x \rightarrow a} f_n(x) = \lim_{x \rightarrow a} \lim_{n \rightarrow \infty} f_n(x)$ .

### 7.3 Convergence of series of functions

**Definition 7.3.1** (Convergence of series of functions). Given  $(f_n) : S \rightarrow \mathbb{R}$  defined on  $S \subset \mathbb{R}$ , the series  $\sum_{n=1}^{\infty} f_n(x)$  **converges (uniformly)** iff the sequence of partial sums  $S_n(x) = \sum_{n=1}^n f_n(x)$  converges (uniformly).

**Theorem 7.3.2** (Weierstrass M-test). Given continuous  $(f_n) : S \rightarrow \mathbb{R}$  defined on  $S \subset \mathbb{R}$ ,

$$\begin{aligned} &\forall x \in S \text{ and } \forall i \in \mathbb{N}, \exists M_1, M_2, \dots \in \mathbb{R} \text{ such that } |f_i(x)| \leq M_i \text{ and } \sum_{i=1}^{\infty} M_i \text{ converges} \\ &\implies \sum_{n=1}^{\infty} f_n(x) \text{ converges uniformly to some continuous } g : S \rightarrow \mathbb{R}. \end{aligned}$$

**Theorem 7.3.3.** If a power series  $f(x) = \sum_{n=1}^{\infty} f_n(x)$  has radius of convergence  $R > 0$  then  $f$  is continuous on  $(-R, R)$ .

## 8 Differentiation

### 8.1 Differentiability

**Definition 8.1.1** (Differentiability). A function  $f : \mathbb{R} \rightarrow \mathbb{R}$  is **differentiable** at  $a \in \mathbb{R}$ , with **derivative**

$$f'(a) = \left. \frac{d}{dx} f(x) \right|_a \text{ iff } \lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} \text{ exists, which we set to } f'(a).$$

$f$  is differentiable on  $S \subseteq \mathbb{R}$ , with derivative  $\frac{d}{dx} f = \frac{df}{dx} = f' : \mathbb{R} \rightarrow \mathbb{R}$ , if it is differentiable at every  $x \in S$ .

**Examples 8.1.2.** The following functions are all differentiable,

- $f(x) = x^n$ , for  $n \in \mathbb{N}$  on  $\mathbb{R}$  with  $f'(x) = nx^{n-1}$ ,
- $f(x) = e^x$  on  $\mathbb{R}$  with  $f'(x) = e^x$ ,
- $f(x) = \ln x$  on  $\mathbb{R}^{>0}$  with  $f'(x) = \frac{1}{x}$ .

**Theorem 8.1.3.**  $f$  is differentiable  $\implies f$  is continuous.

**Theorems 8.1.4** (Operations on derivatives). If  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  are both differentiable at  $x = a \in \mathbb{R}$  then,

1. for all  $c, d \in \mathbb{R}$ ,  $h(x) := c \cdot f(x) + d \cdot g(x)$  is differentiable at  $x = a$  with  $h'(a) = c \cdot f'(a) + d \cdot g'(a)$ ,
2.  $p(x) := f(x) \cdot g(x)$  is differentiable at  $x = a$  with  $p'(a) = f(a) \cdot g'(a) + f'(a) \cdot g(a)$ ,
3. if  $f(a) \neq 0$ ,  $q(x) := \frac{1}{f(x)}$  is differentiable at  $x = a$  with  $q'(a) = -\frac{f'(a)}{[f(a)]^2}$ ,
4. if  $g(a) \neq 0$   $r(x) := \frac{f(x)}{g(x)}$  is differentiable at  $x = a$  with  $r'(a) = \frac{f'(a) \cdot g(a) - f(a) \cdot g'(a)}{[g(a)]^2}$ .

**Theorem 8.1.5** (Chain rule). If  $g, f : \mathbb{R} \rightarrow \mathbb{R}$  are differentiable at  $x = a \in \mathbb{R}$  and  $x = g(a)$  respectively then  $s(x) := f \circ g(x)$  is differentiable at  $x = a$  with  $s'(a) = g'(a) \cdot f'(g(a))$ .

### 8.2 Local extrema and mean values

**Definition 8.2.1** (Local extrema). For a function  $f : S \rightarrow \mathbb{R}$ ,  $f$  has a **local minimum** at  $a \in \mathbb{R}$  iff  $\exists \delta > 0$  such that  $\forall y \in S$  with  $|y - a| < \delta$ ,  $f(y) \geq f(a)$ , and similarly for a **local maximum**.

**Theorem 8.2.2.** If  $f : [a, b] \rightarrow \mathbb{R}$  is differentiable on  $(a, b)$  and has a local maximum or minimum at  $c \in (a, b)$ ,  $f'(c) = 0$ .

**Theorem 8.2.3** (Rolle's theorem). If  $f : [a, b] \rightarrow \mathbb{R}$  is differentiable on  $(a, b)$  with  $f(a) = f(b)$ ,  $\exists c \in (a, b)$  such that  $f'(c) = 0$ .

**Theorem 8.2.4** (Mean value theorem). If  $f : [a, b] \rightarrow \mathbb{R}$  is differentiable on  $(a, b)$ ,  $\exists c \in (a, b)$  such that  $f'(c) = \frac{f(b) - f(a)}{b - a}$ .

**Theorem 8.2.5.** If  $f : [a, b] \rightarrow \mathbb{R}$  is differentiable on  $(a, b)$  with  $f'(x) \geq 0$  for all  $x \in (a, b)$  then  $f$  is monotone increasing. Similar holds for monotone/strictly increasing/decreasing or constant.

**Theorem 8.2.6** (Cauchy's MVT). A similar but slightly more general statement than the MVT: if  $f, g : [a, b] \rightarrow \mathbb{R}$  are differentiable on  $(a, b)$ ,  $\exists c \in (a, b)$  with  $(f(b) - f(a))g'(c) = (g(b) - g(a))f'(c)$ .

### 8.3 L'Hôpital's rule

**Theorem 8.3.1** (L'Hôpital's rule). Given  $f, g : [c, d] \rightarrow \mathbb{R}$  are differentiable on  $(c, d)$  except possibly at some  $a \in (c, d)$  with  $g'(x) \neq 0$  on  $(c, d) \setminus \{a\}$ :

$$\text{if } \lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} g(x) = 0 \text{ or } \infty \text{ and } \lim_{x \rightarrow a} \frac{f'(x)}{g'(x)} = L \text{ then } \lim_{x \rightarrow a} \frac{f(x)}{g(x)} = L.$$

This also applies when taking  $\lim_{x \rightarrow \infty}$ .



**Definition 8.3.2** (Higher derivatives). **Higher derivatives** of  $f : \mathbb{R} \rightarrow \mathbb{R}$  are defined inductively as

$$f^{(n)}(x) := \begin{cases} f(x) & \text{if } x = 0 \\ f^{(n-1)'}(x) & \text{otherwise} \end{cases}.$$

The existence of the  $n$ th derivative of  $f$  requires all lower order derivatives of  $f$  also exist and be differentiable.

**Theorem 8.3.3** (Second derivative test). For a second differentiable function  $f : \mathbb{R} \rightarrow \mathbb{R}$  with  $f'(a) = 0$  for some  $a \in \mathbb{R}$ ,

- $f''(a) > 0 \implies f$  has a local minimum at  $x = a$ ,
- $f''(a) < 0 \implies f$  has a local maximum at  $x = a$ ,
- the test is inconclusive if  $f''(a) = 0$ .

## 8.4 Taylor's theorem

**Definition 8.4.1** (Taylor polynomial of a function). Given  $f : [c, d] \rightarrow \mathbb{R}$  has an order  $n \in \mathbb{N}_0$  derivative at  $x = a \in (c, d)$ , the **Taylor polynomial** of order  $n$  at  $x = a$  is

$$P_n(x) := \sum_{i=0}^n \frac{f^{(i)}(a)}{i!} (x-a)^i = f(a) + \frac{f'(a)}{1!} (x-a) + \frac{f''(a)}{2!} (x-a)^2 + \dots + \frac{f^{(n)}(a)}{n!} (x-a)^n.$$

**Theorem 8.4.2** (Taylor's theorem). Given  $f : [c, d] \rightarrow \mathbb{R}$  has an order  $n+1$ , for some  $n \in \mathbb{N}_0$ , derivative for all  $x \in (c, d)$ . For  $a, b \in [c, d]$  with  $a \neq b$  there exists some  $t$  between  $a$  and  $b$  such that,

$$f(b) = P_n(b) + \frac{f^{(n+1)}(t)}{(n+1)!} (b-a)^{n+1}.$$

This is a further, massive generalisation of the MVT (the case when  $n = 0$ ).

**Definition 8.4.3** (Taylor series of a function). The **Taylor series**,  $P(x)$ , for a function  $f : \mathbb{R} \rightarrow \mathbb{R}$  at  $x = a$  exists if  $f^{(n)}(a)$  exists for all  $n \in \mathbb{N}$  and is given by

$$P(x) := \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x-a)^n = f(a) + \frac{f'(a)}{1!} (x-a) + \frac{f''(a)}{2!} (x-a)^2 + \dots$$

**Definition 8.4.4** (Analytic function). A function  $f : \mathbb{R} \rightarrow \mathbb{R}$  is **analytic** if it equals its Taylor series.

## 8.5 Convexity

**Definition 8.5.1** (Convexity of functions). A function  $f : [a, b] \rightarrow \mathbb{R}$  is **convex** iff

$$\forall c, t, d \in [a, b] \text{ with } c < t < d, f(c) + \frac{f(d) - f(c)}{d - c} (t - c) \geq f(t).$$

**Theorem 8.5.2.** Given the function  $f : [a, b] \rightarrow \mathbb{R}$  with  $f''(x)$  existing on  $(a, b)$ ,  $f$  is convex  $\iff f''(x)$  non-negative on  $(a, b)$ .

## 8.6 Exchange of limits and derivatives

**Theorem 8.6.1** (Criteria for exchange of limits and derivatives). Given  $(f_n)$  is a sequence of functions with  $f_n : [a, b] \rightarrow \mathbb{R}$  differentiable, if  $\lim_{n \rightarrow \infty} f_n(c)$  exists for some  $c \in [a, b]$  and  $(f'_n(x))$  converges uniformly on  $[a, b]$ :  $(f_n)$  converges uniformly to some differentiable  $f$  satisfying  $f'(x) = \lim_{n \rightarrow \infty} f'_n(x)$ .

**Theorem 8.6.2** (Derivatives of power series). Given a power series  $f(x) = \sum_{n=0}^{\infty} a_n x^n$  with radius of convergence  $R > 0$ ,  $f$  has a continuous derivative on  $(-R, R)$  with  $f'(x) = \sum_{n=0}^{\infty} n a_n x^{n-1}$ .

**Corollary 8.6.3.** Given a power series  $f(x) = \sum_{n=0}^{\infty} a_n x^n$  with radius of convergence  $R > 0$ , the Taylor series of  $f$  centered at  $x = 0$  is  $\sum_{n=0}^{\infty} a_n x^n$ .

## 8.7 Trigonometric properties

**Definition 8.7.1** ( $\pi$ ). Let  $S = \{y > 0 : \sin(y) = 0\}$ ,  $\pi := \inf S$ .

**Definition 8.7.2** (Periodic function). A function  $f : \mathbb{R} \rightarrow \mathbb{R}$  is  **$2L$ -periodic** iff  $f(x + 2L) = f(x)$  for all  $x \in \mathbb{R}$ .

**Theorem 8.7.3.**  $\sin$  and  $\cos$  satisfy the following important properties: 1.  $\sin(x)$  is odd, 2.  $\cos(x)$  is even, 3.  $\cos^2(x) + \sin^2(x) = 1$  for all  $x \in \mathbb{R}$ , 4.  $\sin$  and  $\cos$  are  $2\pi$ -periodic functions.

## 9 Integration

### 9.1 Partitions

**Definition 9.1.1** (Partition). A **partition**,  $P$ , of the interval  $[a, b] \subset \mathbb{R}$  is a finite collection of points  $x_0, x_1, \dots, x_n \in [a, b]$  such that  $a = x_0 < x_1 < \dots < x_n = b$ . A partition naturally splits the domain  $[a, b]$  into finitely many closed intervals.

**Definition 9.1.2** (Refinement). Given partitions  $Q, P$ ,  $Q$  is a **refinement** of  $P$ , written  $Q \prec P$ , iff every point of  $P$  is also in  $Q$ .

**Definition 9.1.3** (Common refinement). Given partitions  $P, Q$  the **common refinement** of  $P$  and  $Q$  is the partition  $R$  containing all points in  $P$  or  $Q$ .  $R \prec P$  and  $R \prec Q$ .

### 9.2 Darboux sums

**Definition 9.2.1** (Darboux sums). Given the bounded function  $f : [a, b] \rightarrow \mathbb{R}$  and the partition  $P = \{x_0, x_1, \dots, x_n\}$  of  $[a, b]$ , we will assign to each subintervals generated by  $P$ :

- a length,  $\Delta x_i := x_{i+1} - x_i$ ,
- an infimum,  $m_i := \inf_{x_i \leq t \leq x_{i+1}} f(t)$ ,
- a supremum,  $M_i := \sup_{x_i \leq t \leq x_{i+1}} f(t)$ .

Now define the **lower Darboux sum** and **upper Darboux sum** of  $f$  w.r.t.  $P$  as:

$$L(f, P) := \sum_{i=0}^{n-1} m_i \Delta x_i, \quad U(f, P) := \sum_{i=0}^{n-1} M_i \Delta x_i \quad \text{respectively.}$$

If  $f : [a, b] \rightarrow \mathbb{R}$  is continuous then  $L(f, P)$  and  $U(f, P)$  exist.  $L(f, P)$  is always less than or equal to  $U(f, P)$ .

**Theorem 9.2.2** (Boundedness of refined Darboux sums). If  $f : [a, b] \rightarrow \mathbb{R}$  is bounded with  $Q \prec P$  partitions of  $[a, b]$ ,  $L(f, P) \leq L(f, Q) \leq U(f, Q) \leq U(f, P)$ .

**Theorem 9.2.3.** Given some bounded  $f : [a, b] \rightarrow \mathbb{R}$ , the set  $\{L(f, P) : P \text{ is a partition of } [a, b]\}$  is bounded above by any upper Darboux sum on  $[a, b]$  w.r.t.  $f$ .

### 9.3 Darboux integral

**Definition 9.3.1** (Darboux integrals). Given a bounded function  $f : [a, b] \rightarrow \mathbb{R}$ , the **lower Darboux integral** and **upper Darboux integral** are:

$$\int_a^b f(x) dx := \sup_P L(f, P), \quad \overline{\int_a^b} f(x) dx := \inf_P U(f, P) \quad \text{respectively.}$$

**Definition 9.3.2** (Darboux integrability). If the upper and lower Darboux integral of a bounded function  $f : [a, b] \rightarrow \mathbb{R}$  are equal,  $f$  is **Darboux integrable** on  $[a, b]$  with

$$\int_a^b f(x) dx := \int_a^b f(x) dx = \overline{\int_a^b} f(x) dx.$$

We will now refer to Darboux integrable functions simply as **integrable**.

**Theorem 9.3.3.** A bounded function  $f : [a, b] \rightarrow \mathbb{R}$  is integrable iff  $\forall \epsilon > 0$  there exists a partition  $P$  with  $U(f, P) - L(f, P) < \epsilon$ . Furthermore, given a sequence of partitions  $(P_n)$  if  $\lim_{n \rightarrow \infty} (U(f, P_n) - L(f, P_n)) = 0$  then

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} (L(f, P_n)) = \lim_{n \rightarrow \infty} (U(f, P_n)).$$

**Remark 9.3.4.** For a bounded function  $f : [a, b] \rightarrow \mathbb{R}$ ,  $f$  is integrable if it is, continuous, differentiable, monotone, or discontinuous at finitely many points.

## 9.4 Properties of integration

**Theorem 9.4.1** (Monotonicity). If  $f, g : [a, b] \rightarrow \mathbb{R}$  are integrable with  $f(x) \leq g(x)$  for all  $x \in \mathbb{R}$ ,

$$(1) \quad \int_a^b f(x) dx \leq \int_a^b g(x) dx. \quad (2) \quad m \cdot (b - a) \leq \int_a^b f(x) dx \leq M \cdot (b - a).$$

**Theorem 9.4.2** (Boundedness). If  $f : [a, b] \rightarrow \mathbb{R}$  is integrable with  $m \leq f(x) \leq M$  for all  $x \in \mathbb{R}$ ,

**Theorem 9.4.3** (Linearity). If  $f, g : [a, b] \rightarrow \mathbb{R}$  are integrable, for all  $c, d \in \mathbb{R}$ ,

$$(3) \quad \int_a^b (cf(x) + dg(x)) dx = c \int_a^b f(x) dx + d \int_a^b g(x) dx. \quad (4) \quad \int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx.$$

**Theorem 9.4.4** (Integrability on subdomains).  $f : [a, b] \rightarrow \mathbb{R}$  is integrable iff  $\forall c \in [a, b]$ ,  $f$  is integrable on  $[a, c]$  and  $[c, b]$  with,

**Theorem 9.4.5** (Composition). If  $f : [a, b] \rightarrow [m, M] \subset \mathbb{R}$ ,  $g : [m, M] \rightarrow \mathbb{R}$  are integrable and continuous respectively,  $h(x) := g \circ f(x)$  is integrable on  $[a, b]$ .

**Theorem 9.4.6** (Triangle inequality). If  $f : [a, b] \rightarrow \mathbb{R}$  is integrable then  $|f|$  is integrable on  $[a, b]$  with,

$$(6) \quad \left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx. \quad (7) \quad \int_a^b f(x) dx = \int_a^b g(x) dx.$$

**Theorem 9.4.7** (Finite point differences). If  $f, g : [a, b] \rightarrow \mathbb{R}$  are integrable with  $f(x) = g(x)$  except at finitely many points,

**Theorem 9.4.8** (Products). If  $f, g : [a, b] \rightarrow \mathbb{R}$  are integrable then  $f \cdot g : [a, b] \rightarrow \mathbb{R}$  is integrable.

**Theorem 9.4.9** (Maxima and minima). If  $f, g : [a, b] \rightarrow \mathbb{R}$  are integrable then  $\max(f, g), \min(f, g) : [a, b] \rightarrow \mathbb{R}$  are integrable.

## 9.5 Fundamental theorems of calculus

**Theorem 9.5.1** (Fundamental theorem of calculus 1). Given continuous  $f : [a, b] \rightarrow \mathbb{R}$ , have  $F : [a, b] \rightarrow \mathbb{R}$  with  $F(x) := \int_a^x f(t) dt$ .  $F$  is continuous on  $[a, b]$  and differentiable on  $(a, b)$ .  $F'(x) = f(x)$  for all  $x \in [a, b]$ .

**Theorem 9.5.2** (Fundamental theorem of calculus 2). Given continuous  $f : [a, b] \rightarrow \mathbb{R}$  with continuous derivative on  $(a, b)$ ,  $\int_a^b f'(x) dx = f(b) - f(a)$ .

## 9.6 Methods of integration

**Theorem 9.6.1** (MVT). If  $f : [a, b] \rightarrow \mathbb{R}$  is continuous,  $\exists c \in [a, b]$  such that  $\int_a^b f(x) dx = f(c)(b - a)$ .

**Theorem 9.6.2** (Integration by parts). If  $f, g : [a, b] \rightarrow \mathbb{R}$  have continuous first derivatives,

$$(2) \quad \int_a^b f(x)g'(x) dx = \left[ f(x)g(x) \right]_a^b - \int_a^b f'(x)g(x) dx. \quad (3) \quad \int_{u(c)}^{u(d)} f(x) dx = \int_c^d f(u(x))u'(x) dx.$$

**Theorem 9.6.3** (Integration by substitution). Given continuous  $f : [a, b] \rightarrow \mathbb{R}$  if  $u : [a, b] \rightarrow [c, d]$  has a continuous derivative on  $(c, d)$ ,

## 9.7 Limits and integrals

**Theorem 9.7.1** (Exchanging limits and integrals). If  $f_n : [a, b] \rightarrow \mathbb{R}$  is a sequence of integrable functions converging uniformly to  $f : [a, b] \rightarrow \mathbb{R}$ , then  $f$  is integrable with,

$$\int_a^b f(x) \, dx = \int_a^b \lim_{n \rightarrow \infty} f_n(x) \, dx = \lim_{n \rightarrow \infty} \int_a^b f_n(x) \, dx.$$

**Theorem 9.7.2** (Power series integration). If the power series  $f(x) = \sum_{n=0}^{\infty} a_n x^n$  has radius of convergence  $R > 0$ ,  $f$  is integrable on all closed subintervals of  $(-R, R)$  with

$$\int_0^x f(t) \, dt = \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1} \text{ for all } x \in (-R, R).$$

## 9.8 Improper integrals

**Definition 9.8.1** (Improper integral). Given  $f : (a, b] \rightarrow \mathbb{R}$  integrable on all  $[c, b] \subset (a, b]$ , the **improper integral**,

$$\int_a^b f(x) \, dx = \lim_{c \downarrow a} \int_c^b f(x) \, dx,$$

if the limit exists, otherwise the integral **diverges**; and similarly for other non-closed intervals or those with  $\pm\infty$  as bounds.

**Remark 9.8.2.** When integrating over intervals with multiple undefined points, the integral is split into sums of multiple integrals each with single undefined points on their boundaries.

# Chapter 2

# Linear Algebra

Lectured by Dr Charlotte Kestner  
Typed by Yu Coughlin  
Autumn 2023 & Spring 2024

## Introduction

The following are supplementary reading:

- S Lang, Linear algebra, 1987
- G Strang, Introduction to linear algebra, 2023
- S Axler, Linear Algebra Done Right, 2015

Lecture 1  
Thursday  
10/01/19

# 1 Linear Systems and matrices

## 1.1 Linear systems

**Definition 1.1.1** (Linear system). A **linear system** is a set of linear equations in the same variables.

**Notation 1.1.2.** The follow are all equivalent notation for the same linear system:

$$\begin{aligned}
 & \begin{aligned}
 a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\
 a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\
 \vdots & \\
 a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m
 \end{aligned}
 \iff
 \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}
 \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}
 =
 \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \\
 & \iff
 \left( \begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array} \right).
 \end{aligned}$$

## 1.2 Matrix algebra

**Definition 1.2.1** (Matrix by elements). An  $m \times n$  matrix written as  $A = [a_{ij}]_{m \times n}$  has the element  $a_{ij}$  in the  $i$ th row and  $j$ th column.

**Definition 1.2.2** (Matrix addition). If  $A = [a_{ij}]_{m \times n}$  and  $B = [b_{ij}]_{m \times n}$  then  $A + B := [a_{ij} + b_{ij}]_{m \times n}$ .

**Definition 1.2.3** (Scalar multiplication). If  $A = [a_{ij}]_{m \times n}$  then  $\lambda A := [\lambda a_{ij}]_{m \times n}$ .

**Definition 1.2.4** (Matrix multiplication). If  $A = [a_{ij}]_{p \times q}$  and  $B = [b_{ij}]_{q \times r}$  then  $AB := C = [c_{ij}]_{p \times r}$  where  $c_{ij} = \sum_{k=1}^q a_{ik}b_{kj}$ .

**Theorem 1.2.5.** Matrix multiplication is associative.

**Remark 1.2.6.** Matrix multiplication is not commutative.

## 1.3 EROs

**Definition 1.3.1** (Elementary row operations). The three **elementary row operations (EROs)** that can be performed on augmented matrixes are as follows:

1. Multiply a row by a non-zero scalar.
2. Swap two rows.
3. Add a scalar multiple of a row to another row.

**Remark 1.3.2.** EROs preserve the set of solutions of a linear system. Each ERO has an inverse.

**Definition 1.3.3** (Equivalence of linear systems). Two systems of linear equations are equivalent iff either:

1. They are both inconsistent.
2. (wlog) The augmented matrix of the first system can be transformed to the augmented matrix of the second system with just EROs.

**Definition 1.3.4** (Row echelon form / Echelon form / REF). A matrix is in **row echelon form** if it satisfies the following:

1. All of the zero rows are at the bottom of the matrix,
2. The first non-zero entry in any row is **1**,
3. The first non-zero entry in row  $i$  is strictly to the left of the first non-zero entry in row  $i + 1$ .

**Definition 1.3.5** (Reduced row echelon form / Row reduced echelon form / rREF). A matrix is in **reduced row echelon form** if it is in REF and the first non-zero entry in a row is the only non-zero entry in its column.

## 1.4 Matrices of note

**Definition 1.4.1** (Square matrix). A matrix is **square** iff it has the same number of rows and columns.

**Definition 1.4.2.** A square matrix  $(A = [a_{ij}]_{n \times n})$  is: 1. **Upper triangular** iff  $i > j \implies a_{ij} = 0$ . 2. **Lower triangular** iff  $i < j \implies a_{ij} = 0$ . 3. **Diagonal** iff  $i \neq j \implies a_{ij} = 0$ .

**Definition 1.4.3** (Identity matrix). The **identity matrix** of size  $n$  written  $I_n$ , is the square diagonal matrix of size  $n$  with all diagonal entries equal 1.

**Definition 1.4.4** (Elementary matrix). An **elementary matrix** is a matrix that can be achieved by applying a single ERO to the identity matrix.

**Definition 1.4.5** (Inverse). For a square matrix  $B$  if there exists a matrix  $B^{-1}$  such that  $BB^{-1} = I = B^{-1}B$  then  $B^{-1}$  is the **inverse** of  $B$  and vice versa.

**Definition 1.4.6** (Singular). A matrix without an inverse is **singular**.

**Theorem 1.4.7.** The inverse of a matrix is unique.

**Definition 1.4.8.** A **transpose** of the matrix  $A = [a_{ij}]_{m \times n}$  is  $A^T := [a_{ji}]_{n \times m}$ .

**Theorem 1.4.9.** If the matrix  $A$  has an inverse then its transpose has an inverse with  $(A^T)^{-1} = (A^{-1})^T$ .

**Theorem 1.4.10.** If a matrix  $A \in M_{m \times n}$  can be reduced to  $I_n$  by a sequence of EROs then  $A$  is invertible with  $A^{-1}$  given by applying the same sequence of EROs to  $I_n$ .

**Definition 1.4.11.** A matrix  $A$  is **orthogonal** if it has an inverse with  $A^{-1} = A^T$ .

**Theorem 1.4.12.** An orthogonal matrix  $A$  satisfies the condition  $(Ax) \cdot (Ay) = x \cdot y$ , where  $\cdot$  is the dot product.

## 2 Vector Spaces

The notion of a vector space is a structure designed to generalise that of real vectors, so before developing them we must first produce a generalisation of the real numbers.

### 2.1 Fields

**Definition 2.1.1** (Field). A **field** is a set  $\mathbb{F}$  equipped with the binary operations **addition**  $+: \mathbb{F} \times \mathbb{F} \rightarrow \mathbb{F}$  and **multiplication**  $\cdot: \mathbb{F} \times \mathbb{F} \rightarrow \mathbb{F}$  satisfying the follow axioms:

- A1  $\forall x, y \in \mathbb{F} : x + y = y + x$  (commutativity of addition),
- A2  $\forall x, y, z \in \mathbb{F} : x + (y + z) = (x + y) + z$  (associativity of addition),
- A3  $\exists 0_{\mathbb{F}} \in \mathbb{F}$  such that  $\forall x \in \mathbb{F} : x + 0_{\mathbb{F}} = x$ , (additive identity element),
- A4  $\forall x \in \mathbb{F}, \exists (-x) \in \mathbb{F}$  such that  $x + (-x) = 0_{\mathbb{F}}$ , (additive inverse);
- M1  $\forall x, y \in \mathbb{F} : x \cdot y = y \cdot x$  (commutativity of multiplication),
- M2  $\forall x, y, z \in \mathbb{F} : x \cdot (y \cdot z) = (x \cdot y) \cdot z$  (associativity of multiplication),
- M3  $\exists 1_{\mathbb{F}} \in \mathbb{F}$  such that  $\forall x \in \mathbb{F} : x \cdot 1_{\mathbb{F}} = x$ , (multiplicative identity element),
- M4  $\forall x \in \mathbb{F} \setminus \{0_{\mathbb{F}}\}, \exists x^{-1} \in \mathbb{F}$  such that  $x \cdot x^{-1} = 1_{\mathbb{F}}$ , (multiplicative inverse);
- D  $\forall x, y, z \in \mathbb{F} : x \cdot (y + z) = x \cdot y + x \cdot z$  (distributivity of multiplication over addition).

The field  $(\mathbb{F}, +, \cdot)$  is often referred to as just  $\mathbb{F}$ .

**Example 2.1.2.** The familiar sets  $\mathbb{R}, \mathbb{Q}, \mathbb{C}$  are all fields.

**Theorem 2.1.3.** If  $p \in \mathbb{N}$  is prime with  $\mathbb{F}_p = \{0, 1, \dots, p-1\}$  then  $(\mathbb{F}_p, +_p, \cdot_p)$  is a field.

## 2.2 Vector spaces

**Definition 2.2.1** (Vector space). A **vector space** over a field  $\mathbb{F}$  is a set  $V$  equipped with the binary operations **vector addition**  $\oplus : V \times V \rightarrow V$  and **scalar multiplication**  $\odot : \mathbb{F} \times V \rightarrow V$  satisfying the follow axioms:

- A1  $\forall u, v, w \in V : u \oplus (v \oplus w) = (u \oplus v) \oplus w$  (associativity of addition),
- A2  $\forall u, v \in V : u \oplus v = v \oplus u$  (commutativity of vector addition),
- A3  $\exists 0_V \in V$  such that  $\forall v \in V : v \oplus 0_V = v$ , (vector additive identity element),
- A4  $\forall v \in V, \exists (-v) \in V$  such that  $v \oplus (-v) = 0_V$ , (vector additive inverse),
- A5  $\forall x \in \mathbb{F}, \forall u, v \in V : x \odot (u \oplus v) = (x \odot u) \oplus (x \odot v)$  (vector distributivity 1),
- A6  $\forall x, y \in \mathbb{F}, \forall v \in V : x \cdot (y \odot v) = (x \cdot y) \odot v$  (vector distributivity 2),
- A7  $\forall x, y \in \mathbb{F}, \forall v \in V : (x \cdot y) \odot v = x \odot (y \odot v)$  (scalar multiplication associativity),
- A8  $\forall v \in V : 1_{\mathbb{F}} \odot v = v$ , (scalar multiplication identity element).

If  $V$  is a vector space over  $\mathbb{F}$  we say  $V$  is an  $\mathbb{F}$ -vector space with  $v \in V$  a **vector** and  $x \in \mathbb{F}$  a **scalar**.

## 2.3 Subspaces

**Definition 2.3.1** (Subspace). A subset  $W \subseteq V$  is a **subspace** of  $V$ , denoted  $W \leq V$  iff:

- S1  $W \neq \emptyset$ ,
- S2  $\forall w_1, w_2 \in W : w_1 \oplus w_2 \in W$ ,
- S3  $\forall x \in \mathbb{F}, \forall w \in W : x \odot w \in W$ .

If  $W = \{0_V\}$  then  $W$  is the **trivial subspace**.

**Theorem 2.3.2.** Every subspace of  $V$  contains  $0_V$ .

**Theorem 2.3.3.** If  $U$  and  $W$  are subspaces of  $V$ ,  $U \cap W$  is a subspace of  $V$ .

## 3 Spanning and Linear Independence

Throughout this section, assume  $V$  is an  $\mathbb{F}$ -vector space.

### 3.1 Spanning

**Definition 3.1.1** (Span). Given some set  $\{v_1, v_2, \dots, v_n\} \subseteq V$  define the **span** by,

$$\text{Span}(\{v_1, v_2, \dots, v_n\}) := \{u \in V : u = \sum_{i=1}^n \alpha_i v_i \text{ with } \alpha_i \in \mathbb{F}\}.$$

Note that the span of a subset of  $V$  is always a subspace of  $V$ .

**Remark 3.1.2.** If  $S \subseteq V$  is infinite,  $\text{Span}(S)$  is the set of all **finite** linear combinations of elements of  $S$ .

**Definition 3.1.3** (Spanning sets). If  $S \subseteq V$  and  $\text{Span}(S) = V$ , we say  $S$  **spans**  $V$  or  $S$  is a **spanning set** for  $V$ .

### 3.2 Linear independence

**Definition 3.2.1.** The set  $\{v_1, v_2, \dots, v_n\} \subseteq V$  is **linearly independent** in  $V$  iff:

$$\sum_{i=1}^n \alpha_i v_i = 0_V \iff \alpha_i = 0_{\mathbb{F}} \text{ for all } i \in [1, n].$$

**Theorem 3.2.2.** If  $S = \{v_1, v_2, \dots, v_n\} \subseteq V$  is linearly independent in  $V$  with  $v_{n+1} \in V \setminus \text{Span}(S)$  then  $S \cup \{v_{n+1}\}$  is also linearly independent in  $V$ .



## 4 Bases

### 4.1 Definition

Again, assume  $V$  is an  $\mathbb{F}$ -vector space throughout this section.

**Definition 4.1.1** (Bases). A **basis** for  $V$  is linearly independent, spanning set of  $V$ . If  $V$  has a finite bases then  $V$  is said to be a **finite dimensional** vector space.

**Theorem 4.1.2.** Any  $S \subseteq V$  is a basis for  $V$  iff every vector in  $V$  can be uniquely expressed as a linear combination of the elements of  $S$ .

**Theorem 4.1.3.** If  $V$  is non-trivial and  $S$  is a finite spanning set of  $V$  then  $S$  contains a basis for  $V$ .

**Lemma 4.1.4** (Steinitz Echange Lemma). Given some  $X \subseteq V$  with  $u \in \text{Span}(X)$  but  $u \notin \text{Span}(X \setminus \{v\})$  for some  $v \in X$ , let  $Y = (X \setminus \{v\}) \cup \{u\}$  then  $\text{Span}(X) = \text{Span}(Y)$ .

**Theorem 4.1.5.** Given a LI  $S \subseteq V$  and spanning set  $T \subseteq V$ ,  $|S| \leq |T|$ .

**Corollary 4.1.6.** If  $S$  and  $T$  are both bases for  $V$ ,  $|S| = |T|$ .

### 4.2 Dimension

**Definition 4.2.1** (Dimension of a vector space). If  $V$  is finite dimensional then the **dimension** of  $V$ ,  $\dim V$ , is the size of any basis of  $V$ .

**Definition 4.2.2** (Notable subspaces). Let  $U$  and  $W$  both be subspaces of  $V$ , the **intersection** of  $U$  and  $W$ :

$$U \cap W := \{v \in V : v \in U \text{ and } v \in W\}$$

is a subspace of  $V$ , and the **sum** of  $U$  and  $W$ :

$$U + W := \{u + w : u \in U, w \in W\}$$

is also a subspace of  $V$ .

**Theorem 4.2.3.** Let  $U$  and  $W$  both be subspaces of  $V$ , we have:

$$\dim(U + W) = \dim U + \dim W - \dim(U \cap W).$$

## 5 Matrix rank

**Definition 5.0.1.** Given a field  $\mathbb{F}$  and a matrix  $A \in M_{m \times n}(\mathbb{F})$  we have:

- the **row space** of  $A$ ,  $\text{RSp}(A)$ , as the span of the rows of  $A$ , this is a subspace of  $\mathbb{F}^n$ ,
- the **row rank** of  $A$ , is  $\dim(\text{RSp}(A))$ ,
- the **column space** of  $A$ ,  $\text{CSp}(A)$ , as the span of the columns of  $A$ , this is a subspace of  $\mathbb{F}^m$ ,
- the **column rank** of  $A$ , is  $\dim(\text{CSp}(A))$ .

**Theorem 5.0.2.** For any matrix  $A$ , the row rank of  $A$  is equal to the column rank of  $A$ .

**Definition 5.0.3** (Rank of a matrix). The **rank** of a matrix  $A$ ,  $\text{rank}(A)$ , is equal to the row/column rank of  $A$ .

**Theorem 5.0.4.** Given a field  $\mathbb{F}$  and a matrix  $A \in M_{n \times n}(\mathbb{F})$  with  $\text{rank}(A) = n$ :

- the rows of  $A$  for a basis for  $\mathbb{F}^n$ ,
- the columns of  $A$  for a basis for  $\mathbb{F}^n$ ,
- $A$  is invertible.

## 6 Linear transformations

### 6.1 Definition

**Definition 6.1.1** (Linear transformation). Given  $\mathbb{F}$ -vector spaces  $V$  and  $W$ , let  $T : V \rightarrow W$  be a function,  $T$  is a **linear transformation** iff the following two properties hold:

1.  $T$  **preserves vector addition**:  $\forall v_1, v_2 \in V$  we have  $T(v_1 +_V v_2) = T(v_1) +_W T(v_2)$ ,
2.  $T$  **preserves scalar multiplication**:  $\forall v \in V$  and  $\forall \lambda \in \mathbb{F}$  we have  $\lambda T(v) = T(\lambda v)$ .

**Definition 6.1.2** (Identity transformation). The **identity transformation** of the vector space  $V$  is the linear transformation  $\text{Id}_V : V \rightarrow V$  with  $\text{Id}_V(v) := v$  for all  $v \in V$ .

**Definition 6.1.3** (Linear transformation of a matrix). If  $A \in M_{m \times n}(\mathbb{F})$  then we can define  $T : \mathbb{F}^n \rightarrow \mathbb{F}^m$  by  $T(v) := Av$ ,  $T$  is a linear transformation.

**Theorem 6.1.4.** If  $V$  and  $W$  are  $\mathbb{F}$ -vector spaces,  $T(0_V) = 0_W$  and

$$v = \lambda_1 v_1 + \dots + \lambda_n v_n \iff T(v) = \lambda_1 T(v_1) + \dots + \lambda_n T(v_n).$$

**Theorem 6.1.5.** If  $V$  and  $W$  are  $\mathbb{F}$ -vector spaces,  $\text{Hom}(V, W)$  is the set of linear transformations from  $V$  to  $W$ , with pointwise addition and scalar multiplication  $\text{Hom}(V, W)$  is a  $\mathbb{F}$ -vector space.

### 6.2 Image and kernel

Throughout, assume  $T : V \rightarrow W$  is a linear transformation and  $V, W$  are  $\mathbb{F}$ -vector spaces

**Definition 6.2.1** (Image). We define the **image** of  $T$ , denoted  $\text{Im } T$ , as

$$\text{Im } T := \{w \in W : \exists v \in V, T(v) = w\},$$

with  $\text{Im } T$  being a subspace of  $W$ .

**Definition 6.2.2** (Kernel). We define the **kernel** of  $T$ , denoted  $\ker T$ , as

$$\ker T := \{v \in V : T(v) = 0_W\},$$

with  $\ker T$  being a subspace of  $V$ .

**Theorem 6.2.3.** If  $v_1, v_2 \in V$  then  $T(v_1) = T(v_2) \iff v_1 - v_2 \in \ker T$ .

**Theorem 6.2.4.** If  $\{v_1, v_2, \dots, v_n\}$  is a basis for  $V$ , then  $\text{Im } T = \text{Span}(\{T(v_1), T(v_2), \dots, T(v_n)\})$ .

**Remark 6.2.5.** If  $T$  is the linear transformation for some matrix  $A \in M_{m \times n}(\mathbb{F})$  then,  $\ker T$  is the set of solutions for  $Av = 0$ ,  $\text{Im } T$  is the column space of  $A$ , and  $\dim(\text{Im } T) = \text{rank } A$

### 6.3 Rank nullity

**Theorem 6.3.1** (Rank Nullity Theorem). If  $V$  and  $W$  are finite dimensional  $\mathbb{F}$ -vector spaces and  $T : V \rightarrow W$  is a linear transformation, we have:

$$\dim(\text{Im } T) + \dim(\ker T) = \dim V.$$

## 7 Representations

### 7.1 Matrices of transformations

Throughout this subsection let  $V$  be an  $n$ -dimensional  $\mathbb{F}$ -vector space and  $B = \{e_1, e_2, \dots, e_n\}$  be a basis for  $V$ .

**Definition 7.1.1** (Representation of a vector). Given some  $v \in V$  with  $v = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n$  for  $\lambda_i \in \mathbb{F}$ , we define the  $v$  **with respect to** (w.r.t.)  $B$  as

$$[v]_B := \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} \in \mathbb{F}^n.$$

**Remark 7.1.2.** This must be well defined as all vectors have a unique representation in terms of every basis.

**Definition 7.1.3** (Linear isomorphism). A **linear isomorphism** is a bijective linear transformation.

**Theorem 7.1.4.** The linear transformation  $T : V \rightarrow \mathbb{F}^n$  given by  $T(v) := [v]_B$  is a linear isomorphism.

## 7.2 Matrices of transformations

**Definition 7.2.1** (Representation of a linear transformation). Given finite dimensional  $\mathbb{F}$ -vector spaces  $V$  and  $W$  with bases  $B = \{v_1, v_2, \dots, v_n\}$ ,  $C = \{w_1, w_2, \dots, w_n\}$  respectively, the **matrix of  $T$  w.r.t.  $B$  and  $C$**  denoted  ${}_C[T]_B$  is  $m \times n$  matrix with the  $i$ th column given by  $[T(v_i)]_C$ .  ${}_B[T]_B$  is often shortened to  $[T]_B$ .

**Remark 7.2.2.**  ${}_C[T]_B[v]_B = [T(v)]_C$ , for all  $v \in V$ .

**Theorem 7.2.3.** Given a finite dimensional  $\mathbb{F}$ -vector space  $V$  with bases  $B = \{v_1, v_2, \dots, v_n\}$  and  $C = \{w_1, w_2, \dots, w_n\}$ , if  $v_i = \lambda_{1i} w_1 + \lambda_{2i} w_2 + \cdots + \lambda_{ni} w_n$  and  $P$  is the matrix given by  $P = [\lambda_{ij}]_{n \times n}$ , we have:

- $P = [X]_C$  where  $X$  is the unique linear transformation given by  $X(w_i) = v_i$  for all  $i \in [1, n]$ ,
- $P([v]_B) = [v]_C$  for all  $v \in V$ ,
- $P = {}_C[\text{Id}_V]_B$ .

$P$  is often also called the **change of basis matrix** from  $B$  to  $C$ .

**Corollary 7.2.4.**  $P$  is invertible with  $(P)^{-1} = ({}_C[\text{Id}_V]_B)^{-1} = {}_B[\text{Id}_V]_C$ .

**Theorem 7.2.5.** If  $T : V \rightarrow V$  is a linear transformation  $[T]_C = ({}_C[\text{Id}_V]_B)[T]_B({}_B[\text{Id}_V]_C)$ .

## 8 Determinants

### 8.1 Definition

**Definition 8.1.1** (Minor of matrix). Given a matrix  $A \in M(\mathbb{F})_n$  the  **$ij$ th-minor** of the matrix  $A$ ,  $A_{ij} \in M(\mathbb{F})_n$ , is  $A$  with row  $i$  and column  $j$  removed.

**Definition 8.1.2** (Determinant). The **determinant** of the matrix  $A$  is defined recursively by

$$\det(A) := \begin{cases} a_{11} & \text{if } A \text{ is a matrix with a single row and column} \\ \sum_{j=1}^n (-1)^{j+1} a_{1j} \det(A_{1j}) & \text{otherwise.} \end{cases}.$$

The determinant is only a function on square matrices.

**Theorem 8.1.3.** If a matrix  $A$  is singular,  $\det(A) = 0$ .

**Theorem 8.1.4.** If  $A$  is invertible then the columns of  $A$  are LI.

## 8.2 Properties

**Definition 8.2.1** (EROs). The three ERO's on the matrix  $A$  to form  $A'$  have the following effects on the determinant:

- multiplying a row by  $\lambda \neq 0$ ,  $\det(A') = \lambda \det(A)$ ;
- swapping two rows,  $\det(A') = -\det(A)$ ;
- adding a scalar multiple of one row to another,  $\det(A') = \det(A)$ .

**Definition 8.2.2** (Other miscellaneous properties). For obvious types:

- If  $A$ ,  $B$  and  $C$  all only differ in the  $i$ th row with the  $i$ th row of  $C$  being the sum of the  $i$ th row of  $A$  and  $B$ ,  $\det(C) = \det(A) + \det(B)$ ,
- if a matrix  $A$  has two identical rows,  $\det(A) = 0$ ,
- $\det(AB) = \det(A)\det(B)$ ,
- $\det(A^T) = \det(A)$ ,
- $\det(I_n) = 1$ .

**Definition 8.2.3** (Cofactor). The  $ij$ th cofactor of a matrix  $A$  is defined as,

$$c_{ij} := (-1)^{i+j} \det(A_{ij}).$$

The **matrix of cofactors** of  $A$  is defined as  $C = [c_{ij}]_{n \times n}$  where  $c_{ij}$  is the  $ij$ th cofactor of  $A$ .

**Theorem 8.2.4.** For a matrix  $A$  with matrix of cofactors  $C$ ,  $C^T A = \det(A) I_n$ .

**Theorem 8.2.5** (Cramer's Rule). ugh

**Definition 8.2.6.** The **determinant** of a linear transformation  $T : V \rightarrow V$  where  $B$  is a basis for  $T$ ,  $\det(T) = \det([T]_B)$ . This definition says, rather importantly, that the determinant of the matrix of linear transformation is independent of the basis that linear transformation is represented in.

## 9 Eigen-things

The prefix “eigen” comes from the German word “eigen” which can be roughly translated to mean “proper” or “characteristic”.

### 9.1 Eigenvectors and eigenvalues

**Definition 9.1.1** (Eigenvectors and eigenvalues). Given the finite dimensional  $\mathbb{F}$ -vector space,  $V$ , and the linear transformation  $T : V \rightarrow V$ , we say  $v \in V \setminus \{0_V\}$  is an **eigenvector** of  $T$  if it satisfies the equation  $T(v) = \lambda v$  for some  $\lambda \in \mathbb{F}$ , we call  $\lambda$  the corresponding **eigenvalue**.

**Definition 9.1.2** (Eigenspace). The **eigenspace** of an eigenvalue of a given linear transformation  $T : V \rightarrow V$  is the set of eigenvectors that correspond to said eigenvalue. The eigenspace of any eigenvalue  $\lambda$  of  $T$  is a subspace of  $V$ .

**Remark 9.1.3.** The eigenvectors, eigenvalues and eigenspaces of a matrix are defined obviously and do not depend on which basis the linear transformation is respresented in.

### 9.2 Characteristic polynomial

**Theorem 9.2.1.** If  $D$  is a square diagonal matrix,  $D^k$  is  $D$  with its entries raised to the power of  $k$ .

**Definition 9.2.2** (Characteristic polynomial). Given the finite dimensional  $\mathbb{F}$ -vector space  $V$  with basis  $B$  and the linear transformation  $T : V \rightarrow V$ , we define the **characteristic polynomial** of  $T$ ,  $\chi_T : \mathbb{F} \rightarrow \mathbb{F}$  by  $\chi_T(\lambda) := \det(\lambda I_n - T_B)$ .

**Theorem 9.2.3.** The characteristic polynomial of a linear transformation is independent of the basis it is represented in.

**Remark 9.2.4.** Therefore, the characteristic polynomial of a matrix can be defined as the characteristic polynomial of the linear transformation it represents.

### 9.3 Diagonalisation

**Definition 9.3.1** (Diagonalisability). Given a finite dimensional  $\mathbb{F}$ -vector space  $V$ , a linear transformation  $T : V \rightarrow V$  is **diagonalisable** if there exists a basis for  $V$  consisting of eigenvectors of  $T$ . Similarly the matrix  $A \in M_n(\mathbb{F})$  is **diagonalisable** if there exists a basis to  $\mathbb{F}^n$  as eigenvectors of  $A$ .

**Theorem 9.3.2.** If  $V$  is a  $n$ -dimensional vector space and  $T : V \rightarrow V$  has  $n$  distinct eigenvalues,  $T$  is diagonalisable.

**Theorem 9.3.3.** If a matrix  $A \in M_n(\mathbb{F})$  is diagonalisable, let  $P$  be the matrix with columns as eigenvectors of  $A$  and  $D$  be the diagonal matrix with  $i$ th entry as the corresponding eigenvalue for the  $i$ th column of  $P$ , then  $A = PDP^{-1}$ .

**Theorem 9.3.4** (Precursor of Cayley-Hamilton). If a matrix  $A \in M_n(\mathbb{F})$  is diagonalisable,  $\chi_A(A) = 0$  where  $0$  is the zero-matrix in  $M_n(\mathbb{F})$ .

**Note 9.3.5.** The full **Cayley-Hamilton theorem**, which applies to all  $A \in M_n(\mathbb{F})$ , requires more advanced reasoning, this simpler version only requires the expansion of the polynomial and diagonalisation.

## 10 Orthogonality

### 10.1 Inner product spaces\*

**Definition 10.1.1** (Inner product). Let  $V$  be an  $n$ -dimensional  $\mathbb{F}$ -vector space, a **inner product** on  $V$  is a bilinear map  $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{F}$  satisfying the following:

- $\langle u, v \rangle = \langle v, u \rangle$  for all  $u, v \in V$  (Symmetry),
- $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$  and  $\langle \lambda u, v \rangle = \lambda \langle u, v \rangle$  for all  $u, v, w \in V$  and  $\lambda \in \mathbb{F}$  (Bilinearity),
- $\langle v, v \rangle \geq 0$  for all  $v \in V$  with equality when  $v = 0_V$  (Positive-definite).

Bilinearity must hold in both arguments however it can be derived from a single argument with the symmetry property.

**Definition 10.1.2** (Norm). Given a real-vector space  $V$  with an inner product  $\langle \cdot, \cdot \rangle$  the **norm** induced by the inner product of  $v \in V$  is:

$$\|v\| := \sqrt{\langle v, v \rangle}$$

**Definition 10.1.3** (Orthogonality). Two vectors  $u, v$  in an real or complex vector space  $V$  with inner product  $\langle \cdot, \cdot \rangle$  are **orthogonal** iff:  $\langle u, v \rangle = 0$ .

### 10.2 Orthonormal sets

Throughout the remainder of this section we will assume all vector spaces are over the real or complex numbers and will use the dot product as our inner product with its induced norm.

**Definition 10.2.1** (Orthogonal sets). A set of vectors  $\{v_1, v_2, \dots, v_n\}$  in a vector space is **orthogonal** if it is pairwise orthogonal.

**Definition 10.2.2** (Orthonormal sets). A set of vectors  $\{v_1, v_2, \dots, v_n\}$  in a vector space is **orthonormal** if it is orthogonal and satisfies  $\|u_i\| = 1$  for all  $i \in [1, n]$ .

**Theorem 10.2.3.** The columns of an orthogonal matrix  $P \in M_n(\mathbb{R})$  form an orthonormal set.

### 10.3 Gram-Schmidt process

The Gram-Schmidt process is a method of producing orthonormal bases.

**Algorithm 10.3.1** (Gramm-Shmidt process). Given a LI set  $\{v_1, v_2, \dots, v_r\} \in \mathbb{R}^n$  the **Gramm-Shmidt process** will produce the set of vectors  $\{w_1, w_2, \dots, w_r\} \in \mathbb{R}^n$  by the following:

$$\begin{aligned} w_1 &= v_1, \\ w_2 &= v_2 - \frac{w_1 \cdot v_2}{\|w_1\|^2} w_1, \\ w_3 &= v_3 - \left( \frac{w_1 \cdot v_3}{\|w_1\|^2} w_1 + \frac{w_2 \cdot v_3}{\|w_2\|^2} w_2 \right), \\ &\vdots \\ w_r &= v_r - \sum_{j=1}^{r-1} \frac{w_j \cdot v_r}{\|w_j\|^2} w_j. \end{aligned}$$

Note that each vector is the original vector  $v_i$  with its projection along all of the previous  $w_j$ s subtracted and therefore  $\{w_1, w_2, \dots, w_r\}$  is orthogonal. Finally,  $\{u_1, u_2, \dots, u_r\}$ , where  $u_i = \frac{w_i}{\|w_i\|}$  for all  $i \in [1, r]$ , is an orthonormal set with  $\text{Span}(\{u_1, u_2, \dots, u_r\}) = \text{Span}(\{v_1, v_2, \dots, v_r\})$ .

**Corollary 10.3.2.** Given some vector  $u \in \mathbb{R}^n$  there exists an orthogonal matrix in  $M_n(\mathbb{R})$  with  $u$  as its first column.

## 11 Real symmetric matrices

Throughout this section, unsurprisingly, all matrices will be assumed to be real.

### 11.1 Introduction

**Definition 11.1.1** (Self-adjoint matrices). If a matrix  $A \in M_n(\mathbb{R})$  is symmetric and satisfies  $A(u \cdot v) = (Au) \cdot v$  for all vectors  $u, v \in \mathbb{R}^n$ , we say  $A$  is **self-adjoint** w.r.t. the usual scalar product.

**Theorem 11.1.2.** If  $A \in M_n(\mathbb{R})$  is symmetric with  $\lambda \in \mathbb{C}$  a root of  $\chi_A(x) = 0$ ,  $\lambda \in \mathbb{R}$ .

**Corollary 11.1.3.** Real symmetric matrices have at least 1 real eigenvalue.

**Theorem 11.1.4.** If  $A \in M_n(\mathbb{R})$  is symmetric with discrete eigenvalues  $\lambda, \mu \in \mathbb{R}$ , their corresponding eigenvectors  $u, v \in \mathbb{R}^n$  satisfy  $u \cdot v = 0$ .

### 11.2 Spectral theorem

**Theorem 11.2.1** (Spectral theorem). If  $A \in M_n(\mathbb{R})$  is symmetric, then there exists an orthonormal matrix  $P$  such that  $P^{-1}AP$  is diagonal.

**Corollary 11.2.2.** Appropriately scaled eigenvectors of a symmetric matrix  $A \in M_n(\mathbb{R})$  form an orthonormal basis for  $\mathbb{R}^n$ .

# Chapter 3

## Groups

Lectured by Dr Michele Zordan

Typed by Yu Coughlin

Spring 2024

### Introduction

The following are supplementary reading:

- J B Fraleigh, A first course in abstract algebra, 2014
- R B J T Allenby, Rings, field and groups: an introduction to abstract algebra, 1991
- A W Knap, Basic Algebra, 2006

Lecture 1

Thursday

10/01/19

# 1 Binary operations and groups

**Definition 1.0.1** (Binary operation). Given a set  $G$  a **binary operation** on  $G$  is a mapping  $\cdot : G \times G \rightarrow G$  written  $\cdot(g, h) = g \cdot h$  (and sometimes  $gh$ ) for all  $g, h \in G$ .

**Definition 1.0.2** (Group). A **group** is a pair  $G = (G, \cdot)$ , for some set  $G$  and a binary operation  $\cdot$ , satisfying the following properties:

- G1  $(a \cdot b) \cdot c = a \cdot (b \cdot c)$  for all  $a, b, c \in G$  - the binary operation is **associative**,
- G2  $\exists e \in G$  such that  $\forall g \in G, g \cdot e = e \cdot g = g$  - there is an **identity** element,
- G3  $\forall g \in G, \exists g^{-1} \in G$  such that  $g \cdot g^{-1} = g^{-1} \cdot g = e$  - every element has an **inverse**.

In some literature, the condition of **closure** is also required however this is given in the fact that  $\cdot$  is a binary operation on  $G$ .

**Theorem 1.0.3** (Uniqueness). The identity element for some group  $G$  is unique. The inverse,  $g^{-1}$ , of any element  $g \in G$  is also unique.

**Lemma 1.0.4** (Inverse of product). Given a group  $G$  and the elements  $g_1, g_2, \dots, g_n \in G$  we have,

$$(g_1 g_2 \dots g_n)^{-1} = g_n^{-1} g_{n-1}^{-1} \dots g_1^{-1}.$$

**Definition 1.0.5** (Abelian Group). If a group  $G$  also satisfies the condition  $g \cdot h = h \cdot g$  for all  $g, h \in G$  - **commutativity**, then  $G$  is said to be an **abelian group**.

**Definition 1.0.6** (Powers of elements). Given a group  $G$  and some  $g \in G$  the  $n$ th **power** of  $g$  in  $G$  is defined recursively as,

$$g^n := \begin{cases} e & \text{if } n = 0 \\ g^{n-1}g & \text{if } n > 0. \\ (g^n)^{-1} & \text{if } n < 0 \end{cases}$$

**Definition 1.0.7** (Order of group). The **order** of a group  $G$ , written  $|G|$ , is the cardinality of the underlying set of  $G$ .

**Example 1.0.8** (Symmetric group). The **symmetric group of size  $n$** , denoted  $S_n$ , is the set of bijections on the interval  $[1, n]$ , for  $n \in \mathbb{N}$ , under function composition.

## 2 Subgroups

### 2.1 Subgroups

**Definition 2.1.1** (Subgroup). Given a group  $(G, \cdot)$  and a subset  $H \subseteq G$  we say  $(H, \cdot)$  is a **subgroup** of  $G$ , written  $H \leq G$ , if  $(H, \cdot)$  forms a group and

$$\forall h_1, h_2 \in H : h_1 \cdot h_2 \in H.$$

A subgroup,  $H$ , is a **proper subgroup** if  $H \neq G$ .  $\{e\}$  is the trivial subgroup.

**Theorem 2.1.2** (Subgroup test). Given a group  $(G, \cdot)$ ,  $(H, \cdot)$  is a subgroup iff:

- S1  $H$  is non-empty - **existence**,
- S2 for all  $h_1, h_2 \in H$  we have  $h_1 \cdot h_2 \in H$  - **closure under group operation**,
- S3 for all  $h \in H$  we have  $h^{-1} \in H$  - **closure under inverses**.



## 2.2 Cyclic groups and orders

**Definition 2.2.1** (Cyclic group). We say a group  $G$  is **cyclic** if there is an element  $g \in G$  such that

$$G = \langle g \rangle := \{g^n : n \in \mathbb{N}\}.$$

We say that  $G$  is **generated** by  $g$  or  $g$  is a **generator** of  $G$ .

**Definition 2.2.2** (Order of elements). Given a group  $G$  and some  $g \in G$ , the **order** of  $g$  in  $G$ , written  $\text{ord } g$ , is the smallest positive integer  $n$  such that  $g^n = e$  or  $\infty$  if no such  $n$  exists.

**Theorem 2.2.3.** Suppose  $G$  is a cyclic group generated by  $g$  with  $|G| = n$ ,  $\text{ord } g = |\{e, g, g^2, \dots, g^{n-1}\}| = |G| = n$ .

**Theorem 2.2.4.** Suppose  $G$  is a cyclic group with  $G = \langle g \rangle$ , the three statements:

1.  $H \leq G \implies H$  is cyclic,
2. suppose  $|G| = n$  and  $m \in \mathbb{Z}$  with  $f = \gcd(m, n)$ ,

$$\langle g^m \rangle = \langle g^d \rangle \text{ and } |\langle g^m \rangle| = \frac{n}{d}.$$

In particular,  $\langle g^m \rangle = G$  iff  $\gcd(m, n) = 1$ ,

3. if  $|G| = n$  and  $k \leq n$ , then  $G$  has a subgroup of order  $k$  iff  $k|n$ , this subgroup is  $\langle g^{n/k} \rangle$ .

**Definition 2.2.5** (Euler totient). The **Euler totient** function  $\phi$  is defined as  $\phi(n) := |\{k \in \mathbb{N} : k \leq n \text{ and } \gcd(k, n) = 1\}|$ .

**Corollary 2.2.6.** For  $n \in \mathbb{N}$ :

$$\sum_{d|n} \phi(d) = n.$$

## 2.3 Cosets

**Definition 2.3.1** (Coset). Given a group  $G$  with  $H \leq G$  and  $g \in G$  then

$$gH := \{gh : h \in H\},$$

is a **left coset** of  $H$  in  $G$  (the definition of a **right coset** follows clearly).

**Note 2.3.2.** For the rest of this section, unless specified otherwise, a coset is assumed to be a left-coset.

**Theorem 2.3.3.** Given a group  $G$  with  $H \leq G$ , all cosets of  $H$  in  $G$  have the same size.

**Theorem 2.3.4.** If  $G$  is a finite group with  $H \leq G$ , the left cosets of  $H$  form a partition of  $G$ .

## 2.4 Lagrange's theorem

**Theorem 2.4.1** (Lagrange's theorem). If  $G$  is a finite group and  $H \leq G$ ,  $|H|$  divides  $|G|$ .

**Corollary 2.4.2.** Given a group  $G$  with  $H \leq G$ , the relation  $\sim$  on  $G$  given by:  $g \sim k$  iff  $g^{-1}k \in H$ , is an equivalence relation with equivalence classes given by cosets of  $H$ .

**Corollary 2.4.3.** Given a group  $G$  of order  $n$ , for all  $g \in G$ ,  $\text{ord } g|n$  and  $g^n = e$ .

**Corollary 2.4.4** (Fermat's little theorem). Let  $p$  be prime. If  $x \in \mathbb{Z}$  and  $p \nmid x$ , then  $x^{p-1} \equiv 1 \pmod{p}$ .

## 2.5 Generating groups

**Definition 2.5.1.** Given a group  $G$  with  $S \subseteq G$ ,  $S^{-1} := \{g^{-1} \in G : g \in S\}$ .

**Definition 2.5.2** (Subgroup generated by a set). Let  $G$  be a group with non-empty  $S \subseteq G$ . The **subgroup generated by  $S$**  is defined as

$$\langle S \rangle := \{g_1 g_2 \dots g_k \in G : k \in \mathbb{N} \text{ and } g_i \in S \cup S^{-1} \text{ for all } i \in [1, k]\}.$$

**Lemma 2.5.3.** Given a group  $G$  with non-empty  $S \subseteq G$ ,  $\langle S \rangle \leq G$  and,  $H \leq G$ ,  $S \subseteq H \implies \langle S \rangle \leq H$ . This is equivalent to saying " $\langle S \rangle$  is the smallest subgroup of  $G$  containing  $S$ ".

### 3 Group homomorphisms

**Definition 3.0.1** (Group homomorphism). If  $(G, \cdot)$  and  $(H, *)$  are groups,  $\phi : G \rightarrow H$  is a **group homomorphism** iff  $\phi(g_1) * \phi(g_2) = \phi(g_1 \cdot g_2)$  for all  $g_1, g_2 \in G$ . If  $\phi$  is bijective then it is called a **group isomorphism** with  $G$  and  $H$  being **isomorphic**, written  $G \cong H$ .

**Example 3.0.2.** The **determinant** is a group homomorphism, suppose  $\mathbb{F}$  is a field:

$$\det : \text{GL}(n, \mathbb{F}) \rightarrow (\mathbb{F}^*, \times).$$

**Lemma 3.0.3.** If  $G, H$  are groups with  $\phi : G \rightarrow H$ ,

1.  $\phi(e_G) = e_H$ ,
2.  $\phi(g^{-1})(\phi(g))^{-1}$  for all  $g \in G$ .

**Definition 3.0.4** (Image and kernel of group homomorphism). If  $G, H$  are groups with  $\phi : G \rightarrow H$ , the **image** of  $\phi$  is:

$$\text{im } \phi := \{h \in H : \exists g \in G, h = \phi(g)\}.$$

and the **kernel** of  $\phi$  is

$$\ker \phi := \{g \in G : \phi(g) = e_H\}.$$

These are each subgroups of  $H$  and  $G$  respectively.

**Lemma 3.0.5.** A group homomorphism,  $\phi : G \rightarrow H$ , is injective iff  $\ker \phi = \{e_H\}$ .

**Theorem 3.0.6.** The composition of two compatible group homomorphisms is also a group homomorphism.

**Theorem 3.0.7.** All cyclic groups of the same order are isomorphic.

### 4 Symmetric groups

#### 4.1 Disjoint cycle decomposition

**Definition 4.1.1.** If  $f, g \in S_n$  and  $x \in [1, n]$  then  $f$  **fixes**  $x$  if  $f(x) = x$  and  $f$  **moves**  $x$  otherwise.

**Definition 4.1.2.** The **support** of  $f \in S_n$  is the set of points  $f$  moves,  $\text{supp}(f) := \{x \in [1, n] : f(x) \neq x\}$ .

**Definition 4.1.3.** If  $f, g \in S_n$  satisfy  $\text{supp}(f) \cap \text{supp}(g) = \emptyset$ ,  $f$  and  $g$  are **disjoint**.

**Lemma 4.1.4.** If  $f, g \in S_n$  are disjoint,  $fg = gf$ .

**Definition 4.1.5** (Cycles). If  $f \in S_n$  with  $i_1, i_2, \dots, i_r \in [1, n]$  for some  $r \leq n$  such that,

$$f(i_s) = i_{s+1 \bmod r} \text{ for all } s \in [1, r],$$

with  $f$  fixing all other elements of  $[1, n]$ , then  $f$  is a **cycle of length  $r$**  or an  **$r$ -cycle** and we write  $f = (i_1 i_2 \dots i_r)$ .

**Theorem 4.1.6** (Disjoint cycle form). if  $f \in S_n$  then there exists  $f_1, f_2, \dots, f_k \in S_n$  all with disjoint supports such that  $f = f_1 f_2 \dots f_k$ . If we further have, for all  $i \in [1, k]$ , both  $f_i$  is not a 1-cycle when  $f \neq \text{id}$  and  $\text{supp}(f_i) \subseteq \text{supp}(f)$ . We say  $f$  is in **disjoint cycle form** or **d.c.f.**

**Theorem 4.1.7** (Uniqueness of disjoint cycles). The disjoint cycle form of some  $f \in S_n$  is unique up to rearrangement.

**Theorem 4.1.8.** If  $f \in S_n$  is written in d.c.f as  $f = f_1 f_2 \dots f_k$  where  $f_i$  is an  $r_i$ -cycle for  $i \in [1, k]$  then,

1.  $f^m = \text{id}$  iff  $f_i^m = \text{id}$  for all  $i \in [1, k]$ ,
2.  $\text{ord}(f) = \text{lcm}(r_1, r_2, \dots, r_k)$ .

## 4.2 Alternating groups

**Theorem 4.2.1.** Every permutation in  $S_n$  can be written as the product of 2-cycles.

**Definition 4.2.2** (Sign of a permutation). We define the **sign** of a permutation with the group homomorphism,  $\text{sgn} : S_n \rightarrow \{-1, 1\}$  with  $\text{sgn}(i\ j) := -1$  for all  $i, j \in [1, n]$  with  $i \neq j$ . This is defined over all permutations by the decomposition into 2-cycles, the sign of a permutation is unique. We say  $f \in S_n$  is **even** if  $f \in \ker(\text{sgn})$  and **odd** otherwise.

**Definition 4.2.3** (Alternating group). The **alternating group** of size  $n$  is  $A_n := \ker(\text{sgn})$  with  $A_n \leq S_n$ .

## 4.3 Dihedral groups

**Definition 4.3.1** (Dihedral group). The **dihedral group** of order  $2n$ , denoted  $D_{2n}$ , is the group of symmetries of a regular  $n$ -gon in  $\mathbb{R}^3$  centered at the origin, it is often written at

$$D_{2n} = \{e, r, r^2, \dots, r^{n-1}, s, sr, sr^2, \dots, sr^{n-1}\},$$

where  $r$  is a rotation by  $\frac{2\pi}{n}$  and  $s$  is the reflection along the centre of the polygon and the first vertex.

**Theorem 4.3.2.** The elements of  $D_{2n}$  can be written as elements of  $S_n$  giving  $D_{2n} \leq S_n$ . Specifically,  $r = (1\ 2\ \dots\ n)$  and  $s = (1)(2\ n)(3\ n-1)\dots$  or  $(1\ n)(2\ n-1)\dots$  if  $n$  is odd or even respectively.

# Chapter 4

# Calculus

Lectured by Professor Demetrios Papageorgiou  
Typed by Yu Coughlin  
Autumn 2023

## Introduction

The following are suggested textbooks:

- G F Simmons, Calculus with Analytic Geometry, 1995
- J Stewart, Calculus, 2011
- S Lang, A First Course in Calculus, 1986
- S Lang, Undergraduate Analysis, 1997
- J Marsden and A Weinstein, Calculus I and Calculus II, 1985

**Note.** The actual majority of MATH40004A Calculus was a less formal and more example / application based derivation of the entirety of MATH40002 Analysis. As all of this content can be found in the corresponding document for Analysis, it isn't included in here.

Lecture 1  
Thursday  
10/01/19

# 1 Lengths, volumes and surfaces

## 1.1 Lengths

**Theorem 1.1.1** (Arc length). The **arc length** of the curve  $y = f(x)$  along  $[a, b]$  is given by

$$\int_a^b \sqrt{1 + (f'(x))^2} \, dx$$

**Theorem 1.1.2** (Distance and velocity of parameterised curves). If a curve is parameterised by  $(x(t), y(t), z(t))$ , the **distance travelled** from time  $t_0$  to  $t$  is given by:

$$L(t) = \int_{t_0}^t \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2} \, dt$$

which naturally leads to the velocity at  $t$ :

$$v(t) = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}$$

## 1.2 Volumes and volumes of revolution

**Theorem 1.2.1** (Volume). If the cross sectional area of a shape when cut by a plane at  $x = x_0$  is given by  $A(x_0)$  for all  $x_0 \in [a, b]$ , the volume of the shape is given by

$$V = \int_a^b A(x) \, dx$$

**Theorem 1.2.2** (Disk method). The **volume of revolution** of  $y = f(x)$  about the  $x$ -axis from  $x = a$  to  $x = b$  is given by,

$$V_x = \int_a^b \pi (f(x)^2) \, dx$$

**Theorem 1.2.3** (Shell method). The **volume of revolution** of  $y = f(x)$  about the  $y$ -axis from  $y = a$  to  $y = b$  is given by,

$$V_y = \int_a^b \pi (f^{-1}(x)^2) \, dx = \int_a^b 2\pi x f(x) \, dx$$

## 1.3 Surfaces

**Theorem 1.3.1.** The **surface area of revolution** of  $y = f(x)$  about the  $x$ -axis from  $x = a$  to  $x = b$  is given by,

$$S_x = \int_a^b 2\pi f(x) \sqrt{1 + (f'(x))^2} \, dx$$

**Theorem 1.3.2.** The **surface area of revolution** of  $y = f(x)$  about the  $y$ -axis from  $y = a$  to  $y = b$  is given by,

$$S_y = \int_a^b 2\pi x \sqrt{1 + (f'(x))^2} \, dx$$

## 1.4 Centres of mass

**Theorem 1.4.1** (1D discrete case). If we have a system of  $n$  particles each with mass  $m_k$  and position  $x_k$  we can define the **centre of mass** at  $\bar{x}$  by

$$\bar{x} = \frac{\sum_{k=1}^n m_k x_k}{\sum_{k=1}^n m_k}$$

**Theorem 1.4.2** (2D continuous case). If we have a region limited by  $f(x)$  and  $g(x)$ , give  $g(x) \leq f(x)$  for all  $x \in [a, b]$ , with uniform mass, the coordinates of the **centre of mass**,  $(\bar{x}, \bar{y})$  is

$$\bar{x} = \frac{\int_a^b x(f(x) - g(x)) \, dx}{\int_a^b f(x) - g(x) \, dx} \quad \bar{y} = \frac{\int_a^b \frac{f(x)^2 - g(x)^2}{2} \, dx}{\int_a^b f(x) - g(x) \, dx}$$

**Theorem 1.4.3** (Pappu's theorem). If  $R$  is a reigon with area  $A$  lying on one side of the line  $l$ ,  $V = Ad$  is the volume obtained by rotation  $R$  about  $l$ , where  $d$  is the distance travelled by the **com** when  $R$  is rotated about  $l$ .

## 1.5 Moments of inertia

**Theorem 1.5.1.** Given a curve  $y = f(x)$  in the interval  $[a, b]$ , this is representing a wire in a given shape, and have the density per unit lenght of the wire at a given  $x$  be  $\rho(x)$ , the **moment of inertia** of the curve about the  $x$  and  $y$  axis respectively is given by

$$I_x = \int_a^b \rho(x) f(x)^2 \sqrt{1 + f'(x)^2} \, dx \quad I_y = \int_a^b \rho(x) x^2 \sqrt{1 + f'(x)^2} \, dx$$

## 1.6 Polar coordinates

**Definition 1.6.1** (Polar coordinates). A parameterisation of  $x, y$  is  $r, \theta$  with  $x = r \cos(\theta)$  and  $y = r \sin(\theta)$ .

**Theorem 1.6.2** (Polar arc length). The arc length of a curve,  $r = f(\theta)$  in polar coordinates between angles  $\alpha, \beta$  is given by

$$L = \int_{\alpha}^{\beta} \sqrt{\left(\frac{dr}{d\theta}\right)^2 + r^2} \, d\theta$$

**Theorem 1.6.3** (Polar area). The area of a polar curve,  $r = f(\theta)$  between angles  $\alpha, \beta$  is given by

$$A = \frac{1}{2} \int_{\alpha}^{\beta} f(\theta)^2 \, d\theta$$

# 2 Fourier series

## 2.1 Orthogonal and orthonormal function spaces

**Definition 2.1.1** (Inner product of functions). If  $f, g : [a, b] \rightarrow \mathbb{R}$  are integrable on  $[a, b]$ , the their **inner product** is defined as

$$\langle f, g \rangle := \int_a^b f(x)g(x) \, dx$$

**Definition 2.1.2** (Orthogonal and orthonormal system). If  $\mathcal{S} = \{\phi_0, \phi_1, \dots\}$  is a collection of integrable real functions on  $[a, b]$ , iff  $\langle \phi_n, \phi_m \rangle = 0$  for all  $n \neq m$  then  $\mathcal{S}$  is an **orthogonal system** on  $[a, b]$ . Furthermore,  $\mathcal{S}$  is a **orthonormal system** on  $[a, b]$  iff  $\|\phi_n\| := \langle \phi_n, \phi_n \rangle = 1$  for all  $n$ .

**Theorem 2.1.3.** The system

$$\mathcal{S} = \left\{ \frac{1}{\sqrt{2\pi}}, \frac{\cos(x)}{\sqrt{2\pi}}, \frac{\sin(x)}{\sqrt{2\pi}}, \frac{\cos(2x)}{\sqrt{2\pi}}, \frac{\sin(2x)}{\sqrt{2\pi}}, \dots \right\}$$

is orthonormal on all closed intervals of length  $2\pi$ .

## 2.2 Periodic functions

**Definition 2.2.1** (Periodic function). A function  $f : \mathbb{R} \rightarrow \mathbb{R}$  is **periodic** with period  $T$  iff  $f(x + T) = f(x)$  for all  $x \in \mathbb{R}$ .

**Definition 2.2.2** (Discontinuity). When periodically extending a function, if  $\lim_{x \rightarrow \xi+} f(x) \neq \lim_{x \rightarrow \xi-} f(x)$ , we

set  $f(\xi) := \frac{1}{2} \left[ \lim_{x \rightarrow \xi+} f(x) + \lim_{x \rightarrow \xi-} f(x) \right]$

**Theorem 2.2.3** (Integral over period). If  $f(x)$  is a  $T$  periodic function, for all  $a, b \in \mathbb{R}$  we have

$$\int_{a+T}^{b+T} f(x) dx = \int_a^b f(x) dx$$

## 2.3 Trigonometric polynomials

**Definition 2.3.1** (Trigonometric polynomial). A **trigonometric polynomial** is a function in the form

$$S_n(x) = \frac{1}{2}a_0 + \sum_{k=1}^n (a_k \cos(kx) + b_k \sin(kx))$$

**Theorem 2.3.2.** Using euler's identity we can rewrite a trigonometric polynomial

$$S_n(x) = \frac{1}{2}a_0 + \sum_{k=1}^n (a_k \cos(kx) + b_k \sin(kx)) \text{ as } \sum_{k=-n}^n (\gamma_k e^{ikx}) \text{ where } \gamma_k = \begin{cases} \frac{1}{2}a_0 & \text{if } k = 0 \\ \frac{1}{2}(a_k - ib_k) & \text{if } k \in [1, n] \\ \gamma_k^* & \text{otherwise} \end{cases}$$

## 2.4 Fourier series

[Fourier series]

**Definition 2.4.1** (Fourier series). If  $f(x)$  is  $2L$  periodic then its **Fourier series** is given by

$$f(x) := \frac{1}{2}a_0 + \sum_{n=1}^{\infty} \left[ a_n \cos\left(\frac{n\pi x}{L}\right) + b_n \sin\left(\frac{n\pi x}{L}\right) \right] \text{ where}$$

$$a_n := \frac{1}{\pi} \int_{-L}^L f(x) \cos\left(\frac{n\pi x}{L}\right) dx, \quad b_n := \frac{1}{\pi} \int_{-L}^L f(x) \sin\left(\frac{n\pi x}{L}\right) dx$$

**Lemma 2.4.2** (Riemann-Lebesgue). If the function  $f(x)$  is integrable on  $[a, b]$  then

$$I_\lambda := \int_a^b g(x) \sin(\lambda x) dx \rightarrow 0 \text{ as } \lambda \rightarrow \infty$$

**Theorem 2.4.3** (Parseval's). If  $f(x)$  is periodic on  $2\pi$  and is represented by its Fourier series,

$$\frac{1}{\pi} \int_{-\pi}^{\pi} f^2(x) dx = \frac{1}{2}a_0^2 + \sum_{n=0}^{\infty} (a_n^2 + b_n^2)$$

**3 Plane curves****4 Laplace transform**



# Chapter 5

# Differential Equations

Lectured by Professor Demetrios Papageorgiou  
Typed by Yu Coughlin  
Spring 2024

## Introduction

The following are suggested textbooks:

- G F Simmons, Calculus with Analytic Geometry, 1995
- J Stewart, Calculus, 2011
- S Lang, A First Course in Calculus, 1986
- S Lang, Undergraduate Analysis, 1997
- J Marsden and A Weinstein, Calculus I and Calculus II, 1985

Lecture 1  
Thursday  
10/01/19

- 1    Fourier transform
- 2    Ordinary differential equations
- 3    Qualitative analysis
- 4    Bifurcations
- 5    Multivariate calculus
- 6    Partial differential equations

# Chapter 6

# Probability

Lectured by Professor Almut Veraart  
Typed by Yu Coughlin  
Autumn 2023

## Introduction

The following are complementary reading for the course.

- G. Grimmett and D. J. A. Welsh, Probability: An Introduction, 1986
- J. K. Blitzstein and J. Hwang, Introduction to Probability, 2019
- D. F. Anderson et al, Introduction to Probability, 2018
- S. M. Ross, Introduction to Probability Models, 2014
- G. Grimmett and D. Stirzaker, Probability and Random Processes, 2001
- G. Grimmett and D. Stirzaker, One Thousand Exercises in Probability, 2009

Lecture 1  
Monday  
30/10/2023

## 0.1 Sample spaces and set theory

**Definition 0.1.1.** The **sample space**  $\Omega$  is the set of all possible outcomes of an experiment. An element of the sample space  $\omega \in \Omega$  is a **sample point**.

**Examples 0.1.2.** When flipping a coin  $\Omega = \{H, T\}$ . When rolling a standard die  $\Omega = \{1, 2, 3, 4, 5, 6\}$ .

**Definition 0.1.3.** Subsets of  $\Omega$  are collections of sample points and called **events**.

Suppose events  $A, B \subseteq \Omega$ :

- $A \cup B$  is the event that  $A$  or  $B$  or both occur,
- $A \cap B$  is the event that  $A$  and  $B$  both occur,
- $A^c = \bar{A}$  is the event that occurs iff  $A$  does not occur.

Let  $\mathcal{I}$  be a general index set with  $A_i \subseteq \Omega, \forall i \in \mathcal{I}$  and  $B \subseteq \Omega$ . The following identities hold.

$$\left(\bigcup_{i \in \mathcal{I}} A_i\right)^c = \bigcap_{i \in \mathcal{I}} A_i^c, \quad \left(\bigcap_{i \in \mathcal{I}} A_i\right)^c = \bigcup_{i \in \mathcal{I}} A_i^c, \quad B \cap \left(\bigcup_{i \in \mathcal{I}} A_i\right) = \bigcap_{i \in \mathcal{I}} (A_i \cup B), \quad B \cup \left(\bigcap_{i \in \mathcal{I}} A_i\right) = \bigcap_{i \in \mathcal{I}} (A_i \cap B).$$

These are **De Morgan's Laws** and **Distributivity** respectively.

Lecture 2  
Tuesday  
31/10/2023

## 0.2 Interpretation of probability

**Definition 0.2.1.** The **Cardinality** of a set, denoted  $\text{card}(A)$  or  $|A|$  is the number of elements in the set  $A$ .

**Definition 0.2.2.** Two sets have the same cardinality iff there exists a bijection between the them.

**Definition 0.2.3.**  $A$  is **finite** if it has as finite numbers of elements,  $A$  is **countably infinite** if there exists a bijection  $f: A \rightarrow \mathbb{N}$ ,  $A$  is **countable** if it is finite or countable infinite,  $A$  is **uncountable** or **uncountable infinite** if it isn't countable.

Samples spaces can be countable or uncountable.

**Definition 0.2.4** (Naive probability). Suppose  $|A| < \infty$  and we want to assign a probability to  $A \subseteq \Omega$ .

$$P_{\text{Naive}}(A) := \frac{|A|}{|\Omega|} \implies P(A^c) = 1 - P(A).$$

This Naive example does not consider when  $|A|$  is infinite but of finite area.

**Example 0.2.5.** Let  $\Omega = \{(x, y) \in \mathbb{R}^2, x^2 + y^2 = 1\}$  and  $A \subseteq \Omega$ . Define:

$$P(A) := \frac{\text{area of } A}{\text{area of } \Omega}$$

In the case where  $A = \{(x, y) \in \mathbb{R}^2, x^2 + y^2 = 0.5^2\}$  we have  $P(A) = 0.25$

**Remark 0.2.6.** For classical / naive probability we require  $|\Omega| < \infty$  or the “area” of  $\Omega$  be finite.

**Definition 0.2.7** (Limiting frequency). Consider  $n_{\text{total}}$  repetitions of an experiment and  $n_A$  the number of time  $A$  occurs.

$$P(A) := \lim_{n_{\text{total}} \rightarrow \infty} \frac{n_A}{n_{\text{total}}}$$

Unfortunately,  $n_{\text{total}} \rightarrow \infty$  is often hard to conceive with finite representations not necessarily being representative.

**Definition 0.2.8** (Subjective probability). For an event  $A$  assign the probability  $P(A)$  based on our own personal beliefs. The subjective probability need not be the same for different individuals, and despite its appearance it remains a valid interpretation of probability.

**Remark 0.2.9.** All three interpretations of probability depend of assumptions about the experiment.

Lecture 3  
Friday  
03/11/2023

# 1 Counting

## 1.1 Multiplication principle

Computing naive probabilities often requires some combinatorics.

**Definition 1.1.1** (Multiplication principle). If we perform an experiment  $A$  that has  $a$  possible outcomes and an experiment  $B$  with  $b$  possible outcomes (in any order) then the number of outcomes of the **compound experiment** will be  $ab$ .

**Remark 1.1.2.** When dealing with repetitions of the same experiment (with sample space  $\Omega$ , the sample space is given by the Cartesian product of the individual samples spaces.

$$\Omega_1 \times \Omega_2 \times \cdots \times \Omega_n := \{(\omega_1, \omega_2, \dots, \omega_n) : \omega_i \in \Omega_i\}.$$

The cardinality of this samples space follows from the multiplication principle.

## 1.2 Power sets

**Definition 1.2.1** (Power Set). Given a set  $A$  its **power set** is defined as:

$$\mathcal{P}(A) := \{X : X \subseteq A\}.$$

**Theorem 1.2.2.** If  $A$  is a finite set,  $|\mathcal{P}(A)| = 2^{|A|}$ .

## 1.3 Combinatorial coefficients

**Definition 1.3.1** (Factorial). Let  $n \in \mathbb{N}$  the **factorial** of  $n$  is defined as:

$$n! := \prod_{i=1}^n i.$$

**Definition 1.3.2** (Descending factorial). Let  $k, n \in \mathbb{N}$  with  $k \leq n$  the **descending factorial** denoted  $(n)_k$  is defined as:

$$(n)_k := n(n-1) \cdots (n-k+1) = \prod_{i=0}^{k-1} (n-i) = \prod_{j=n-k+1}^n j = \frac{n!}{(n-k)!}.$$

**Definition 1.3.3** (Binomial coefficient). Let  $k, n \in \mathbb{N}_0$  the **binomial coefficient** is the number of subsets of size  $k$  of a set  $n$ :

$$\binom{n}{k} := \begin{cases} \frac{n(n-1) \cdots (n-(k-1))}{k!} = \frac{(n)_k}{k!} = \frac{n!}{(n-k)!k!} & \text{if } k \leq n \\ 0 & \text{otherwise.} \end{cases}$$

## 1.4 Sampling with and without replacement

“Definitions” given in the context of drawing balls from an urn,  $S = \{1, 2, \dots, n\}$ .

**Definition 1.4.1** (Ordered sampling with replacement). Take out a ball from  $S$ , note its number, put it back; repeat this  $k$  times. The sample space for this experiment is  $\Omega = S^k$ .

**Definition 1.4.2** (Ordered sampling without replacement). Take out a ball from  $S$ , note its number but **do not** put it back; repeat  $k < n$  times. There are  $|\Omega| = (n)_k$  possible outcomes.

**Definition 1.4.3** (Unordered sampling without replacement). We take  $k$  balls out of the urn, there are  $\binom{n}{k}$  possibilities.

**Definition 1.4.4** (Unordered sampling with replacement). We take  $k$  balls out of the urn, with the stars and bars argument: there must be  $k$  stars divided by  $n-1$  bars giving us:

$$|\Omega| = \binom{n+k-1}{k} = \binom{n+k-1}{n-1}.$$

Lecture 4  
Monday  
06/11/2023

Lecture 5  
Tuesday  
07/11/2023

## 2 Axiomatic probability

### 2.1 Event space

We do not always want to consider all subsets of  $\Omega$  so denote  $\mathcal{F} \subseteq \mathcal{P}(\Omega)$  the **event space**, which contains the events we are allowed to consider.  $\mathcal{F}$  must always be a  $\sigma$ -algebra.

**Definition 2.1.1** (Algebra).  $\mathcal{F}$  is an **algebra** (or a field) on  $\Omega$  iff:

1.  $\emptyset \in \mathcal{F}$ ,
2.  $A \in \mathcal{F} \implies A^c \in \mathcal{F}$ ,
3.  $A, B \in \mathcal{F} \implies A \cup B \in \mathcal{F}$ .

**Definition 2.1.2** ( $\sigma$ -algebra).  $\mathcal{F}$  is a  **$\sigma$ -algebra** (or a  $\sigma$ -field) on  $\Omega$  iff:

1.  $\emptyset \in \mathcal{F}$ ;
2.  $A \in \mathcal{F} \implies A^c \in \mathcal{F}$ ,
3. For all  $i$  in some countable indexing set  $\mathcal{I}$ ,  $A_i \in \mathcal{F} \implies \bigcup_{i \in \mathcal{I}} A_i \in \mathcal{F}$ .

**Remark 2.1.3.** 1. Any algebra is closed under finite unions and finite intersections,

2. Any  $\sigma$ -algebra is closed under countable intersections,

3. Any ( $\sigma$ -)algebra on  $\Omega$  contains  $\Omega$ .

**Definition 2.1.4** (Trivial sigma algebra). The **trivial sigma algebra** on  $\Omega$  is defined as  $\mathcal{F}_{trivial} := \{\emptyset, \Omega\}$ .

**Example 2.1.5** (Smallest  $\sigma$ -algebra of an element). For some  $A \subseteq \Omega$ , the sigma algebra  $\mathcal{F}_A := \{\emptyset, A, A^c, \Omega\}$  is the smallest  $\sigma$ -algebra on  $\Omega$  (smallest cardinality) that contains  $A$ .

Lecture 6  
Friday  
10/11/2023

### 2.2 Probability measure

**Definition 2.2.1** (Probability measure). A mapping  $P : \mathcal{F} \rightarrow \mathbb{R}$  is a **probability measure** on  $(\Omega, \mathcal{F})$  iff:

1.  $P(A) \geq 0$  for all  $A \in \mathcal{F}$ ;
2.  $P(\Omega) = 1$ ;
3. for a countable, disjoint sequence of events  $(A_i)_{i \in \mathcal{I}}$  on an indexing set  $\mathcal{I}$ :

$$P\left(\bigcup_{i \in \mathcal{I}} A_i\right) = \sum_{i \in \mathcal{I}} P(A_i).$$

### 2.3 Probability space

**Definition 2.3.1** (Probability space). A **probability space** is a triple  $(\Omega, \mathcal{F}, P)$ , with  $\Omega$  a sample space,  $\mathcal{F}$  a  $\sigma$ -algebra on  $\Omega$ , and  $P$  a probability measure on  $(\Omega, \mathcal{F})$ .

**Corollary 2.3.2.** For  $A, B \in \mathcal{F}$ :

1.  $P(A^c) = 1 - P(A)$ ,
2.  $A \subseteq B \implies P(A) \leq P(B)$ ,
3.  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .

Lecture 7  
Monday  
13/11/2023

## 3 Conditional probability

**Definition 3.0.1** (Conditional probability measure). Consider the probability space  $(\Omega, \mathcal{F}, P)$  and some event  $B \in \mathcal{F}$  with  $P(B) > 0$ , we construct the probability measure  $Q$  on  $(\Omega, \mathcal{F})$  by

$$Q(A) := \frac{P(A \cap B)}{P(B)}.$$

Denote the **conditional probability** of  $A$  given  $B$  by  $P(A|B) = Q(A)$ .

Lecture 8  
Tuesday  
14/11/2023

### 3.1 Bayes' rule and total probability

**Theorem 3.1.1** (Bayes' rule). For  $A, B \in \mathcal{F}$  with  $P(A) > 0, P(B) > 0$  we have,

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

**Definition 3.1.2** (Partition of a set). A partition of some set  $\Omega$  is a collection  $\{B_i, i \in \mathcal{I}\}$  for some countable index set  $\mathcal{I}$  with  $B_i \cap B_j = \emptyset$  for all  $i, j \in \mathcal{I}$  with  $i \neq j$  and  $\bigcup_{i \in \mathcal{I}} B_i = \Omega$ .

**Theorem 3.1.3** (Total probability). Given some partition  $\{B_i, i \in \mathcal{I}\}$  of  $\Omega$  with  $P(B_i) > 0$  for all  $i \in \mathcal{I}$  and some event  $A \in \mathcal{F}$ ,

$$P(A) = \sum_{i \in \mathcal{I}} P(A \cap B_i) = \sum_{i \in \mathcal{I}} P(A|B_i)P(B_i).$$

These two theorems can then be combined to form the following.

**Theorem 3.1.4** (Bayes' rule with extra conditioning). For events  $A, B, E \in \mathcal{F}$  with  $P(A \cap E) > 0, P(B \cap E) > 0$  we have

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

**Theorem 3.1.5** (Total probability with extra conditioning). Given events  $A, E \in \mathcal{I}$  with  $P(E) > 0$  and some partition  $\{B_i, i \in \mathcal{I}\}$  of  $\Omega$  with  $P(B_i \cap E) > 0$  for all  $i \in \mathcal{I}$ ,

$$P(A|E) = \sum_{i \in \mathcal{I}} \frac{P(A \cap B_i \cap E)}{P(E)} = \sum_{i \in \mathcal{I}} P(A|B_i \cap E)P(B_i|E).$$

Lecture 9  
Friday  
17/11/2023

## 4 Independence

### 4.1 Event independence

Two events  $A, B \in \mathcal{F}$  will be independent iff the occurrence of one does not effect the probability the other occurs, i.e  $P(A|B) = P(A)$  and vice versa.

**Definition 4.1.1** (Independent events). Two events  $A, B \in \mathcal{F}$  are said to be **independent** iff

$$P(A \cap B) = P(A)P(B),$$

and **dependent** otherwise.

**Corollary 4.1.2.** If  $A$  and  $B$  are independent then so are all pairs of their complements.

**Definition 4.1.3** (General independence). A finite collection of events  $\{A_1, A_2, \dots, A_n\}$  is independent iff

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2) \dots P(A_n),$$

and similarly a countably or uncountably infinite collection of events is independent iff each finite subcollection is independent.

Lecture 10  
Monday  
20/11/2023

### 4.2 Conditional independence

**Definition 4.2.1** (Conditional independence). Given the events  $A, B, C \in \mathcal{F}$  with  $P(C) > 0$  we say  $A$  and  $B$  are **conditional independent** given  $C$  iff,

$$P(A \cap B|C) = P(A|C)P(B|C).$$

### 4.3 Product rule for general independence

The upcoming subsection may seem disparate, they are however necessary parts to the omitted proof of the product rule for general independence and therefore deemed relevant.

**Definition 4.3.1** (Set difference). Given two set  $A, B \in \Omega$  the **set difference** of  $A$  and  $B$  is defined as,  $A \setminus B := A \cap B^c$ .

**Lemma 4.3.2.** Any countable union of sets can be written as a countable union of disjoint sets.

**Definition 4.3.3** (Increasing and decreasing sets). A sequence of sets  $(A_i)_{i=1}^{\infty}$  is said to **increase** to  $A$  (written  $A_i \uparrow A$ ) iff  $A_1 \subseteq A_2 \subseteq \dots$  and  $\bigcup_{i=1}^{\infty} A_i = A$ . The definition for a sequence of sets  $(B_i)_{i=1}^{\infty}$  to **decrease** to a set  $B$  ( $B_i \downarrow B$ ) is defined similarly.

**Theorem 4.3.4** (Continuity property of probability measures). If  $A_1, A_2, \dots \in \mathcal{F}$  with  $A_i \uparrow A$  or  $A_i \downarrow A$  for some  $A \in \mathcal{F}$ ,

$$\lim_{i \rightarrow \infty} P(A_i) = P(\lim_{i \rightarrow \infty} A_i) = P(A).$$

**Theorem 4.3.5** (Product rule for general independence). Given a countably infinite set of independent events  $A_1, A_2, \dots \in \mathcal{F}$ ,

$$P\left(\bigcap_{i=1}^{\infty} A_i\right) = \prod_{i=1}^{\infty} P(A_i).$$

## 5 Discrete random variables

### 5.1 Images and their properties

throughout this subsection we will be considering the function  $f : \mathcal{X} \rightarrow \mathcal{Y}$ .

**Definition 5.1.1** (Image). For some subset  $A \subseteq \mathcal{X}$  we define the **image** of  $A$  under  $f$  by,

$$f(A) := \{y \in \mathcal{Y} : \exists x \in A, y = f(x)\} = \{f(x) : x \in A\}.$$

When  $A = \mathcal{X}$ ,  $f(\mathcal{X}) = \text{im } f$ .

**Definition 5.1.2** (Pre-image). For some subset  $B \subseteq \mathcal{Y}$  we now define the **pre-image** of  $B$  under  $f$  by,

$$f^{-1}(B) := \{x \in \mathcal{X} : f(x) \in B\}.$$

Despite the similar notation to the inverse function of  $f$  they are not the same thing. Notably, the pre-image under  $f$  always exists while the inverse function need not exist.

**Lemma 5.1.3.** For a collection of subsets  $B_i \in \mathcal{F}$  for  $i$  in some indexing set  $\mathcal{I}$  we have,

$$f^{-1}\left(\bigcup_{i \in \mathcal{I}} B_i\right) = \bigcup_{i \in \mathcal{I}} f^{-1}(B_i).$$

### 5.2 DRVs and their distributions

**Definition 5.2.1** (Discrete random variable). A **discrete random variable** (DRV) on the probability space  $(\Omega, \mathcal{F}, P)$  is a function  $X : \Omega \rightarrow \mathbb{R}$  that satisfies the following properties:

- $\text{im } X = \{X(\omega) : \omega \in \Omega\}$  must be a countable subset of  $\mathbb{R}$ ,
- $X^{-1}(x) \in \mathcal{F}$  for all  $x \in \mathbb{R}$ .

**Remark 5.2.2.** The nomenclature of  $X$  being discrete stems from the fact that its image is a countable subset of  $\mathbb{R}$  and so can be mapped to  $\mathbb{N}$  which we see as being discrete.

**Definition 5.2.3** (Probability mass function). The **probability mass function** (pmf) of a DRV  $X$  is defined as a function  $p_X : \mathbb{R} \rightarrow [0, 1]$  such that,

$$p_X(x) := P(X^{-1}(x)).$$

This is commonly denoted by  $p_X(x) = P(X = x)$ .

**Remark 5.2.4.** Some useful properties of the pmf extending from the definition are:

- $x \notin \text{im } X \implies p_X(x) = 0$ ,
- For  $x_1, x_2 \in \text{im } X$  with  $x_1 \neq x_2$ ,  $X^{-1}(x_1) \cap X^{-1}(x_2) = \emptyset$ ,
- $\sum_{x \in \text{im } X} p_X(x) = \sum_{x \in \mathbb{R}} p_X(x) = 1$ .

**Theorem 5.2.5.** Suppose  $\mathcal{I}$  is some indexing set and  $S = \{s_i \in \mathbb{R} : i \in \mathcal{I}\}$  is countable and  $\{\pi_i : i \in \mathcal{I}\}$  is a collection such that  $\pi_i \geq 0$  for all  $i \in \mathcal{I}$  and  $\sum_{i \in \mathcal{I}} \pi_i = 1$ . Then there exists some probability space  $(\Omega, \mathcal{F}, P)$  and a DRV  $X$  on said probability space such that  $p_X(s_i) = \pi_i$  for all  $i \in \mathcal{I}$  and  $p_X(s) = 0$  otherwise.

Lecture 11  
Tuesday  
21/11/2023



## 6 Common DRV's

All DRV's within this section will be considered over the probability space  $(\Omega, \mathcal{F}, P)$ .

### 6.1 Bernoulli distribution

**Definition 6.1.1** (Bernoulli distribution). A DRV  $X$  is said to have **Bernoulli distribution** with parameter  $p \in (0, 1)$  if  $\text{im } X = \{0, 1\}$  with  $p_X(1) = p$ . This is denoted by  $X \sim \text{Bern}(p)$ .

**Definition 6.1.2** (Indicator variable). Given some event  $A \in \mathcal{F}$  the **indicator variable** of the event  $A$  is given by,

$$\mathbb{I}_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}.$$

**Remark 6.1.3.**  $\mathbb{I}_A \sim \text{Bern}(P(A))$ .

### 6.2 Binomial distribution

**Definition 6.2.1** (Binomial distribution). Consider a sequence of  $n \in \mathbb{N}$  iid Bernoulli trials with parameter  $p$ , count the number of successes and denote this by the random variable  $X$  then  $\text{im } X = [0, n]$  and,

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x \in [0, n].$$

We say  $X$  follows a **binomial distribution** and this is denoted by  $X \sim \text{Bin}(n, p)$ .

### 6.3 Hypergeometric distribution

As we have done previously, consider of urn of  $N \in \mathbb{N}$  balls with  $K \in \mathbb{N}$  of these being white and the remainder being black from which we will draw  $n \in \mathbb{N}$  balls and want to consider the DRV ( $X$ ) for the number of white balls drawn. When drawing with replacement we have  $X \sim \text{Bin}(n, K/N)$ . However, when drawing without replacement  $X$  follows the hypergeometric distribution.

**Definition 6.3.1** (Hypergeometric distribution). A DRV  $X$  follows the **hypergeometric distribution** with three parameters  $N \in \mathbb{N}_0, K \in \mathbb{N}, n \in [0, N]$  if  $\text{im } X = [0, \min(n, K)]$  and,

$$p_X(x) = \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}} \quad \text{for } x \in [0, K].$$

**Lemma 6.3.2** (Vandemonde's identity). **Vandemonde's identity** is an important tool in the derivation of the pmf for the hypergeometric distribution and so is included here. The identity is as follows, for  $k, m, n \in \mathbb{N}$  with  $k \leq m + n$ , we have:

$$\binom{m+n}{k} = \sum_{i=0}^k \binom{m}{i} \binom{n}{k-i}.$$

### 6.4 Discrete uniform distribution

**Definition 6.4.1** (Discrete uniform distribution). A DRV  $X$  follows the **discrete uniform distribution** over a nonempty set of numbers  $C$ , denoted  $X \sim \text{DUnif}(C)$ , if  $\text{im } X = C$  and,

$$p_X(x) = \begin{cases} \frac{1}{\text{card}(C)} & \text{for } x \in C \\ 0 & \text{otherwise} \end{cases}.$$

### 6.5 Poisson distribution

The poisson distribution is commonly used for modelling the number of events occurring in a certain time period. Its pdf is derived by taking the  $\lim_{n \rightarrow \infty} p_X(x)$  where  $X \sim \text{Bin}(n, \frac{\lambda}{n})$  for some  $\lambda \in \mathbb{R}$ .

**Definition 6.5.1** (Poisson distribution). A DRV  $X$  follows the **poisson distribution** with parameter  $\lambda \in \mathbb{R}^{>0}$ , denoted  $X \sim \text{Poi}(\lambda)$ , if  $\text{im } X = \mathbb{N}_0$  and,

$$p_X(x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad \text{for } x \in \mathbb{N}_0.$$

## 6.6 Geometric distribution

**Definition 6.6.1** (Geometric distribution). A DRV  $X$  follows the **geometric distribution** with parameter  $p \in (0, 1)$ , denoted  $X \sim \text{Geom}(p)$ , if  $\text{im } X = \mathbb{N}$  and,

$$p_X(x) = (1 - p)^x p \quad \text{for } x \in \mathbb{N}.$$

This can be seen as counting the number of Bernoulli trials with parameter  $p$  that occur before a failure.

## 6.7 Negative binomial distribution

**Definition 6.7.1** (Generalised binomial coefficient). Let  $\alpha \in \mathbb{C}$  and  $k \in \mathbb{N}$  and define the **generalised binomial coefficient** by,

$$\binom{\alpha}{k} := \frac{\alpha(\alpha - 1) \dots (\alpha - k + 1)}{k!}.$$

**Lemma 6.7.2.** For  $x \in \mathbb{N}_0$  and  $r \in \mathbb{N}$  the following identity holds,

$$\binom{x + r - 1}{r - 1} = (-1)^x \binom{-r}{x}.$$

The generalised binomial coefficient as well as this lemma are necessary to have a well defined and valid pdf for the negative binomial distribution.

**Definition 6.7.3** (Negative binomial distribution). A DRV  $X$  follows the **negative binomial distribution** with parameters  $r \in \mathbb{N}$  and  $p \in (0, 1)$ , denoted  $X \sim \text{NBin}(r, p)$ , if  $\text{im } X = \mathbb{N}_0$  and,

$$p_X(x) = \binom{x + r - 1}{r - 1} p^r (1 - p)^x \quad \text{for } x \in \mathbb{N}_0.$$

This is the distribution of the number of failed ii Bernoulli trials with parameter  $p$  before  $r$  successes have occurred.

# 7 Continuous random variables

## 7.1 General random variables and their distributions

**Definition 7.1.1** (Random variable). A **random variable (RV)** on the probability space  $(\Omega, \mathcal{F}, P)$  is a mapping  $X : \Omega \rightarrow \mathbb{R}$  such that  $X^{-1}((-\infty, x]) = \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$  for all  $x \in \mathbb{R}$ . By taking the countable union of pre-images of all  $\omega \leq x$  in  $\mathcal{F}$ , it can be seen that a DRV satisfies this condition.

**Definition 7.1.2** (Cumulative distribution function). For some RV  $X$  on the probability space  $(\Omega, \mathcal{F}, P)$ , the **cumulative distribution function (CDF)** of  $X$  is defined as the mapping  $F_X : \mathbb{R} \rightarrow [0, 1]$  given by,

$$F_X(x) = P(X^{-1}((-\infty, x])),$$

often denoted  $F_X(x) = P(X \leq x)$ .

**Theorem 7.1.3** (cdf properties). For some RV  $X$  on the probability space  $(\Omega, \mathcal{F}, P)$  the following properties hold:

1.  $F_X$  is monotonically non-decreasing,
2.  $F_X$  is right-continuous ( $(x_n) \downarrow x \implies F_X(x_n) \rightarrow F_X(x)$  as  $n \rightarrow \infty$ ),
3.  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  and  $\lim_{x \rightarrow \infty} F_X(x) = 1$ .

**Theorem 7.1.4.** For  $a, b \in \mathbb{R}$  if  $a < b$ , then  $P(a < X \leq b) = F_X(b) - F_X(a)$ .

**Remark 7.1.5.** In general the cdf of an RV is not left continuous.

## 7.2 CRVs and pdfs

**Definition 7.2.1** (Continuous random variable). A random variable  $X$  on the probability space  $(\Omega, \mathcal{F}, P)$  is called a **continuous random variable (CRV)** iff its cdf can be written as:

$$F_X(x) = \int_{-\infty}^x f_X(u) du \quad \text{for all } x \in \mathbb{R},$$

where  $f_X : \mathbb{R} \rightarrow \mathbb{R}$  satisfies:  $f_X(u) \geq 0$  for all  $u \in \mathbb{R}$  and  $\int_{-\infty}^{\infty} f_X(u) du = 1$ . We call  $f_X$  the **probability density function (pdf)** of  $X$ .

**Theorem 7.2.2.** If  $X$  is a CRV on the probability space  $(\Omega, \mathcal{F}, P)$  with pdf  $f_X$ ,  $P(X = x) = 0$  for all  $x \in \mathbb{R}$ .

**Theorem 7.2.3.** With the same conditions,  $P(a \leq X \leq b) = \int_a^b f_X(u) du$  for all  $a, b \in \mathbb{R}$  with  $a \leq b$ .

**Remark 7.2.4.** Combining the results from this section leads to the conclusion that the cdf of a CRV is continuous.

## 8 Common CRVs

All CRVs  $X$  within this section will be considered over the probability space  $(\Omega, \mathcal{F}, P)$  with the natural notation for their pdf and cdf. These distributions will be uniquely identified by their pdfs.

### 8.1 Uniform distribution

**Definition 8.1.1** (Uniform distribution). A CRV  $X$  follows the **uniform distribution** on the interval  $(a, b)$  for  $a, b \in \mathbb{R}$  with  $a < b$ , denoted  $X \sim U(a, b)$  if it satisfies:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases}, \quad F_X(x) = \begin{cases} 0 & \text{if } x \leq a \\ \frac{x-a}{b-a} & \text{if } a < x < b \\ 1 & \text{if } x \geq b \end{cases}.$$

### 8.2 Exponential distribution

**Definition 8.2.1** (Exponential distribution). A CRV  $X$  follows the **exponential distribution** with parameter  $\lambda \in \mathbb{R}^{>0}$ , denoted  $X \sim \text{Exp}(\lambda)$  if it satisfies:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}, \quad F_X(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 - e^{-\lambda x} & \text{if } x > 0 \end{cases}.$$

### 8.3 Gamma distribution

**Definition 8.3.1** (Gamma function). For  $t \in \mathbb{R}$  with  $t > 0$  we define the **gamma function** by,

$$\Gamma(t) := \int_0^{\infty} x^{t-1} e^{-x} dx.$$

One of the gamma function's many interesting properties is that  $\Gamma(t) = (t-1)\Gamma(t-1)$  for  $t > 1$ .

**Definition 8.3.2** (Gamma distribution). A CRV  $X$  follows the **gamma distribution** with shape and rate parameter  $\alpha, \beta \in \mathbb{R}^{>0}$  respectively, denoted  $X \sim \text{Gamma}(\alpha, \beta)$  if it satisfies:

$$f_X(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}.$$

Its cdf cannot be written in a closed form so must be left as an integral of the pdf or approximated.

## 8.4 Chi-squared distribution

**Definition 8.4.1** (Chi-squared distribution). A CRV  $X$  follows the **chi-squared distribution** with  $n \in \mathbb{N}$  degrees of freedom, denoted  $X \sim \chi^2(n)$  if it satisfies:

$$f_X(x) = \begin{cases} \frac{1}{2\Gamma(n/2)} \left(\frac{x}{2}\right)^{n/2-1} e^{-x/2} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}.$$

Its cdf can also not be written in a closed form. The  $\chi^2(n)$  distribution is the same as the **Gamma** $(\frac{n}{2}, \frac{1}{2})$  distribution.

## 8.5 F-distribution

These pdfs are getting tough.

**Definition 8.5.1** (F-distribution). A CRV  $X$  follows the **f-distribution** with  $d_1, d_2 \in \mathbb{R}^{>0}$  degrees of freedom, denoted  $X \sim F(d_1, d_2)$  if it satisfies:

$$f_X(x) = \begin{cases} \frac{\Gamma\left(\frac{d_1+d_2}{2}\right) \left(\frac{d_1}{d_2}\right)^{d_1/2} x^{d_1/2-1}}{\Gamma\left(\frac{d_1}{2}\right) \Gamma\left(\frac{d_2}{2}\right) \left(1 + \frac{d_1}{d_2}x\right)^{(d_1+d_2)/2}} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}.$$

Its cdf can also not be written in a closed form. It is important to note that  $d_1, d_2$  are not restricted to integer values, and that  $X = \frac{X_1/d_1}{X_2/d_2}$  where  $X_1 \sim \chi^2(d_1)$  and  $X_2 \sim \chi^2(d_2)$ .

## 8.6 Beta distribution

**Definition 8.6.1** (Beta function). For  $\alpha, \beta \in \mathbb{R}^{>0}$  we define the **beta function** by,

$$B(\alpha, \beta) := \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

**Definition 8.6.2** (Beta distribution). A CRV  $X$  follows the **beta distribution** with parameters  $\alpha, \beta \in \mathbb{R}^{>0}$ , denoted  $X \sim \text{Beta}(\alpha, \beta)$  if it satisfies:

$$f_X(x) = \begin{cases} \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

Its cdf can also not be written in a closed form.

## 8.7 Normal distribution

**Definition 8.7.1** (Standard normal distribution). A CRV  $X$  follows the **standard normal distribution** or **Gaussian distribution**, denoted  $X \sim N(0, 1)$  if it satisfies,

$$f_X(x) = \phi(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{for } x \in \mathbb{R}, \quad F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \quad \text{for } x \in \mathbb{R}.$$

Where, once again, there is no explicit formula for the cdf.

**Definition 8.7.2** (Normal distribution). A CRV  $X$  follows the **normal distribution** with mean  $\mu \in \mathbb{R}$  and variance  $\sigma^2$  for  $\sigma \in \mathbb{R}^{>0}$  denoted  $X \sim N(\mu, \sigma^2)$  if it satisfies,

$$f_X(x) = \phi(x) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{for } x \in \mathbb{R}.$$

## 8.8 Cauchy distribution

**Definition 8.8.1** (Cauchy distribution). A CRV  $X$  follows the **Cauchy distribution** if it satisfies,

$$f_X(x) = \frac{1}{\pi(1+x^2)} \quad \text{for } x \in \mathbb{R}, \quad F_X(x) = \frac{1}{\pi} \arctan(x) + \frac{1}{2} \quad \text{for } x \in \mathbb{R}.$$

If  $X, Y \sim N(0, 1)$ , then  $Z = X/Y$  follows the Cauchy distribution.

## 8.9 Student t-distribution

**Definition 8.9.1** ((Student's) t-distribution). A CRV  $X$  follows the **Student t-distribution** with  $\nu \in \mathbb{R}^{>0}$  degrees of freedom if it satisfies,

$$f_X(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \quad \text{for } x \in \mathbb{R}.$$

Its cdf cannot be written in a closed form.

**Remark 8.9.2.** Not all RVs are either discrete or continuous.

## 9 Transformations of random variables

### 9.1 DRVs

**Theorem 9.1.1.** Let  $X$  be a DRV on  $(\Omega, \mathcal{F}, P)$  and let  $g : \mathbb{R} \rightarrow \mathbb{R}$  be a deterministic function, then  $Y = g(X)$  is a DRV with pmf:

$$p_Y(y) = \sum_{\{x \in \text{im } X : g(x)=y\}} p_X(x) \quad \text{if } y \in \text{im } Y \text{ and } 0 \text{ otherwise.}$$

### 9.2 CRVs

**Theorem 9.2.1.** Let  $X$  be a CRV on  $(\Omega, \mathcal{F}, P)$  and let  $g : \mathbb{R} \rightarrow \mathbb{R}$  be a strictly monotonic and differentiable function with inverse  $g^{-1} : \mathbb{R} \rightarrow \mathbb{R}$ , then  $Y = g(X)$  is a CRV with pdf:

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} [g^{-1}(y)] \right| \quad \text{for all } y \in \mathbb{R}.$$

**Remark 9.2.2.** The term  $\left| \frac{d}{dy} [g^{-1}(y)] \right|$  is often called the **Jacobian** of the transformation.

## 10 Expectation of random variables

Throughout this section, unless otherwise specified, all infinite sums will be assumed to converge absolutely and all integrals will be assumed to be  $< \infty$ .

### 10.1 Definition

**Definition 10.1.1** (Expectation of a DRV). Let  $X$  be a DRV on  $(\Omega, \mathcal{F}, P)$  then the **expectation** of  $X$  is defined by,

$$E(X) := \sum_{x \in \text{im } X} xp_X(x).$$

**Definition 10.1.2** (Expectation of a CRV). Let  $X$  be a CRV on  $(\Omega, \mathcal{F}, P)$  then the **expectation** of  $X$  is defined by,

$$E(X) := \int_{-\infty}^{\infty} xf_X(x)dx.$$

### 10.2 LOTUS

**Theorem 10.2.1** (Discrete LOTUS). Let  $X$  be a DRV on  $(\Omega, \mathcal{F}, P)$  and  $g : \mathbb{R} \rightarrow \mathbb{R}$ , we have,

$$E(g(X)) = \sum_{x \in \text{im } X} g(x)p_X(x).$$

**Theorem 10.2.2** (Continuous LOTUS). Let  $X$  be a CRV on  $(\Omega, \mathcal{F}, P)$  and  $g : \mathbb{R} \rightarrow \mathbb{R}$ , we have,

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

Note that this is one of the few theorems throughout the course given without proof.

**Theorem 10.2.3.** If  $X$  is non-negative then  $E(X) \geq 0$ .

### 10.3 Variance

**Definition 10.3.1** (Variance). Let  $X$  be a discrete/continuous random variable, then the **variance** of  $X$  is defined by,

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

**Theorem 10.3.2.** For a discrete/continuous random variable with finite variance,

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

## 11 Multivariate random variables

### 11.1 Multivariate distributions

**Definition 11.1.1** (Joint distribution function). Consider the sequence of random variables  $X_1, X_2, \dots, X_n$  all on  $(\Omega, \mathcal{F}, P)$  and write  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  and  $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ . Then the **joint distribution function** of  $\mathbf{X}$  is  $F_{\mathbf{X}} : \mathbb{R}^n \rightarrow [0, 1]$  defined by:

$$F_{\mathbf{X}}(\mathbf{x}) := P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) \quad \text{for all } \mathbf{x} \in \mathbb{R}^n.$$

### 11.2 Independence

**Definition 11.2.1** (Pairwise independence for  $n$  random variables). We call the sequence of RVs,  $X_1, X_2, \dots, X_n$ , **pairwise independent** iff,

$$F_{X_i, X_j}(x_i, x_j) = F_{X_i}(x_i)F_{X_j}(x_j) \quad \text{for all } x_i, x_j \in \mathbb{R} \text{ with } i \neq j.$$

**Definition 11.2.2** (Independence of a family of random variables). Given some indexing set  $\mathcal{I} \subset \mathbb{R}$ , a family of random variables  $\{X_i : i \in \mathcal{I}\}$  is **independent** iff for all finite  $\mathcal{J} \subseteq \mathcal{I}$ :

$$P\left(\bigcap_{j \in \mathcal{J}} \{X_j \leq x_j\}\right) = \prod_{j \in \mathcal{J}} P(\{X_j \leq x_j\}) \quad \text{for all } (x_j)_{j \in \mathcal{J}} \text{ with } x_j \in \mathbb{R}.$$

(All finite subfamilies of the family of random variables is independent by the natural definition)

### 11.3 Multivariate DRVs

**Definition 11.3.1** (Joint probability mass functions). Let  $X_1, X_2, \dots, X_n$  all be DRVs on  $(\Omega, \mathcal{F}, P)$  that form the random vector  $\mathbf{X}$ , their **joint probability mass function**,  $p_{\mathbf{X}} : \mathbb{R}^n \rightarrow [0, 1]$  is defined as,

$$p_{\mathbf{X}}(x_1, x_2, \dots, x_n) := P(\{\omega \in \Omega : X_1(\omega) = x_1, X_2(\omega) = x_2, \dots, X_n(\omega) = x_n\}) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n).$$

The **marginal probability mass function** of  $X_i \in \mathbf{X}$  is given by,

$$p_{X_i}(k) = \sum_{(x_1, x_2, \dots, x_n) \in \mathbb{R}^n} \cdots \sum_{(x_1, x_2, \dots, x_n) \in \mathbb{R}^n} p_{\mathbf{X}}(x_1, x_2, \dots, x_{i-1}, k, x_{i+1}, \dots, x_n).$$

It can be obtained that for any sufficiently "nice" set  $A \in \mathbb{R}^n$ ,

$$P(\mathbf{X} \in A) = \sum_{(x_1, x_2, \dots, x_n) \in A} \cdots \sum_{(x_1, x_2, \dots, x_n) \in A} P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n).$$

**Definition 11.3.2** (Independence of DRVs). Given some indexing set  $\mathcal{I} \in \mathbb{R}$  a family of DRVs,  $\{X_i : i \in \mathcal{I}\}$  with joint pmf  $p_{\mathbf{X}}$ , is **independent** iff for all finite  $\mathcal{J} \subseteq \mathcal{I}$ :

$$p_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n p_{X_i}(x_i) \quad \text{for all } \mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n.$$

## 11.4 Multivariate CRVs

**Definition 11.4.1** (Continuous random vector). The random vector  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  is a **continuous random vector** iff,

$$F_{\mathbf{X}}(\mathbf{x}) = \int \cdots \int_B f_{\mathbf{X}}(\mathbf{x}) dx_1 dx_2 \cdots dx_n \quad \text{with } B = (\infty, x_1] \times (\infty, x_2] \times \cdots \times (\infty, x_n], \quad \text{for all } \mathbf{x} \in \mathbb{R}^n;$$

for some  $f_{\mathbf{X}} : \mathbb{R}^n \rightarrow \mathbb{R}$  satisfying:  $f_{\mathbf{X}}(\mathbf{x}) \geq 0$  for all  $\mathbf{x} \in \mathbb{R}^n$  and  $\int \cdots \int_{\mathbb{R}^n} f_{\mathbf{X}}(\mathbf{x}) dx_1 dx_2 \cdots dx_n = 1$ .

Note that  $f_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} F_{\mathbf{X}}(\mathbf{x})$  and  $P(\mathbf{X} \in A) = \int \cdots \int_A f_{\mathbf{X}}(\mathbf{x}) d^n \mathbf{x}$ .

**Definition 11.4.2** (Independence of CRVs). Given some indexing set  $\mathcal{I} \in \mathbb{R}$  a family of CRVs,  $\{X_i : i \in \mathcal{I}\}$  with joint pdf  $f_{\mathbf{X}}$ , is **independent** iff for all finite  $\mathcal{J} \in \mathcal{I}$ :

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n f_{X_i}(x_i) \quad \text{for all } \mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n.$$

## 11.5 Transformations of random vector

**Definition 11.5.1** (Transformation). We are going to **transform** the random vector  $\mathbf{X}$  with joint pdf  $f_{\mathbf{X}}$  to  $\mathbf{U} = (u_1(\mathbf{X}), u_2(\mathbf{X}), \dots, u_n(\mathbf{X}))$  with  $u_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$  for all  $i \in [1, n]$ . Firstly, define  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  by  $T(\mathbf{x}) = (u_1(\mathbf{x}), u_2(\mathbf{x}), \dots, u_n(\mathbf{x}))$  and assume  $T$  is bijective on  $D = \{\mathbf{x} \in \mathbb{R}^n : f_{\mathbf{X}}(\mathbf{x}) > 0\}$  with range  $S \subseteq \mathbb{R}^n$ . Secondly, have the Jacobian determinant of  $T^{-1} : S \rightarrow D$ ,  $J(\mathbf{u}) = \det([a_{ij}]_{m \times n})$  with  $a_{ij} = \frac{\partial x_i}{\partial u_j}$ . Finally, define:

$$f_{\mathbf{U}}(\mathbf{u}) := \begin{cases} f_{\mathbf{X}}(T^{-1}(\mathbf{u})) |J(\mathbf{u})| & \text{if } \mathbf{u} \in S \\ 0 & \text{otherwise} \end{cases}.$$

## 11.6 Multivariate LOTUS

**Theorem 11.6.1** (Discrete multivariate LOTUS). If  $X_1, X_2, \dots, X_n$  are DRVs on  $(\Omega, \mathcal{F}, P)$  and form the random vector  $\mathbf{X}$  with  $g : \mathbb{R}^n \rightarrow \mathbb{R}$ , then  $Y = g(\mathbf{X})$  is a DRV on  $(\Omega, \mathcal{F}, P)$  with expectation,

$$E(g(\mathbf{X})) = \sum_{x_i \in \text{im } X_i} \cdots \sum g(\mathbf{X}) P(\mathbf{X} = \mathbf{x}) \quad \text{for all } \mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n.$$

**Theorem 11.6.2** (Continuous multivariate LOTUS). If  $X_1, X_2, \dots, X_n$  are DRVs on  $(\Omega, \mathcal{F}, P)$  and form the random vector  $\mathbf{X}$  with  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  we have,

$$E(h(\mathbf{X})) = \int \cdots \int_{\mathbb{R}^n} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) dx_1 dx_2 \cdots dx_n \quad \text{for all } \mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n.$$

## 11.7 Covariance

**Definition 11.7.1** (Covariance). Given two random variable  $X$  and  $Y$  on the same probability space with expectations  $\mu_X$  and  $\mu_Y$  respectively. The **covariance** of  $X$  and  $Y$  is defined as,

$$\text{Cov}(X, Y) := E[(X - \mu_X)(Y - \mu_Y)] \quad \text{assuming both expectation take finite values.}$$

**Definition 11.7.2** (Correlation). Given the same  $X$  and  $Y$  the **correlation** of  $X$  and  $Y$  is defined as,

$$\text{Cor}(X, Y) := \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

**Theorem 11.7.3.** For jointly discrete/continuous RVs with finite expectations the following hold:

1. when  $X = Y$ ,  $\text{Cov}(X, Y) = E[(X - \mu_X)^2] = \text{Var}(X)$ ,
2.  $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$ ,
3. when  $X$  and  $Y$  are independent,  $E(XY) = E(X)E(Y)$ ,
4. when variances are also finite,  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$ .

## 12 Generating functions

### 12.1 Probability generating functions

**Definition 12.1.1** (Probability generating functions). Given a DRV  $X$  with  $\text{im}(X) \subseteq \mathbb{N}_0$ , denote,

$$\mathcal{S}_X := \left\{ s \in \mathbb{R} : \sum_{x=0}^{\infty} |s|^x \mathbb{P}(X = x) < \infty \right\},$$

and define the **probability generating function (pgf)** of  $X$  as the function  $G_X : \mathcal{S}_X \rightarrow \mathbb{R}$  given by,

$$G_X(s) := \mathbb{E}(s^X) = \sum_{x=0}^{\infty} s^x \mathbb{P}(X = x),$$

noting that the pgf is well defined for  $|s| < 1$  and  $G_X(0) = \mathbb{P}(X = 0)$  and  $G_X(1) = 1$ .

**Theorem 12.1.2** (Uniqueness of pgfs). Given two DRVs  $X$  and  $Y$  with pgfs  $G_X$  and  $G_Y$  respectively,

$$G_X(s) = G_Y(s) \quad \text{for all } s \in \mathcal{S}_X \cap \mathcal{S}_Y \iff p_X(x) = p_Y(x) \quad \text{for all } x \in \mathbb{N}_0.$$

**Theorem 12.1.3.** Let  $X, Y$  be independent DRVs with  $\text{im } X, \text{im } Y \in \mathbb{N}_0$ , then

$$G_{X+Y}(s) = G_X(s)G_Y(s) \quad \text{for all } s \in \mathcal{S}_X \cap \mathcal{S}_Y.$$

**Theorem 12.1.4** (Pgfs of sum of independent DRVs). Given a collection of  $n$  independent DRVs  $X_1, X_2, \dots, X_n$ ,

$$G_{\sum_{i=1}^n X_i}(s) = \prod_{i=1}^n G_{X_i}(s) \quad \text{for all } s \in \bigcap_{i=1}^n \mathcal{S}_{X_i}.$$

**Theorem 12.1.5** (Moments). Given a DRV  $X$  with  $\text{im } X \subseteq \mathbb{N}_0$ , the  $k$ th derivative of  $G_X$ , for  $k \in \mathbb{N}$  is given by,

$$\left. \frac{d^k}{ds^k} G_X(s) \right|_{s=1} = G_X^{(k)}(1) = \mathbb{E}[X(X-1)\dots(X-k+1)].$$

### 12.2 Common pgfs

**Example 12.2.1** (Bernoulli distribution). Let  $X \sim \text{Bern}(p)$ , then  $G_X(s) = 1 - p + sp$  for all  $s \in \mathbb{R}$ .

**Example 12.2.2** (Binomial distribution). Let  $X \sim \text{Bin}(n, p)$ , then  $G_X(s) = (1 - p + sp)^n$  for all  $s \in \mathbb{R}$ .

**Example 12.2.3** (Poisson distribution). Let  $X \sim \text{Poi}(\lambda)$ , then  $G_X(s) = \exp(\lambda(s-1))$  for all  $s \in \mathbb{R}$ .

### 12.3 Moment generating functions

**Definition 12.3.1** (Moment generating function). Let  $X$  be a RV, its **moment generating function (mgf)** is defined as,

$$M_X(t) = \mathbb{E}(e^{tX}),$$

assuming the expectation exists in some neighbourhood of zero.

**Remark 12.3.2.** If  $X$  is a RV with a mgf,  $M_X(t) = \mathbb{E}(e^{tX}) = G_X(e^t)$

**Theorem 12.3.3.** If  $X$  is a RV with a mgf, the  $k$ th moment of  $X$  is  $\mathbb{E}(X^k) = M_X^{(k)}(0)$ .

**Theorem 12.3.4.** If  $X_1, X_2, \dots, X_n$  are a family of independent RVs on the same probability space with mgfs  $M_{X_1}, M_{X_2}, \dots, M_{X_n}$  respectively, we have,

$$M_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n M_{X_i}(t).$$

**Theorem 12.3.5** (Characterisation by mgf). If the RVs  $X, Y$  have existent mgfs  $M_X, M_Y$  respectively such that  $M_X(t) = M_Y(t)$  for all  $t$  in some neighbourhood of 0, we have,

$$F_X(u) = F_Y(u) \quad \text{for all } u.$$



## 13 Conditional distribution and expectation

### 13.1 Discrete: Conditional expectation and total expectation

**Definition 13.1.1** (Condition distribution and expectation of a DRV). Given a DRV  $Y$  on  $(\Omega, \mathcal{F}, P)$  and some event  $B \in \mathcal{F}$  with  $P(B) > 0$ , the **conditional distribution** of  $Y$  given  $B$  is defined as,

$$P(Y = y|B) := \frac{P(\{Y = y\} \cap B)}{P(B)} \quad \text{for } y \in \mathbb{R};$$

with the **conditional expectation** of  $Y$  given  $B$  defined as,

$$E(Y|B) := \sum_{i \in \text{im } Y} eP(Y = y|B).$$

**Theorem 13.1.2** (Discrete law of total expectation). Given a DRV  $Y$  on  $(\Omega, \mathcal{F}, P)$  and some partition  $\{B_i : i \in \mathcal{I}\}$  of  $\Omega$  with  $P(B_i) > 0$  for all  $i \in \mathcal{I}$  we have,

$$E(Y) = \sum_{i \in \mathcal{I}} E(Y|B_i)P(B_i).$$

### 13.2 Conditioning on a DRV

**Definition 13.2.1** (Conditional probability mass function). Given two joint DRVs  $(X, Y)$ , the **conditional probability mass function** of  $Y$  given  $X = x$  is given by,

$$p_{Y|X}(y|x) := \frac{p_{X,Y}(x, y)}{p_X(x)} \quad \text{for } y \in \mathbb{R}.$$

**Theorem 13.2.2** (Conditional expectation). Given two joint DRVs  $(X, Y)$ , the **conditional expectation** of  $Y$  given  $X = x$  is given by,

$$E(Y|X = x) = \sum_{y \in \text{im } Y} yp_{Y|X}(y|x).$$

Independence, LOTUS and a Bayes' rule formulation all follow naturally from this as they do for the non-distribution settings.

### 13.3 Continuous: Conditional density, distribution and expectation

**Definition 13.3.1** (Conditional distribution and conditional density). For two joint CRVs  $(X, Y)$  the **conditional density** of  $Y$  given  $X = x$  is define as,

$$f_{Y|X}(y|x) := \frac{f_{X,Y}(x, y)}{f_X(x)} \quad \text{for all } y, x \in \mathbb{R} \text{ with } f_X(x) > 0;$$

with the corresponding **conditional distribution** of  $Y$  given  $X = x$  defined as,

$$F_{Y|X=x}(y|x) := \frac{1}{f_X(x)} \int_{-\infty}^y f_{X,Y}(x, t) dt \quad \text{for all } y, x \in \mathbb{R} \text{ with } f_X(x) > 0.$$

Where, once again, familiar formulations for independence and Bayes' rule can be easily derived.

**Definition 13.3.2** (Conditional expectation). Given two joint CRVs  $(X, Y)$ , the **conditional expectation** of  $Y$  given  $X = x$  is defined as,

$$E(Y|X = x) := \int_{-\infty}^{\infty} yf_{Y|X}(y|x)dy \quad \text{provided } f_X(x) > 0.$$

**Theorem 13.3.3** (Continuous law of total expectation). Given two joint CRVs  $(X, Y)$  with  $E(|Y|) < \infty$ , we have,

$$E(Y) = \int_{\{x: f_X(x) > 0\}} E(Y|X = x)f_X(x)dx.$$

# Chapter 7

## Statistics

### 1 Introduction

The following are references.

- E Artin, Galois theory, 1994
- A Grothendieck and M Raynaud, Revêtements étales et groupe fondamental, 2002
- I N Herstein, Topics in algebra, 1975
- M Reid, Galois theory, 2014

**Notation.** If  $K$  is a field, or a ring, I denote the **ring of polynomials** with coefficients in  $K$ .

Lecture 1  
Thursday  
10/01/19

## 2 Central tendency and dispersion

- 2.1 Mean, variance and moments
- 2.2 Parameter estimation
- 2.3 Other measures of central tendency
- 2.4 Sampling from normal RVs

## 3 Hypothesis testing

- 3.1 Introduction
- 3.2 Single sample hypothesis testing
- 3.3 Distribution of p-values
- 3.4 Errors
- 3.5 Two sample hypothesis testing
- 3.6 Multiple hypothesis testing

## 4 Covariance and Correlations

- 4.1 Covariance
- 4.2 Correlation

## 5 Statistical models

- 5.1 Definitions
- 5.2 Likelihood
- 5.3 Linear regression

## 6 Bayesian inference

- 6.1 Definitions
- 6.2 Conjugate pair distributions
- 6.3 Intractable posteriors
- 6.4 Choosing a prior

## 7 Bootstrap

- 7.1 Empirical distribution
- 7.2 Bootstrap procedure

# Chapter 8

## Computation

### 1 Introduction

The following are references.

- E Artin, Galois theory, 1994
- A Grothendieck and M Raynaud, Revêtements étales et groupe fondamental, 2002
- I N Herstein, Topics in algebra, 1975
- M Reid, Galois theory, 2014

**Notation.** If  $K$  is a field, or a ring, I denote the **ring of polynomials** with coefficients in  $K$ .

Lecture 1  
Thursday  
10/01/19

# Chapter 9

## Applied Mathematics

### 1 Introduction

The following are references.

- E Artin, Galois theory, 1994
- A Grothendieck and M Raynaud, Revêtements étales et groupe fondamental, 2002
- I N Herstein, Topics in algebra, 1975
- M Reid, Galois theory, 2014

**Notation.** If  $K$  is a field, or a ring, I denote the **ring of polynomials** with coefficients in  $K$ .

Lecture 1  
Thursday  
10/01/19