Mathematics Notes

March 11, 2025

Contents

1 Logic, Metric Spaces, and Set Theory **1**
1.1 Metric Spaces . . . 8

2 Algebra **11**
2.1 Divisibility in $\mathbb{Z}$ . . . 12
2.2 Congruence and Congruence Classes . . . 15
2.3 Rings . . . 18

3 Linear Algebra **22**

4 Analysis **25**
4.1 Sequences . . . 25
4.2 Series . . . 35
4.3 Continuity . . . 37

5 Probability **39**
5.1 Axioms of Probability . . . 40

6 Advanced Risk and Portfolio Management **46**
6.1 Data Science . . . 46
6.1.1 Probabilistic Framework . . . 46
6.1.2 Mean-Covariance Framework . . . 46
6.1.3 Linear Models . . . 47
6.1.4 Machine Learning . . . 47
6.1.5 Estimation . . . 47
6.1.6 Inference . . . 47
6.1.7 Sequential Decisions . . . 47
6.2 Quantiative Finance . . . 47
6.2.1 Financial Engineering . . . 47
6.2.2 Risk Management . . . 47
6.2.3 Portfolio Management . . . 47

1 Logic, Metric Spaces, and Set Theory

Why study analysis or mathematics in general? If you intend to reason and navigate the complexities of any system, circumstance, task, or structure, the patterns of reasoning covered in mathematics equips you with the skill of understanding and making inferences or deductions in and about complex systems. So we will study systems at an abstracted level so that our conclusions and hard work are applicable and will aid us in any vocation whether we really notice it or not. Before we begin the rigorous study of calculus, which is the system used to understand and gain insight to abstract dynamic magnitudes. To build this system, we need to first discuss what type of *connections* this systems structure allows.

The first *axiom* of the system is that a *mathematical statement* is either true or false. A mathematical statement is a relationship that is shown through a type of *expression(s)*. An expression is a sequence of mathematical symbols, concepts, and objects that produce some other mathematical object. One can make statements out of expressions by using *relations* such as $=$, $<$, $\geq$, $\in$, $\subset$ or by using *properties* such as "is prime", "is invertible", "is continuous". Then one can make a compound statement from other statements by using *logical connectives*. We show some of these below,

Conjunction: If X is a statement and Y is a statement then the statement "X and Y" is a true statement if X and Y are both true. Notice though that this only concerns truth, where the artist of the mathematics must bring the connotations that illustrate more information that just "X and Y". For example, "X and also Y", or "both X and Y", or even "X but Y". Notice that X but Y suggests that the statements X and Y are in contrast to each other, while X and Y suggests that they support each other. We can find such reinterpretations of every logical connective.

Disjunction: If X is a statement and Y is a statement then the statement "X or Y" is true if either X or Y is true, or both. The reason we include the "X and Y" part is because when we are talking about X or Y we want to be talking about *all of* X or Y, instead of talking about X and not Y or Y and not X. So talking about the *exclusive* "or" (the one that doesn't include "and") is basically talking about two statements.

Negation: The statement "X is not true" or "X is false" is called the *negation* of X and is true if and only if X is false and is false if and only if X is true. Negations convert "and" into "or" and vice versa. For instance, the negation of "Jane Doe has black hair and Jane Doe has blue eyes" is "Jane Doe doesn't have black hair or doesn't have blue eyes". Notice how important the "inclusive or" is here to interpret the meaning of this statement.

If and only if: If X is a statement and Y is a statement, we say that "X is true if and only if Y is true", whenever X is true, Y also has to be true, and whenever Y is true, X must too be true. This is sort of like a logical equivalence. So if we were trying to pin down some type of abstract causal structure of some system an if and only if statement tells me that X and Y will always cause each other.

Implication: If X is a statement and Y is a statement then if we want to know whether (using some abstract notion of "cause") X causes, implies, or leads to Y then we are trying to prove an *implication* which is given by "if X then Y" (the implication of X to Y). So for X to truly *imply* Y, we need that when X is true Y is also true, if X is false then whether Y is true or false doesn't matter. So the only way to disprove an implication is is by showing that when the hypothesis is true, the conclusion is false. One can also think of the statement "if X, then Y " as "Y is at least as true as X"—if X is true, then Y also has to be true, but if X is false, Y could be as false as X, but it could also be true.

Variables and Quantifiers: Notice when we talk about some abstract, general, X and Y, the truth of the statements involving them depends on the context of X and Y. More precisely, X and Y are *variables* since they are variables that are set to obey some properties but the actual value of them hasn't been specified yet. Then *quantifiers* allow us to talk about the different values of these variables. We can say that there exists X where, say, X implies Y is true, this is denoted $\exists$. Or we can say for all X (denoted $\forall$), X implies Y. **Equality:** Out of the different relations we have discussed, *equality* is the most obvious. We need to be able to express the relationship of equality. We will present the axioms of equality, called an *equivalence relation*

Definition 1.1 (Equivalence Relation)**.** Given elements x, y, z in any set with the relation $=$ defined, we have

1. (Reflexivity): Given any object x, we have $x = x$.
2. (Symmetry): Given any two objects x and y of the same type, if $x = y$ then $y = x$
3. (Transitive): Given any three objects x, y, z of the same type, if $x = y$ and $y = z$, then $x = z$.
4. (Substitution): Given any two objects x and y of the same type, if $x = y$, then $f(x) = f(y)$ for all functions or operations f. Similarly, for any property $P(x)$ depending on x, if $x = y$, then $P(x)$ and $P(y)$ are equivalent statements.

Definition 1.2. A *set* is a well-defined collection of distinct objects, called *elements* or *members* considered as a single entity unified under the defining properties of the set. The membership of an element x in a set S is denoted by $x \in S$, while non-membership is written as $x \notin S$. A set containing no elements is called the *empty set*, denoted $\emptyset$.

Proposition 1.1. *Let A, B, C be sets, and let X be a set containing A, B, C as subsets.*

1. *(Minimal element) We have* $A \cup \emptyset = A$ *and* $A \cap \emptyset = \emptyset$
2. *(Maximal element) We have* $A \cup X = X$ *and* $A \cap X = A$.
3. *(Identity) We have* $A \cup A = A$ *and* $A \cap A = A$
4. *(Commutativity) We have* $A \cup B = B \cup A$ *and* $A \cap B = B \cap A$
5. *(Associativity) We have* $(A \cup B) \cup C = A \cup (B \cup C)$ *and* $(A \cap B) \cap C = A \cap (B \cap C)$
6. *(Distributivity) We have* $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ *and* $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
7. *(Partition) We have* $A \cup (X \setminus A) = X$ *and* $A \cap (X \setminus A) = \emptyset$
8. *(De Morgan Laws) We have* $X \setminus (A \cup B) = (X \setminus A) \cap (X \setminus B)$ *and* $X \setminus (A \cap B) = (X \setminus A) \cup (X \setminus B)$

Definition 1.3. An *ordered set* is a set S together with an ordering relation, denoted $<$, such that

1. *(trichotomy)* $\forall x, y \in S,$ exactly one of $x < y, x = y,$ or $y < x$ holds.
2. *(transitivity)* If $x, y, z \in S$ such that $x < y$ and $y < z \implies x < z$.

Well ordering property of $\mathbb{N}$: Every nonempty subset of $\mathbb{N}$ has a least element.

Definition 1.4. We define the natural numbers $\{1, 2, 3, 4, \dots\}$ to be a set $\mathbb{N}$ with the *successor function* S defined on it. The successor function $S : \mathbb{N} \to \mathbb{N}$, is defined by the following axioms,
N1: $1 \in \mathbb{N}$
N2: If $n \in \mathbb{N}$ then its successor $n + 1 \in \mathbb{N}$
N3: 1 is not the successor of any element in $\mathbb{N}$
N4: If n and m in $\mathbb{N}$ have the same successor, then $n = m$.
N5: A subset of $\mathbb{N}$ that contains 1, and contains $n + 1$ whenever it contains n, must be equivalent to $\mathbb{N}$.

Theorem 1.1 (Principle of induction)**.** *Let* $P(n)$ *be a statement depending on a natural number* n. *Suppose that*

(i) (basis statement) $P(1)$ *is true.*

(ii) (induction step) If $P(n)$ *is true, then* $P(n + 1)$ *is true.*

Then $P(n)$ *is true for all* $n \in \mathbb{N}$.

Proof. Let S be the set of natural numbers n for which $P(n)$ is not true. Suppose for contradiction that S is nonempty. Then S has a least element by the well-ordering property. Call $m \in S$ the least element of S. We know $1 \notin S$ by hypothesis. So $m > 1$, and $m - 1$ is a natural number as well. Since m is the least element of S, we know that $P(m - 1)$ is true. But the induction step says that $P(m - 1 + 1) = P(m)$ is true, contradicting the statement that $m \in S$. Therefore, S is empty and $P(n)$ is true for all $n \in \mathbb{N}$.

$\square$

Definition 1.5. A set F is called a *field* if it has two operations defined on it, addition $x + y$ and multiplication xy, and if it satisfies the following axioms:

(A1) If $x \in F$ and $y \in F$, then $x + y \in F$.

(A2) *(commutativity of addition)* $x + y = y + x$ for all $x, y \in F$.

(A3) *(associativity of addition)* $(x + y) + z = x + (y + z)$ for all $x, y, z \in F$.

(A4) There exists an element $0 \in F$ such that $0 + x = x$ for all $x \in F$.

(A5) For every element $x \in F$, there exists an element $-x \in F$ such that $x + (-x) = 0$.

(M1) If $x \in F$ and $y \in F$, then $xy \in F$.

(M2) *(commutativity of multiplication)* $xy = yx$ for all $x, y \in F$.

(M3) *(associativity of multiplication)* $(xy)z = x(yz)$ for all $x, y, z \in F$.

(M4) There exists an element $1 \in F$ (with $1 \neq 0$) such that $1x = x$ for all $x \in F$.

(M5) For every $x \in F$ such that $x \neq 0$, there exists an element $1/x \in F$ such that $x(1/x) = 1$.

(D) *(distributive law)* $x(y + z) = xy + xz$ for all $x, y, z \in F$.

Definition 1.6. A field F is said to be an *ordered field* if F is also an ordered set such that

(i) For $x, y, z \in F$, $x < y$ implies $x + z < y + z$.

(ii) For $x, y \in F$, $x > 0$ and $y > 0$ implies $xy > 0$.

If $x > 0$, we say x is *positive*. If $x < 0$, we say x is *negative*. We also say x is *nonnegative* if $x \geq 0$, and x is *nonpositive* if $x \leq 0$.

Proposition 1.2. *Let F be an ordered field and $x, y, z, w \in F$. Then*

(i) If $x > 0$, then $-x < 0$ (and vice versa).

(ii) If $x > 0$ and $y < z$, then $xy < xz$.

(iii) If $x < 0$ and $y < z$, then $xy > xz$.

(iv) If $x \neq 0$, then $x^2 > 0$.

(v) If $0 < x < y$, then $0 < 1/y < 1/x$.

(vi) If $0 < x < y$, then $x^2 < y^2$.

(vii) If $x \leq y$ and $z \leq w$, then $x + z \leq y + w$.

Note that (iv) implies, in particular, that $1 > 0$.

Proof. Let us prove (i). The inequality $x > 0$ implies by item (i) of the definition of ordered fields that $x + (-x) > 0 + (-x)$. Apply the algebraic properties of fields to obtain $0 > -x$. The "vice versa" follows by a similar calculation.

For (ii), note that $y < z$ implies $0 < z - y$ by item (i) of the definition of ordered fields. Apply item (ii) of the definition of ordered fields to obtain $0 < x(z - y)$. By algebraic properties, $0 < xz - xy$. Again, by item (i) of the definition, $xy < xz$.

Part (iii) is left as an exercise.

To prove part (iv), first suppose $x > 0$. By item (ii) of the definition of ordered fields, $x^2 > 0$ (use $y = x$). If $x < 0$, we use part (iii) of this proposition, where we plug in $y = x$ and $z = 0$.

To prove part (v), notice that $1/y$ cannot be equal to zero (why?). Suppose $1/y < 0$, then $-1/y > 0$ by (i). Apply part (ii) of the definition (as $x > 0$) to obtain $x(-1/y) > 0$ or $-1 > 0$, which contradicts $1 > 0$ by using part (i) again. Hence $1/y > 0$. Similarly, $1/x > 0$. Thus $(1/x)(1/y)x < (1/x)(1/y)y$.

By algebraic properties, $1/y < 1/x$.

Parts (vi) and (vii) are left as exercises.

□

Definition 1.7. Let $E \subset S$, where S is an ordered set.

(i) If $\exists b \in S$ such that $x \leq b$, $\forall x \in E \implies E$ is *bounded above* and b is an *upper bound* of E.

(ii) If $\exists b \in S$ such that $x \geq b$, $\forall x \in E \implies E$ is *bounded below* and b is a *lower bound* of E.

(iii) If $\exists b_0$ an upper bound of E such that $b_0 \leq b$, $\forall$ upper bounds b of E, then b_0 is called the *least upper bound* or the *supremum* of E. We write:

$$\sup E := b_0.$$

(iv) If $\exists b_0$ a lower bound of E such that $b_0 \geq b$, $\forall$ lower bounds b of E, then b_0 is called the *greatest lower bound* or the *infimum* of E. We write

$$\inf E := b_0.$$

When a set E is both bounded above and bounded below, we say simply that E is *bounded.*

Definition 1.8 (Least Upper Bound Property)**.** An ordered set S has the *least-upper-bound property* if every nonempty subset $E \subset S$ that is bounded above has a least upper bound, that is, $\sup E$ exists in S.

The *least-upper-bound property* is sometimes called the *completeness property* or the *Dedekind completeness property.*

Remark 1.1. So since A is a subset of an ordered field that has the least upper bound property, which states that every set bounded above with the least upper bound property is bounde

Proposition 1.3. *Let F be an ordered field with the least-upper-bound property. Let $A \subset F$ be a nonempty set that is bounded below. Then* $\inf A$ *exists.*

Proof. Let $B = \{-a \mid a \in A\}$. Then since A is bounded above with the least upper bound property, $\exists \sup A = b \in F$. Thus $\forall a \in A, a \leq b$ which implies $-b \leq -a$, which means that B is bounded below by $-b$. Now suppose $\exists M \in F$ such that

$$\forall -a \in B, \quad -b \leq M \leq -a \implies b \geq -M \geq a$$

Since this is contradicts $b = \sup A$. Therefore we have found that B is bounded below by $-b$ and $-b$ is greater than every other lower bound, so $\inf B$ exists.

□

Exercise 1.1. *Let S be an ordered set, and let $B \subseteq S$ be a subset that is bounded above and below. Suppose that $A \subseteq B$ is a nonempty subset and that both* $\inf A$ *and* $\sup A$ *exist. Then we have the inequalities:*

$$\inf B \leq \inf A \leq \sup A \leq \sup B.$$

Proof. Let $B \subset S$ be bounded above and below, and let $A \subset B$ be nonempty. By definition of greatest lower bound, every lower bound of B is also a lower bound of A (since $A \subset B$), and hence $\inf B \leq \inf A$. Also, every upper bound of B is an upper bound of A, so $\sup A \leq \sup B$. Furthermore, because A is nonempty, for any $x \in A$ we have $\inf A \leq x \leq \sup A$, which ensures $\inf A \leq \sup A$. Combining these gives

$$\inf B \quad \leq \quad \inf A \quad \leq \quad \sup A \quad \leq \quad \sup B,$$

as required. □

Remark 1.2. Notice that it seems like we are being imprecise about the infs and sups across subsets. We are actually using the definition, try to contradict and show that $\sup A > \sup B$.

Proposition 1.4 (The Supremum is the least upper bound)**.** *Let $S \subset \mathbb{R}$ be nonempty, and $L \in \mathbb{R} \cup \{\infty, -\infty\}$. Then*

$$\sup S \leq L \iff s \leq L \quad \forall s \in S.$$

Proof. Suppose $\sup S \leq L$. Then by transitivity of ordering 1.3

$$s \leq \sup S \leq L \quad \forall s \in S$$

Which shows $s \leq L$.
Conversely, suppose for some $L \in \mathbb{R} \cup \{\infty, -\infty\}$ we have $s \leq L$, $\forall s \in S$. Since we can say that L is in the set of extended reals that bound the set S where $\sup S$ is the least element, so we have

$$s \leq \sup S \leq L \quad \forall s \in S.$$

□

Exercise 1.2. *Let $A, B \subset \mathbb{R}$ be nonempty sets such that $x \leq y$ whenever $x \in A$ and $y \in B$. Assume A is bounded above, B is bounded below, and $\sup A \leq \inf B$. Then it follows that A is bounded below, B is bounded above, and moreover:*

$$\sup A \leq \inf B.$$

This inequality confirms that the upper bound of A does not exceed the lower bound of B, effectively placing A entirely below or at most touching B.

Exercise 1.3. *If S and T are nonempty subsets of $\mathbb{R}$ and $T \subseteq S$, then $\sup T \leq \sup S$ and $\inf T \geq \inf S$. Note that the supremum and infimum could be finite or infinite.*

Proof. Suppose nonempty sets $T \subseteq S \subseteq \mathbb{R}$ exist. Then $\forall t \in T$, $\exists s_1, s_2 \in S$ such that $s_1 \leq t \leq s_2$. Then,

$$\inf S \leq s_1 \leq \inf T \leq t, \quad \forall t, \implies \inf T \geq \inf S.$$

$$t \leq \sup T \leq s_2 \leq \sup S, \quad \forall t, \implies \sup T \leq \sup S.$$

This states that every upper/lower bound of S is also an upper/lower bound of T so the maximum/minimum of such bounds must too satisfy the inequality. Which is exactly what we wanted to prove. Note that the inequalities above also hold if the sets are unbounded. We can see this by considering an example,

$$\text{If } \sup T = \infty \implies \sup S = \infty$$

but the converse does not hold, as T could just be a finite subset.

□

Exercise 1.4. *Let A and B be two nonempty bounded sets of real numbers, and let $C = \{a + b : a \in A, b \in B\}$ and $D = \{ab : a \in A, b \in B\}$. Then*

1. $\sup C = \sup A + \sup B$ *and* $\inf C = \inf A + \inf B$.
2. $\sup D = (\sup A)(\sup B)$ *and* $\inf D = (\inf A)(\inf B)$.

Definition 1.9. A *function* $f : A \to B$ is a subset f of $A \times B$ such that for each $x \in A$, there exists a unique $y \in B$ for which $(x, y) \in f$. We write $f(x) = y$. Sometimes the set f is called the *graph* of the function rather than the function itself.

The set A is called the *domain* of f (and sometimes confusingly denoted $D(f)$). The set

$$R(f) := \{y \in B : \text{there exists an } x \in A \text{ such that } f(x) = y\}$$

is called the *range* of f. The set B is called the *codomain* of f.

Definition 1.10. Consider a function $f : A \to B$. Define the *image* (or *direct image*) of a subset $C \subset A$ as

$$f(C) := \{f(x) \in B : x \in C\}.$$

Define the *inverse image* of a subset $D \subset B$ as

$$f^{-1}(D) := \{x \in A : f(x) \in D\}.$$

In particular, $R(f) = f(A)$, the range is the direct image of the domain A.

Theorem 1.2. *Let $f : A \to B$ be a function. Then the inverse relation f^{-1} is a function from B to A if and only if f is bijective. Furthermore, if f is bijective, then f^-1 is also bijective.*

Proposition 1.5. *Consider $f : A \to B$. Let C, D be subsets of B. Then*

$$f^{-1}(C \cup D) = f^{-1}(C) \cup f^{-1}(D),$$

$$f^{-1}(C \cap D) = f^{-1}(C) \cap f^{-1}(D),$$

$$f^{-1}(C^c) = (f^{-1}(C))^c.$$

Read the last line of the proposition as $f^{-1}(B \setminus C) = A \setminus f^{-1}(C)$.

Proposition 1.6. *Consider $f : A \to B$. Let C, D be subsets of A. Then*

$$f(C \cup D) = f(C) \cup f(D),$$
$$f(C \cap D) \subseteq f(C) \cap f(D).$$

Definition 1.11. Let $f : A \to B$ be a function. The function f is said to be *injective* or *one-to-one* if

$$f(x_1) = f(x_2) \text{ implies } x_1 = x_2.$$

In other words, f is injective if for all $y \in B$, the set $f^{-1}(\{y\})$ is empty or consists of a single element. We call such an f an *injection*.

If $f(A) = B$, then we say f is *surjective* or *onto*. In other words, f is surjective if the range and the codomain of f are equal. We call such an f a *surjection*.

If f is both surjective and injective, then we say f is *bijective* or that f is a *bijection*.

Definition 1.12. Let $f : A \to B$ and $g : B \to C$ be functions. Then we define the composition as $(g \circ f)(x) = g(f(x))$. So we first use f to map from A to B, then take the value of f in B and input into g and use it to map to C.

Proposition 1.7. *If $f : A \to B$ and $g : B \to C$ are bijective functions, then $f \circ g$ is bijective.*

Definition 1.13. Let A and B be sets. We say A and B have the same *cardinality* when there exists a bijection $f : A \to B$.

We denote by $|A|$ the equivalence class of all sets with the same cardinality as A, and we simply call $|A|$ the *cardinality* of A.

Definition 1.14. We write

$$|A| \leq |B|$$

if there exists an injection from A to B.

We write $|A| = |B|$ if A and B have the same cardinality.

We write $|A| < |B|$ if $|A| \leq |B|$, but A and B do not have the same cardinality.

If $|A| \leq |\mathbb{N}$ then we say that A is countable. If $|A| = |\mathbb{R}|$ then we say that A is uncountable.

Theorem 1.3. *If there exists a bijective function between two sets A and B, then we have that the cardanalities, 1.13, are equivalenet.*

Exercise 1.5. *Let S be a nonempty collection of nonempty sets. A realation R is defined on S by A R B if there exists a bijective function from A to B. Then R is an equivalence relation 1.1.*

Proposition 1.8. *The set $\mathbb{Z}$ is countable*

Proposition 1.9. *Every infinite subset of a countable set is also countable*

Proposition 1.10. *If A and B are countable, then $A \times B$ is countable*

Theorem 1.4. *The set $\mathbb{Q}$ is countable*

Theorem 1.5. *The open interval $(0, 1)$ of real numbers is uncountable.*

Theorem 1.6. $|(0, 1)| = |\mathbb{R}|$

Theorem 1.7. $|\mathcal{P}(A)| = |2^A|$

Lemma 1.1. *Let $f : A \to B$ and $g : C \to D$ be one-to-one functions, where $A \cap C = \emptyset$, and where the function $h : A \cup C \to B \cup D$ is defined by*

$$h(x) = \begin{cases} f(x) & \text{if } x \in A, \\ g(x) & \text{if } x \in C. \end{cases}$$

If $B \cap D = \emptyset$, then h is also a one-to-one function. Consequently, if f and g are bijective functions, then h is a bijective function.

Theorem 1.8. *Let A and B be nonempty sets such that $B \subseteq A$. If there exists an injective function from A to B, then there exists a bijective function from A to B.*

Theorem 1.9 (**Schröder-Bernstein Theorem**)**.** *If A and B are sets such that $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$.*

Theorem 1.10. $|\mathcal{P}(\mathbb{N})| = |\mathbb{R}|$

1.1 Metric Spaces

Definition 1.15. Let X be a set, and let $d : X \times X \to \mathbb{R}$ be a function such that for all $x, y, z \in X$:

1. $d(x, y) \geq 0$ (nonnegativity)
2. $d(x, y) = 0$ if and only if $x = y$ (identity of indiscernibles)
3. $d(x, y) = d(y, x)$ (symmetry)
4. $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality)

The pair (X, d) is called a *metric space*. The function d is called the *metric* or the *distance function*. Sometimes we write just X as the metric space instead of (X, d) if the metric is clear from context.

Lemma 1.2. *(Cauchy-Schwarz inequality). Suppose* $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, $y = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$. *Then*

$$\left(\sum_{k=1}^{n} x_k y_k\right)^2 \leq \left(\sum_{k=1}^{n} x_k^2\right)\left(\sum_{k=1}^{n} y_k^2\right).$$

Proposition 1.11. *Let* (X, d) *be a metric space and* $Y \subset X$. *Then the restriction* $d|_{Y\times Y}$ *is a metric on* Y.

Definition 1.16. If (X, d) is a metric space, $Y \subset X$, and $d' := d|_{Y\times Y}$, then (Y, d') is said to be a *subspace* of (X, d).

Definition 1.17. Let (X, d) be a metric space. A subset $S \subset X$ is said to be *bounded* if there exists a $p \in X$ and a $B \in \mathbb{R}$ such that

$$d(p, x) \leq B \quad \text{for all } x \in S.$$

We say (X, d) is *bounded* if X itself is a bounded subset.

Definition 1.18. Let (X, d) be a metric space, $x \in X$, and $\delta > 0$. Define the *open ball*, or simply *ball*, of radius δ around x as

$$B(x, \delta) := \{y \in X : d(x, y) < \delta\}.$$

Define the *closed ball* as

$$C(x, \delta) := \{y \in X : d(x, y) \leq \delta\}.$$

When dealing with different metric spaces, it is sometimes vital to emphasize which metric space the ball is in. We do this by writing $B_X(x, \delta) := B(x, \delta)$ or $C_X(x, \delta) := C(x, \delta)$.

Definition 1.19. Let (X, d) be a metric space. A subset $V \subset X$ is *open* if for every $x \in V$, there exists a $\delta > 0$ such that $B(x, \delta) \subset V$. A subset $E \subset X$ is *closed* if the complement $E^c = X \setminus E$ is open. When the ambient space X is not clear from context, we say V is *open in* X and E is *closed in* X. If $x \in V$ and V is open, then we say V is an *open neighborhood* of x (or sometimes just *neighborhood*).

Proposition 1.12. *Let* (X, d) *be a metric space.*

1. $\emptyset$ *and* X *are open.*
2. *If* $V_1, V_2, \ldots, V_k$ *are open subsets of* X, *then*

   $$\bigcap_{j=1}^{k} V_j$$

   is also open. That is, a finite intersection of open sets is open.
3. *If* $\{V_\lambda\}_{\lambda\in I}$ *is an arbitrary collection of open subsets of* X, *then*

   $$\bigcup_{\lambda\in I} V_\lambda$$

   is also open. That is, a union of open sets is open.

Proposition 1.13. *Let (X, d) be a metric space.*

1. *$\emptyset$ and X are closed.*
2. *If $\{E_\lambda\}_{\lambda \in I}$ is an arbitrary collection of closed subsets of X, then*
$$\bigcap_{\lambda \in I} E_\lambda$$
is also closed. That is, an intersection of closed sets is closed.
3. *If $E_1, E_2, \ldots, E_k$ are closed subsets of X, then*
$$\bigcup_{j=1}^{k} E_j$$
is also closed. That is, a finite union of closed sets is closed.

Proposition 1.14. *Let (X, d) be a metric space, $x \in X$, and $\delta > 0$. Then $B(x, \delta)$ is open and $C(x, \delta)$ is closed.*

Proposition 1.15. *Suppose (X, d) is a metric space, and $Y \subset X$. Then $U \subset Y$ is open in Y (in the subspace topology) if and only if there exists an open set $V \subset X$ (so open in X) such that $V \cap Y = U$.*

Proposition 1.16. *Suppose (X, d) is a metric space, $V \subset X$ is open, and $E \subset X$ is closed.*

1. *$U \subset V$ is open in the subspace topology if and only if U is open in X.*
2. *$F \subset E$ is closed in the subspace topology if and only if F is closed in X.*

Definition 1.20. A nonempty metric space (X, d) is *connected* if the only subsets of X that are both open and closed (so-called *clopen* subsets) are $\emptyset$ and X itself. If a nonempty (X, d) is not connected, we say it is *disconnected*.

When we apply the term *connected* to a nonempty subset $A \subset X$, we mean that A with the subspace topology is connected.

In other words, a nonempty X is connected if whenever we write $X = X_1 \cup X_2$ where $X_1 \cap X_2 = \emptyset$ and X_1 and X_2 are open, then either $X_1 = \emptyset$ or $X_2 = \emptyset$. So to show X is disconnected, we need to find nonempty disjoint open sets X_1 and X_2 whose union is X.

Proposition 1.17. *Let (X, d) be a metric space. A nonempty set $S \subset X$ is disconnected if and only if there exist open sets U_1 and U_2 in X such that $U_1 \cap U_2 \cap S = \emptyset$, $U_1 \cap S \neq \emptyset$, $U_2 \cap S \neq \emptyset$, and*
$$S = (U_1 \cap S) \cup (U_2 \cap S).$$

Proposition 1.18. *A nonempty set $S \subset \mathbb{R}$ is connected if and only if S is an interval or a single point.*

Definition 1.21. Let (X, d) be a metric space and $A \subset X$. The *closure* of A is the set
$$\bar{A} := \bigcap \{E \subset X : E \text{ is closed and } A \subset E\}.$$
That is, $\bar{A}$ is the intersection of all closed sets that contain A.

Proposition 1.19. *Let (X, d) be a metric space and $A \subset X$. The closure $\bar{A}$ is closed, and $A \subset \bar{A}$. Furthermore, if A is closed, then $\bar{A} = A$.*

Proposition 1.20. *Let (X, d) be a metric space and $A \subset X$. Then $x \in \bar{A}$ if and only if for every $\delta > 0$, $B(x, \delta) \cap A \neq \emptyset$.*

Definition 1.22. Let (X, d) be a metric space and $A \subset X$. The *interior* of A is the set
$$A^\circ := \{x \in A : \text{there exists a } \delta > 0 \text{ such that } B(x, \delta) \subset A\}.$$
The *boundary* of A is the set
$$\partial A := \bar{A} \setminus A^\circ.$$

Proposition 1.21. *Let (X, d) be a metric space and $A \subset X$. Then A° is open and ∂A is closed.*

Proposition 1.22. *Let (X, d) be a metric space and $A \subset X$. Then $x \in \partial A$ if and only if for every $\delta > 0$, $B(x, \delta) \cap A$ and $B(x, \delta) \cap A^c$ are both nonempty.*

Corollary 1.1. *Let (X, d) be a metric space and $A \subset X$. Then*

$$\partial A = \bar{A} \cap \overline{A^c}.$$

Proposition 1.23. *Let (X, d) be a metric space and $\{x_n\}_{n=1}^{\infty}$ a sequence in X. Then $\{x_n\}_{n=1}^{\infty}$ converges to $p \in X$ if and only if for every open neighborhood U of p, there exists an $M \in \mathbb{N}$ such that for all $n \geq M$, we have $x_n \in U$.*

Proof. *Suppose $\{x_n\}_{n=1}^{\infty}$ converges to p. Let U be an open neighborhood of p, then there exists an $\epsilon > 0$ such that $B(p, \epsilon) \subset U$. As the sequence converges, find an $M \in \mathbb{N}$ such that for all $n \geq M$, we have $d(p, x_n) < \epsilon$, or in other words $x_n \in B(p, \epsilon) \subset U$.*

Conversely, given $\epsilon > 0$, let $U := B(p, \epsilon)$ be the neighborhood of p. Then there is an $M \in \mathbb{N}$ such that for $n \geq M$, we have $x_n \in U = B(p, \epsilon)$, or in other words, $d(p, x_n) < \epsilon$. $\square$

A closed set contains the limits of its convergent sequences.

Proposition 1.24. *Let (X, d) be a metric space and $A \subset X$. Then $p \in \bar{A}$ if and only if there exists a sequence $\{x_n\}_{n=1}^{\infty}$ of elements in A such that*

$$\lim_{n \to \infty} x_n = p.$$

Definition 1.23. We say a metric space (X, d) is *complete* or *Cauchy-complete* if every Cauchy sequence $\{x_n\}_{n=1}^{\infty}$ in X converges to a $p \in X$.

Proposition 1.25. *The space $\mathbb{R}^n$ with the standard metric is a complete metric space.*

Proposition 1.26. *The space of continuous functions $C([a, b], \mathbb{R})$ with the uniform norm as metric is a complete metric space.*

Definition 1.24. Let (X, d) be a metric space and $K \subset X$. The set K is said to be *compact* if for every collection of open sets $\{U_\lambda\}_{\lambda \in I}$ such that

$$K \subset \bigcup_{\lambda \in I} U_\lambda,$$

there exists a finite subset $\{\lambda_1, \lambda_2, \ldots, \lambda_m\} \subset I$ such that

$$K \subset \bigcup_{j=1}^{m} U_{\lambda_j}.$$

A collection of open sets $\{U_\lambda\}_{\lambda \in I}$ as above is said to be an *open cover* of K. A way to say that K is compact is to say that *every open cover of K has a finite subcover.*

Proposition 1.27. *Let (X, d) be a metric space. If $K \subset X$ is compact, then K is closed and bounded.*

Lemma 1.3. *(Lebesgue covering lemma). Let (X, d) be a metric space and $K \subset X$. Suppose every sequence in K has a subsequence convergent in K. Given an open cover $\{U_\lambda\}_{\lambda \in I}$ of K, there exists a $\delta > 0$ such that for every $x \in K$, there exists a $\lambda \in I$ with $B(x, \delta) \subset U_\lambda$.*

Theorem 1.11. *Let (X, d) be a metric space. Then $K \subset X$ is compact if and only if every sequence in K has a subsequence converging to a point in K.*

Proposition 1.28. *Let (X, d) be a metric space and let $K \subset X$ be compact. If $E \subset K$ is a closed set, then E is compact.*

Theorem 1.12. *(Heine-Borel theorem). A closed bounded subset $K \subset \mathbb{R}^n$ is compact.*

So subsets of $\mathbb{R}^n$ are compact if and only if they are closed and bounded, a condition that is much easier to check. Let us reiterate that the Heine-Borel theorem only holds for $\mathbb{R}^n$ and not for metric spaces in general. The theorem does not hold even for subspaces of $\mathbb{R}^n$, just in $\mathbb{R}^n$ itself. In general, compact implies closed and bounded, but not vice versa.

Definition 1.25. Let (X, d_X) and (Y, d_Y) be metric spaces and $c \in X$. Then $f : X \to Y$ is *continuous* at c if for every $\epsilon > 0$ there is a $\delta > 0$ such that whenever $x \in X$ and $d_X(x, c) < \delta$, then $d_Y(f(x), f(c)) < \epsilon$.

When $f : X \to Y$ is continuous at all $c \in X$, we simply say that f is a *continuous function*.

Proposition 1.29. *Let (X, d_X) and (Y, d_Y) be metric spaces. Then $f : X \to Y$ is continuous at $c \in X$ if and only if for every sequence $\{x_n\}_{n=1}^{\infty}$ in X converging to c, the sequence $\{f(x_n)\}_{n=1}^{\infty}$ converges to $f(c)$.*

Lemma 1.4. *Let (X, d_X) and (Y, d_Y) be metric spaces and $f : X \to Y$ a continuous function. If $K \subset X$ is a compact set, then $f(K)$ is a compact set.*

Theorem 1.13. *Let (X, d) be a nonempty compact metric space and let $f : X \to \mathbb{R}$ be continuous. Then f is bounded and in fact f achieves an absolute minimum and an absolute maximum on X.*

Proof. As X is compact and f is continuous, $f(X) \subset \mathbb{R}$ is compact. Hence $f(X)$ is closed and bounded. In particular, $\sup f(X) \in f(X)$ and $\inf f(X) \in f(X)$, because both the sup and the inf can be achieved by sequences in $f(X)$ and $f(X)$ is closed. Therefore, there is some $x \in X$ such that $f(x) = \sup f(X)$ and some $y \in X$ such that $f(y) = \inf f(X)$. □

Lemma 1.5. *Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f : X \to Y$ is continuous at $c \in X$ if and only if for every open neighborhood U of $f(c)$ in Y, the set $f^{-1}(U)$ contains an open neighborhood of c in X.*

Theorem 1.14. *Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f : X \to Y$ is continuous if and only if for every open $U \subset Y$, $f^{-1}(U)$ is open in X.*

2 Algebra

Definition 2.1. A number is called an *algebraic number* if it satisfies a polynomial equation

$$c_n x^n + c_{n-1} x^{n-1} + \ldots + c_1 x + c_0 = 0$$

where the coefficients $c_0, c_1, \ldots, c_n$ are integers and $c_n \neq 0$ and $n \geq 1$.

Theorem 2.1 (Rational Zeros Theorem)**.** *Suppose $c_0, c_1, \ldots, c_n$ are integers and $r \in \mathbb{Q}$ satisfies the polynomial*

$$c_n x^n + c_{n-1} x^{n-1} + \ldots + c_1 x + c_0 = 0$$

where $n \geq 1, c_n \neq 0,$ and $c_0 \neq 0$. Let $r = \frac{m}{d}$, where $m, d \in \mathbb{Z}$ such that $\gcd(m, d) = 1$ and $d \neq 0$. *Then $m \mid c_0$* and $d \mid c_n$.

Proof. Let $x = r = m/d$ be a solution to the polynomial. Then,

$$\begin{gathered} c_n x^n + c_{n-1} x^{n-1} + \ldots + c_1 x + c_0 = 0, \\ c_n \left(\frac{m^n}{d^n}\right) + c_{n-1} \left(\frac{m^{n-1}}{d^{n-1}}\right) + \ldots + c_1 \left(\frac{m}{d}\right) + c_0 = 0. \\ c_n m^n + c_{n-1} m^{n-1} d + \cdots + c_1 m d^{n-1} + c_0 d^n = 0 \end{gathered}$$

Then rearranging, we see

$$c_0 d^n = -m \left(c_n m^{n-1} + c_{n-1} m^{n-2} d + \cdots c_1 d^{n-1}\right)$$

Since $gcd(m, d) = 1$, we know that $gcd(m, d^n) = 1$, and thus m divides c_0. Now rearranging again, we see

$$c_n m^n = -d \left(c_{n-1} m^{n-1} + \cdots + c_1 m d^{n-2} + c_0 d^{n-1}\right)$$

Thus, d divides c_n.

□

Remark 2.1. The result above states that given a polynomial with integer coefficients, a constant term, and a nonzero leading coefficient, if the polynomial is going to have rational roots, then the numerator of the root will divide the constant and the denominator will divide the leading coefficient. Note that often the leading coefficient is 1 so we typically only ensure the numerator divides the constant. Also note that we are not saying this rational is always a root, we are only saying that if a rational is a root, it has the form described above.

2.1 Divisibility in $\mathbb{Z}$

We start by defining the integers. This ordered set will be our object of study. SAY MORE HERE

Definition 2.2 ($\mathbb{Z}$)**.** The set of integers is any ordered set equipped with two operations $+, \cdot$ that satisfy the following axioms. $\forall a, b, c \in \mathbb{Z}$:

1. If $a, b \in \mathbb{Z}$, then $a + b \in \mathbb{Z}$ *[Closure for addition]*
2. $a + (b + c) = (a + b) + c$ *[Associative addition]*
3. $a + b = b + a$ *[Commutative addition]*
4. $a + 0 = a = 0 + a$ *[Additive identity]*
5. For each $a \in \mathbb{Z}$, the equation $a + x = 0$ has a solution in $\mathbb{Z}$.
6. If $a, b \in \mathbb{Z}$, then $ab \in \mathbb{Z}$ *[Closure for multiplication]*
7. $a(bc) = (ab)c$ *[Associative multiplication]*
8. $a(b + c) = ab + ac$ and
   $(a + b)c = ac + bc$ *[Distributive laws]*
9. $ab = ba$ *[Commutative multiplication]*
10. $a \cdot 1 = a = 1 \cdot a$ *[Multiplicative identity]*
11. If $ab = 0$, then $a = 0$ or $b = 0$.

Remark 2.2. The below result is foundational to all of number theory and abstract algebra. It is the idea that given some number a to know how b *fits into* a we will take as many copies or multiples of b. We want to show existence and uniqueness. To show existence, we will show that such an r satisfying the hypothesis exists.
So we will consider numbers of the form $r = a - bq$. So we make a set of this form and show that it is nonempty. Then we will let the unique q, r correspond to the min of the set.

Theorem 2.2 (Division Algorithm)**.** *Let $a, b \in \mathbb{Z}$ with $b > 0$. Then there exist unique $q, r \in \mathbb{Z}$ such that*

$$a = bq + r \text{ and } 0 \leq r < b.$$

Proof. Consider,

$$S = \{a - bx \mid \forall x \in \mathbb{Z}\}$$

We start by showing S is nonempty.
Observe that $|a| \in S$ since we can let $x = 0$ which gives $0 \leq a$, which tells us positive a is in S.
Now let $r = \min S$. We know r exists by the Well Ordering Axiom. Then let $x = q$ correspond to r. We will now show that $r < b$.
By contradiction, suppose $r > b$. Then this gives us that there is at least one factor of b in r.

$$a = bq + r = b(q + 1) + r' \implies r' \in S \text{ and } r' < r$$

which contradicts that $r = \min S$, thus q and r exist.
Now we show uniqueness. Suppose there exists r' and q' such that

$$a = bq + r = bq' + r' \implies r' - r = b(q - q').$$

Since we have that both r and r' are less than b, this gives

$$| r' - r | < b \implies | b(q - q') | < b \implies | q - q' | < 1$$

Then since the difference $q - q'$ is an integer, we have that $q = q' \implies r = r'$. ☐

Definition 2.3 (Greatest Common Divisor)**.** For any two nonzero integers a and b, the *greatest common divisor* $gcd(a, b)$ is the unique positive integers d such that

1. $d \mid a$ and $d \mid b$
2. If $\exists c \in \mathbb{Z}$ such that $c \mid a$ and $c \mid b$, then $c \leq d$.

Remark 2.3. The greatest common divisor between any two integers will prove to be an important topic. When broken down, it is essentially a set of the shared factors of a and b. Why would that be so? Because if d is the greatest magnitude greater than 0 that divides both a and b then every other divisor that is greater than 0 but be *contained* in the magnitude of d. This will be helpful as a sort of relation between the integers and their *intersection with respect to divisibility.*

Theorem 2.3 (Bezout's Identity)**.** *Let a and b be integers, not both 0, and let $d = gcd(a, b)$. Then there exists integers u and v such that*

$$gcd(a, b) = d = au + bv$$

Remark 2.4. *Why would this make sense?* So recall that the *gcd* is the largest *positive* divisor, then it would be plausible that the smallest positive integer linear combination of a and b is largest factor that is shared amongst a and b. That is, through linear combinations, we can remove the multiples and factors of a and b that they dont have in their *intersection*, then the magnitude that remains would be the *gcd*. Also notice the usefulness of this result. This allows us to relate the divisibility structure of a and b to any combination that is made with them. *Why does the gcd have to be the least positive element?* First consider if the smallest positive linear combination was greater than, say, a. Since the *gcd* divides both a and b, the smallest linear combo must be less than both a and b. If the smallest linear combo was smaller than the *gcd* then we would have that factors of a and b combine to something positive but less than the greatest factor they have in common.

Proof. Let $S = \{au + bv \mid u, v \in \mathbb{Z}\}$. We will first show that S contains positive integers. Let $u = a$ and $v = b$, then we have $a^2 + b^2 \in S$. Thus there exists positive integers in S. Let $t = \min S$, which we know exists because $S \subset \mathbb{Z}$ so by well ordering axiom there must exist a least positive element. Define $d = gcd(a, b)$. We want to show that $t = d$. We will start by showing $t \mid a$ and $t \mid b$.

$$\text{By 2.2,} \quad a = tq + r \implies r = a - tq \implies r = a - aqu - bqv \implies r = a(1 - qu) + b(-qv)$$

Thus $r \in S$, but since, by the hypothesis of 2.2 $\quad r < t = \min S$. This implies that

☐

Remark 2.5. So we hypothesized that the *gcd* was going to be a linear combination of a and b because it is the greatest factor of them both, so in a way, they can both construct it. We then hypothesized in the proof that the *gcd* is the least positive multiple. So to show that t is the *gcd*, we show that it divides them both and is the greatest such integer to do so. To show that t divides a and b, we show, using 2.2 that the remainder must be in S, but that would mean the remainder is less than t so that gives us what we are looking for.

Proposition 2.1. *Let $a, b, x, y \in \mathbb{Z}$. Then*

$$ax + by = c \iff gcd(a, b) \mid c.$$

Proof. Suppose $ax + by = c$ and let $d = gcd(a, b)$. Then

$$\exists k, l \in \mathbb{Z} \text{ such that } c = dkx + dly \implies d \mid c.$$

Conversely, assume $d = gcd(a, b) \mid c$. That is, $\exists k \in \mathbb{Z}$ such that $dk = c$. Then

$$c = dk = a(kx) + b(ky) \implies \exists u, v \in \mathbb{Z}, \; c = au + bv.$$

This concludes the proof. ☐

Proposition 2.2. *Let $a, b, c \in \mathbb{Z}$. If $a \mid bc$ and $\gcd(a, b) = 1$, then $a \mid c$.*

Proof. Suppose $a \mid bc$ and $gcd(a, b) = 1$. Then $\exists k \in \mathbb{Z}$ such that $ak = bc$. Also by 2.3,

$$\exists u, v \in \mathbb{Z} \text{ such that } 1 = au + bv$$

$$\implies c = acu + bcv \implies c = ac + akv.$$

Thus $a \mid c$. $\square$

Remark 2.6. This proposition is insightful to how the *gcd* will be used often. Notice we have that a divides a product but it shares no factors with b, who is also in the product. Thus the only factors it must share with the product must be with c. So we would expect to have that a divides c.

Exercise 2.1. *Let $a, b, c \in \mathbb{Z}$. Suppose $\gcd(a, b) = 1$. If $a|c$ and $b|c$, then $ab|c$.*

Exercise 2.2. *Let $a, b, c \in \mathbb{Z}$. Then $\forall t \in \mathbb{Z}$ all of the following hold*

1. $\gcd(a, b) = \gcd(a, b + at)$
2. $\gcd(ta, tb) = t \gcd(a, b) \quad \text{for } t > 0$
3. $\gcd(a, \gcd(b, c)) = gcd(gcd(a, b), c)$
4. $\gcd(a, c) = 1 \implies \gcd(ab, c) = \gcd(b, c)$

Exercise 2.3. *Let $a, b, c \in \mathbb{Z}$. If $\gcd(a, c) = 1$ and $\gcd(b, c) = 1$, then $\gcd(ab, c) = 1$*

Exercise 2.4. *A positive integer is divisible by 3 $\iff$ the sum of its digits is divisible by 3.*

Theorem 2.4. *Let $p \in \mathbb{Z}$ with $p \neq 0, 1, -1$. Then p is prime if and only if p has the following property*

$$\text{whenever } p \mid bc, \text{ then } p \mid b \text{ or } p \mid c$$

Remark 2.7. This is obvious in comparison to 2.2 since a prime is coprime to every integer. Thus we will lean on that proof heavily.

Proof. Suppose p is prime and consider $p \mid bc$. Since p is prime, if $p \mid b$ then the theorem is proved, if $p \nmid b$ then since p is prime, $gcd(p, b) = 1$. By 2.2 this gives us that $p \mid b$ or $p \mid c$.
Conversely, by the contrapositive, suppose p is not prime. Then if $p \mid bc$ then to have $p \mid b$ or $p \mid c$ we would need that $gcd(p, b) = 1$, $\forall b \in \mathbb{Z}$. But this would mean that p is prime.

$\square$

Theorem 2.5 (Fundamental Theoerem of Arithmetic)**.** *Every integer $n \neq 0, 1, -1$ has a unique prime factorization.*

Proof. First we will show existence of the factorization.
Let $S = \{n \in \mathbb{N} \mid n > 1 \text{ and } \nexists \text{ primes } p_1 p_2 \cdots p_n \text{ such that } p_1 p_2 \cdots p_n = n\}$. Then assume, by contradiction, that S is nonempty. Then by the well ordering axiom, let $n = \min S$. Since n is not prime, $\exists a, b \in \mathbb{Z}$ such that $ab = n$. Then this means $a \mid n$ and $b \mid n$. Since $a, b \leq n$, we have that a and b have prime factorizations. Thus n has a prime factorization. This proves the existence of a prime factorization for all integers.
Now we will show that this factorization is unique.
By $\square$

Exercise 2.5. *If $n > 1$ has no positive prime faster less than or equal to $\sqrt{n}$, then s is prime.*

Exercise 2.6. $a|b \iff a^n|b^n$

2.2 Congruence and Congruence Classes

Remark 2.8. The concepts below intend to study the structure that arithmetic and divisibility have among the integers. We do this by making our object of focus the remainder that an integer leaves after being divided. If some integer a leaves behind the same remainder as some other integer b when divided by n, then their difference $a - b$ is divisible by n. If we use their unique representation from 2.2, then

$$a - b = nq_1 + r - nq_2 - r = n(q_1 - q_2)$$

Why do we care about the divisibility structure? We will soon see that what we see as divisibility among numbers can actually be abstracted and shown to be an example of a more general concept. The concepts discussed later will show that the properties we find out about the integers actually are very similar properties that the more general elements share with each other.

Definition 2.4 (Congruence $\pmod{n}$)**.** Let $a, b, n \in \mathbb{Z}$ with $n > 0$. Then a is congruent to b modulo n if $n \mid a - b$. This is denoted $a \equiv b \pmod{n}$

Theorem 2.6 (Congruence $\in$ Equivalence Relations)**.** *Let n be a positive integer, then $\forall a, b, c \in \mathbb{Z}$,*

1. $a \equiv a \pmod{n}$
2. *If* $a \equiv b \pmod{n}$*, then* $b \equiv a \pmod{n}$
3. *If* $a \equiv b \pmod{n}$ *and* $b \equiv c \pmod{n}$*, then* $a \equiv c \pmod{n}$*.*

Proof. The proof of (1) and (2) is straightforward after seeing the proof of (3). If $a \equiv b \mod n$ and $b \equiv c \mod n$ then we can write

$$\begin{aligned} \exists k, l \in \mathbb{Z} : \quad a - b = nk \quad &\text{and} \quad b - c = nl \\ \implies \quad b = a - nk \quad &\text{and} \quad b = c + nl \\ \implies \quad & a - c = n(k + l). \end{aligned}$$

Thus $a \equiv c \mod n$.

□

Proposition 2.3 (Modulo Arithmetic)**.** *If* $a \equiv b \pmod{n}$ *and* $c \equiv d \pmod{n}$*, then*

1. $a + c \equiv b + d \pmod{n}$
2. $ac \equiv bd \pmod{n}$

Proof. (1) : Since $a \equiv b$ and $c \equiv d$ we have, by definition, $a - b = nk$ and $c - d = nl$. Adding these, we obtain $(a + c) - (b + d) = n(k + l) \implies a + c \equiv b + d$.
(2) : So we want $ac \equiv bd$, or equivalently, we want to find $k \in \mathbb{Z}$ such that $ac - bd = nk$. Then to use the hypothesis we do,

$$ac - bd = ac - bc + bc - bd = c(a - b) + b(c - d) = c(nk) + b(nl) = n(ck + bl).$$

Thus, $ac \equiv bd \mod n$.

□

Definition 2.5 (Congruence Class)**.** Let $a, n \in \mathbb{Z}$ be integers with $n > 0$. The *congruence class* of a modulo n (denoted $[a]$) is the set of all integers that are congruent to a modulo n, that is,

$$[a] = \{b \mid b \in \mathbb{Z} \quad \text{and} \quad b \equiv a \pmod{n}\}.$$

Recall $b \equiv a \pmod{n}$ means that $b - a = kn$ for some integer k or, equivalently, that $b = a + kn$. Thus

$$[a] = \{b \mid b \equiv a \pmod{n}\} = \{b \mid b = a + kn \text{ with } k \in \mathbb{Z}\} = \{a + kn \mid k \in \mathbb{Z}\}$$

Theorem 2.7 (Congruence Class Equality)**.** $a \equiv c \pmod{n}$ *if and only if* $[a] = [c]$.

Proof. Suppose $a \equiv c$, we want to show that $[a] \subset [c]$ and $[c] \subset [a]$, so also suppose that $x \in [a]$. Then by definition of $[a]$, $x \in [a] \implies x \equiv a$, then by transitivity, we have that $x \equiv a$ and $a \equiv c \implies x \equiv c \implies x \in [c]$. Suppose instead that $x \in [c]$. Then again by transitivity we obtain that $x \in [a]$.
Suppose $[a] = [c]$. Then by definition of $[a]$, $a \equiv a$ but since $[a] = [c]$, we have that $a \equiv a \implies a \in [c] \implies a \equiv c$. □

Corollary 2.1. *Two congruence classes modulo n are either disjoint or identical.*

Proof. If $[a]$ and $[c]$ are disjoint, there is nothing to prove. Suppose that $[a] \cap [c]$ is nonempty. Then there is an integer b with $b \in [a]$ and $b \in [c]$. Then, by the definition of congruence class, $b \equiv a \pmod{n}$ and $b \equiv c \pmod{n}$. Therefore, by symmetry and transitivity, $a \equiv c \pmod{n}$. Then by 2.7 we have that, $[a] = [c]$. □

Exercise 2.7. *Let $n > 1$ be an integer and consider congruence modulo n.*

1. *If a is any integer and r is the remainder when a is divided by n, then $[a] = [r]$.*
2. *There are exactly n distinct congruence classes, namely, $[0], [1], [2], \dots, [n-1]$.*

Proof. (1) : Suppose a is an integer and r is the remainder when a is divided by n, then from 2.2 we have, $a = nk + r$ or $a - r = nk \implies a \equiv r \implies [a] = [r]$. Where the last implication used 2.7.
(2) : From (1) we know that any given integer will be the same congruence class as its remainder r where $0 \leq r < n$, thus there are $n-1$ such possible remainders. We also have from 2.1 that each class is disjoint, thus there are $n-1$ possible equivalence classes.

□

Definition 2.6. The set of all congruences classes modulo n is denoted $\mathbb{Z}_n$.
Note that an element of $\mathbb{Z}_n$ is a class, the set of integers that it is congruent to, not a single integer.

Exercise 2.8. *If a, b are integers such that $a \equiv b \pmod{p}$ for every positive prime p, then $a = b$.*

Remark 2.9. We will continue to study division in the integers at this abstracted level by using the concept that equivalence is defined by having the same remainder when divided by a number. The congruence class $\mathbb{Z}_n$ is a set consisting of other sets. These other sets are the sets of integers that are congruent modulo n, and the numbers that are congruent modulo n are the ones that have the same remainder when divided by n. Now we can define relations between classes more effectively.

Theorem 2.8. *If $[a] = [b]$ and $[c] = [d]$ in $\mathbb{Z}_n$, then*

$$[a + c] = [b + d] \quad \textit{and} \quad [ac] = [bd].$$

Proof. From 2.7 we have that $a \equiv b$ and $c \equiv d$. Then from 2.3 we have

$$a + c \equiv b + d \quad \text{and} \quad ac \equiv bd$$

Then from 2.7 again we have $[a + c] = [b + d]$ and $[ac] = [bd]$. □ □

Definition 2.7 (Operations in $\mathbb{Z}_n$)**.** We define addition $+$ and multiplication $\cdot$ in $\mathbb{Z}_n$ by

$$[a] \oplus [c] = [a + c] \quad \text{and} \quad [a] \odot [c] = [ac].$$

Proposition 2.4. *For any classes $[a], [b], [c]$ in $\mathbb{Z}_n$,*

1. *If $[a] \in \mathbb{Z}_n$ and $[b] \in \mathbb{Z}_n$, then $[a] \oplus [b] \in \mathbb{Z}_n$.*
2. $[a] \oplus ([b] \oplus [c]) = ([a] \oplus [b]) \oplus [c]$.
3. $[a] \oplus [b] = [b] \oplus [a]$.
4. $[a] \oplus [0] = [a] = [0] \oplus [a]$.
5. *For each $[a]$ in $\mathbb{Z}_n$, the equation $[a] \oplus x = [0]$ has a solution in $\mathbb{Z}_n$.*

6. *If $[a] \in \mathbb{Z}_n$ and $[b] \in \mathbb{Z}_n$, then $[a] \odot [b] \in \mathbb{Z}_n$.*
7. $[a] \odot ([b] \odot [c]) = ([a] \odot [b]) \odot [c]$.
8. $[a] \odot ([b] \oplus [c]) = [a] \odot [b] \oplus [a] \odot [c]$ *and* $([a] \oplus [b]) \odot [c] = [a] \odot [c] \oplus [b] \odot [c]$.
9. $[a] \odot [b] = [b] \odot [a]$.
10. $[a] \odot [1] = [a] = [1] \odot [a]$.

Remark 2.10 (Change of Notation). From now on, to denote an element in $\mathbb{Z}_n$ we will just denote it by its integer form. That is, when we say we are *in* $\mathbb{Z}_n$, then we will write $[a]_n$ as a. This is just for notational convenience, nothing has changed.

Remark 2.11. After some work with the integers modulo n, we start to notice a pattern, when the integers are modulo a prime number, the $\mathbb{Z}_n$ product of nonzero elements is always nonzero. So the distinction is that when $a \neq 0$ the equation $ax = 1$ has a solution in $\mathbb{Z}$ if and only if $a = 1$ or $a = -1$, but for the multiplication in $\mathbb{Z}_p$ where p is a prime, the equation always has a solution.

Theorem 2.9. *If $p > 1$ is an integer, then the following are equivalent:*

1. *p is prime.*
2. *For any $a \neq 0$ in $\mathbb{Z}_p$, the equation $ax = 1$ has a solution in $\mathbb{Z}_p$.*
3. *Whenever $bc = 0$ in $\mathbb{Z}_p$, then $b = 0$ or $c = 0$.*

Corollary 2.2. *Let a and n be integers with $n > 1$. Then*
The equation $[a]x = [1]$ has a solution in $\mathbb{Z}_n$ if and only if $\gcd(a, n) = 1$ in $\mathbb{Z}$.

Definition 2.8 (Units)**.** For any $a \in \mathbb{Z}_n$, if $\exists b \in \mathbb{Z}_n$ such that $ab = 1$, then a is a *unit*. In this case, we say b is the *inverse* of a.

Definition 2.9 (Zero Divisors)**.** Suppose $a \in \mathbb{Z}_n$ and $a \neq 0$. If $\exists c \in \mathbb{Z}_n$ such that $c \neq 0$ and $ac = 0$.

Exercise 2.9. *Let $n > 1$ be an integer and let a, b be integers. Define $d = \gcd(a, n)$. Consider the linear congruence*

$$[a]x = [b] \quad \textit{in } \mathbb{Z}_n.$$

1. *Show that the congruence has at least one solution if and only if $d \mid b$. Conclude that no solution exists when $d \nmid b$.*
2. *Assume $d \mid b$. Use Bézout's identity to find integers u, v such that*

   $$au + nv = d.$$

   Show that

   $$x = \left[\frac{b}{d}u\right]$$

   is a solution in $\mathbb{Z}_n$.
3. *Prove that every solution is of the form*

   $$x = \left[\frac{b}{d}u + k\frac{n}{d}\right], \quad k \in \{0, 1, \ldots, d-1\}.$$

   Show that these d solutions are pairwise distinct.
4. *Conclude that if $d \mid b$, there are exactly d distinct solutions, and otherwise, there are none. Explain how this fully classifies solutions to linear congruences.*
5. *Solve the congruences:*

   $$13x = 9 \quad \textit{in } \mathbb{Z}_{24}, \quad \textit{and} \quad 25x = 10 \quad \textit{in } \mathbb{Z}_{65}.$$
6. *Show that if $\gcd(a, n) = 1$, then $[a]$ is invertible in $\mathbb{Z}_n$, ensuring a unique solution to $[a]x = [b]$. Relate this to computing the inverse of $[a]$ in $\mathbb{Z}_n$.*

2.3 Rings

We now generalize the properties we have found consistent across the number-like systems we have studied.

Definition 2.10 (Ring)**.** A ring is a nonempty set R equipped with two operations $+, \cdot$ that satisfy the following axioms. $\forall a, b, c \in R$:

1. If $a \in R$ and $b \in R$, then $a + b \in R$. [Closure under Addition]
2. $a + (b + c) = (a + b) + c$ [Associativity of Addition]
3. $a + b = b + a$ [Commutativity of Addition]
4. There exists an element $0_R \in R$ such that $a + 0_R = a = 0_R + a, \ \forall a \in R$ [Additive identity]
5. For each $a \in R$, $a + x = 0_R$ has a solution in R, that is, $x \in R$ [Additive Inverse]
6. If $a \in R$ and $b \in R$, then $ab \in R$ [Closure under Multiplication]
7. $a(bc) = (ab)c$ [Associativity of Multiplication]
8. $a(b + c) = ab + ac$ and $(a + b)c = ac + bc$ [Distributive Law]
   The additional axioms below come from the definitions that are to follow. These definitions are the specific types of rings.
9. $ab = ba \quad \forall a, b \in R$ [Commutative Ring]
10. $\exists 1_R \in R$ such that $a1_R = a = 1_R a \quad \forall a \in R$. [Identity]
11. A commutative ring, with identity such that $ab = 0 \implies a = 0$ or $b = 0$. [Integral Domain]
12. A commutative ring, with identity such that $\forall a \neq 0 \in R$, $ax = 1$ has a solution in R. [Field]

Definition 2.11 (Commutative Ring)**.** A commutative ring is a ring R that satisfies the additional axiom: commutative multiplication

$$ab = ba \quad \forall a, b \in R.$$

Definition 2.12 (Multiplicative Identity)**.** A ring with indentity is a ring R that contains an element 1_R that satisfies the additional axiom: multiplicative identity

$$a1_R = a = 1_R a \quad \forall a \in R.$$

Definition 2.13 (Integral Domain)**.** An integral domain is a commutative ring R with identity $1_R \neq 0_R$ that satisfies the additional axiom

$$\text{Whenever } a, b \in R \text{ and } ab = 0_R, \text{ then } a = 0_R \text{ or } b = 0_R.$$

Definition 2.14 (Field)**.** A field is a commutative ring R with identity $1_R \neq 0_R$ that satisfies the axiom

$$\text{For each } a \neq 0_R \in R, \quad ax = 1_R \text{ has a solution in } R$$

Remark 2.12. Note that these operations don't have to adhere to what we think of as addition and multiplication of two numbers....

Proposition 2.5. *Let R and S be rings. Define addition and multiplication on the Cartesian product $R \times S$ by*

$$(r, s) + (r', s') = (r + r', s + s') \quad \text{and} \quad (r, s)(r', s') = (rr', ss').$$

Then $R \times S$ is a ring. If R and S are both commutative, then so is $R \times S$. If both R and S have an identity, then so does $R \times S$.

Theorem 2.10 (Subring)**.** *Suppose that R is a ring and that S is a subset of R such that:*

1. *S is closed under addition (if $a, b \in S$, then $a + b \in S$);*
2. *S is closed under multiplication (if $a, b \in S$, then $ab \in S$);*
3. *$0_R \in S$;*
4. *If $a \in S$, then the solution of the equation $a + x = 0_R$ is in S.*

Then S is a subring of R.

Proof. In order for S to be a subring of R, we only need to check that the axioms for rings hold. Additonally, we need that the additive identity of S is the same one that is in R. We need only check that axioms, from definition (2.10), 1,6,4, and 5 hold since axioms 2,3,7, and 8 hold for all elements of R. □

Theorem 2.11. *For any element a in a ring R, the equation $a + x = 0_R$ has a unique solution.*

Proof. From axiom 5 in definition (2.10), we know $a+x = 0_R$ has at least one solution, call it u. Then suppose v is another solution. Then we have

$$v = v + 0_R = v + (a + u) = (v + a) + u = 0_R + u = u.$$

So $v = u$ and so the solution is always unique in any ring. □

Theorem 2.12. *If $a + b = a + c$ in a ring R, then $b = c$.*

Proof. Using associativity from (2.10) we have

$$a + c = a + b \implies (c + a) - a = (b + a) - a \implies c + (a - a) = b + (a - a) \implies b = c.$$

□

Proposition 2.6. *For any elements a and b of a ring R,*

1. *$a \cdot 0_R = 0_R = 0_R \cdot a$. In particular, $0_R \cdot 0_R = 0_R$.*
2. *$a(-b) = -ab \quad$ and $\quad (-a)b = -ab$.*
3. *$-(-a) = a$.*
4. *$-(a + b) = (-a) + (-b)$.*
5. *$-(a - b) = -a + b$.*
6. *$(-a)(-b) = ab$.*

If R has an identity, then

7. *$(-1_R)a = -a$.*

Proof. (1): Since $0 + 0 = 0$, using the distributive law, we have

$$\begin{aligned} a \cdot 0 + a \cdot 0 = a(0 + 0) = a \cdot 0 = a \cdot 0 + 0 \\ \implies a \cdot 0 + a \cdot 0 = a \cdot 0 + 0 \implies a \cdot 0 = 0. \end{aligned}$$

Note that the last implication uses 2.12.

(2): Since $-ab$ is the unique solution to $ab+x = 0$, any other solution is equivalent to $-ab$ by 2.12. So we have

$$ab + a(-b) = a(b - b) = a \cdot 0 = 0 \implies -ab = a(-b).$$

(3): Again from (2.12) we know $-(-a)$ is the unique solution of $-a + x = 0$, but a is also a solution,

thus $a = -(-a)$.
(4): Since $-(a+b)$ is the unique solution of $(a+b)+x = 0$ and since addition is commutative, we have

$$(a+b) + (-a) + (-b) = (a-a) + (b-b) = 0 + 0 = 0 \implies (-a) + (-b) = -(a+b).$$

(5): By parts (4) and (3) above, we have

$$-(a-b) = (-a) + (-(-b)) = -a + b.$$

(6): By parts (2) and (3) above,

$$(-a)(-b) = -(a(-b)) = -(-ab) = ab$$

(7): By (2), we have

$$(-1)a = -(1a) = -a.$$

□

Exercise 2.10. *Let $n, m \in \mathbb{N}$, if R is a ring with $a \in R$, then*

$$a^n = aaa \cdots a \quad \text{(n factors)}$$
$$a^n a^m = a^{m+n} \text{ and } (a^m)^n = a^{mn}$$

Remark 2.13. Now with subtraction formally defined, we can revisit theorem **??** and see if we can find a simpler method for checking subrings.

Proposition 2.7 (Subring)**.** *Let S be a nonempty subset of a ring R such that:*

1. *S is closed under subtraction (if $a, b \in S$, then $a - b \in S$);*
2. *S is closed under multiplication (if $a, b \in S$, then $ab \in S$).*

Then S is a subring of R.

Proof. We will show that this is equivalent to the hypotheses of theorem 2.10. This means we only need to show that closure under subtraction implies (1) S is closed under addition, (2) $0 \in S$, and (3) if $a \in S$ then $x \in S$, where $a + x = 0$.
(2): Since S is nonempty and is closed under subtraction, we have that $c \in S$ exists so that $c - c = 0 \in S$. Thus $0 \in S$.
(3): Since $-a$ is the solution of $a + x = 0$, we just need that $-a \in S$. Again, since S is closed under subtraction, we have $0 - a = -a \in S$.
(1) By part (3) above, we have that $-b \in S$, and so from closure of subtraction $a - b \in S \implies a - (-b) = a + b \in S$. Where the equality used (2.6).

□

Definition 2.15. An element a in a ring R with identity is called a *unit* if there exists $u \in R$ such that $au = 1_R = ua$. In this case, the element u is called the (multiplicative) inverse of a and is denoted a^{-1}. Note that we already defined this in 2.8.

Definition 2.16. An element a in a ring R is a **zero divisor** provided that:

1. $a \neq 0_R$.
2. There exists a nonzero element c in R such that $ac = 0_R$ or $ca = 0_R$.

Note that we already defined this in 2.9.

Theorem 2.13. *Cancellation is valid in any integral domain R: If $a \neq 0_R$ and $ab = ac$ in R, then $b = c$.*

Proof. Since $ab = ac$ and since all rings are closed under subtraction (2.12) we have

$$ab - ac = a(b - c) = 0$$

since S is an integral domain (2.13) we have $a = 0$ or $b - c = 0$, but by hypothesis $a \neq 0$, thus $b = c$. $\square$

Theorem 2.14. *Every field F is an integral domain.*

Proof. Since both fields and integral domains are commutative rings with identity, we only need to show that the existence of a solution $x \in R$ in $ax = 1$ implies that whenever $ab = 0$ then $a = 0$ or $b = 0$. Suppose $b \neq 0$ and $ab = 0$. By definition (2.14) we have $b^{-1} \in R$ such that $bb^{-1} = 1$. Then

$$a = a1 = a(bb^{-1}) = (ab)b^{-1} = 0b^{-1} = 0$$

. $\square$

Theorem 2.15. *Every finite integral domain R is a field.*

Proof. Since R is an integral domain (2.13) it has no zero divisors (2.16). Let $R' = R\ 0$ and let $f : R' \to R'$ be the mapping $f(x) = ax$ for some fixed $a \in R'$. Now if $f(x) = f(y)$ or $ax = ay$ then by cancellation for integral domains (2.13) $x = y$, thus f is injective. But since $R(R')$ is finite, we have that f is also surjective. So fixing any $a \in R$, we have $\forall y \in R'$, $\exists x \in R'$ such that $ax = y$. Letting $y = 1$ we see that $\forall a \in R'$, $\exists x \in R'$ such that $ax = 1$. $\square$

Remark 2.14. Consider the subset $\{0, 2, 4, 6, 8\}$ of $\mathbb{Z}_{10}$ along with the set $\mathbb{Z}_5 = \{0, 1, 2, 3, 4, 5\}$. Notice that the multiplication and addition amongst the subset of $\mathbb{Z}_1 0$ and amongst the elements in $\mathbb{Z}_5$ are analogous in that the only thing changing is the labels of the numbers. Meaning, for the elements of $\mathbb{Z}_5$, if we relabel 0 as 0, 1 as 6, 2 as 2, 3 as 8, and 4 as 4, we see that the two sets are actually identical (after relabeling).

The above is an example of having two structures and finding that for however multiplication and addition are defined, every element along with the stucture those elements build (with operations) can be paired off with elements of another structure. This is an isomorphism and is defined rigorously below.

Definition 2.17 (Isomorphism)**.** A ring R is isomorphic to a ring S (in symbols, $R \cong S$) if there is a function $f : R \to S$ such that all of the below hold:

1. f is injective;
2. f is surjective;
3. $f(a + b) = f(a) + f(b) \quad$ and $\quad f(ab) = f(a)f(b)$ for all $a, b \in R$.

In this case, the function f is called an **isomorphism**.

Remark 2.15. Now if we have that two rings are almost isomorphic but there does not exist a bijection amongst the elements, then we basically have only an isomorphism between the structures only. This implies the operations, the things that build the structure, must satisfy the below.

Definition 2.18 (Homomorphism)**.** Let R and S be rings. A function $f : R \to S$ is said to be a **homomorphism** if

$$f(a + b) = f(a) + f(b) \quad \text{and} \quad f(ab) = f(a)f(b) \quad \text{for all } a, b \in R.$$

Theorem 2.16. *Let $f : R \to S$ be a homomorphism of rings. Then*

1. $f(0_R) = 0_S$.
2. $f(-a) = -f(a)$ *for every* $a \in R$.
3. $f(a - b) = f(a) - f(b)$ *for all* $a, b \in R$.

If R is a ring with identity and f is surjective, then

4. S *is a ring with identity* $f(1_R)$.

5. *Whenever* u *is a unit in* R*, then* $f(u)$ *is a unit in* S *and* $f(u)^{-1} = f(u^{-1})$.

Proof. (1): Since f is a homomorphism we have $f(0_R) + f(0_R) = f(0_R + 0_R) = f(0_R) = f(0_R) + 0_S$. So this means, $f(0_R) = 0_S$.
(2): Let $a \in R$, then $f(a) + f(-a) = f(a - a) = f(0_R) = 0_S$ by (1). Since $-f(a)$ is the solution to the equation $f(a) + x = 0_S$, and since we have that $f(-a)$ is also a solution. By 2.12, we have $-f(a) = f(-a)$.
(3): $f(a-b) = f(a+(-b)) = f(a)+f(-b) = f(a)-f(b)$. Note that we used (2) on the third equality.
(4): Since f is surjective, we have that $\forall s \in S$, $\exists r \in R$ such that $s = f(r)$. Thus,

$$sf(1_R) = f(r)f(1_R) = f(r \cdot 1_R) = f(r) = s \implies f(1_R) = 1_S.$$

(5): Using (4), we see that

$$1_S = f(1_R) = f(uu^{-1}) = f(u)f(u^{-1}) \implies f(u)f(u^{-1}) = 1_S$$

So the multiplicative inverse of any $f(u)$ is $f(u^{-1})$ and since we denote the inverse of $f(u)$ as $f(u)^{-1}$, we see $f(u^{-1}) = f(u)^{-1}$.

□

Corollary 2.3. *If* $f : R \to S$ *is a homomorphism of rings, then the image of* f *is a subring of* S.

Proof. Denote the image of f by $Im(f)$. $Im(f)$ is nonempty because $0_S = f(0_R) \in Im(f)$ by theorem 2.16. Then by definition we have that if $f(a), f(b) \in Im(f)$ then $f(a)f(b) = f(ab) \in Im(f)$ and $f(a) - f(b) = f(a - b) \in Im(f)$, again by 2.16. Thus, $Im(f)$ is a subring of S by 2.7.

□

Remark 2.16. Suppose there is some property amongst the elements of R and their is an isomorphism between R and S. Then we say the property is *preserved* under the isomorphism f if that property is carried over, or, also seen in S. For example, suppose R is a commutative ring (2.11) and $f : R \to S$ is an isomorphism. Then $\forall a, b \in R$, we have $ab = ba \in R$. Therefore, in S we have

$$f(a)f(b) = f(ab) = f(ba) = f(b)f(a).$$

Which means S is also a commutative ring. So we see here that the structure of commutative rings are preserved under isomorphisms.

3 Linear Algebra

Definition 3.1. Let F be a field. A **vector space** over F is a set V equipped with two operations:

- **Vector addition**: A function $+ : V \times V \to V$ assigning to each pair $(v, w) \in V \times V$ a sum $v + w \in V$.
- **Scalar multiplication**: A function $\cdot : F \times V \to V$ assigning to each scalar $a \in F$ and vector $v \in V$ a product $av \in V$.

These operations satisfy the following axioms for all $u, v, w \in V$ and all $a, b \in F$:

1. **Axioms for Vector Addition:**
   (a) **Closure**: $v + w \in V$.
   (b) **Associativity**: $u + (v + w) = (u + v) + w$.
   (c) **Commutativity**: $v + w = w + v$.
   (d) **Existence of Additive Identity**: There exists an element $0 \in V$ such that $v + 0 = v$ for all $v \in V$.

   (e) **Existence of Additive Inverses**: For each $v \in V$, there exists $-v \in V$ such that $v + (-v) = 0$.

2. **Axioms for Scalar Multiplication:**

   (a) **Closure**: $av \in V$ for all $a \in F$ and $v \in V$.

   (b) **Distributivity over Vector Addition**: $a(v + w) = av + aw$.

   (c) **Distributivity over Scalar Addition**: $(a + b)v = av + bv$.

   (d) **Associativity**: $(ab)v = a(bv)$.

   (e) **Multiplicative Identity**: There exists a scalar $1 \in F$ such that $1v = v$ for all $v \in V$.

Definition 3.2 (Subspace)**.** Let V be a vector space, and let W be a subset of V. We define W to be a *subspace* if W satisfies the following conditions:

1. If v, w are elements of W, their sum $v + w$ is also an element of W.

2. If v is an element of W and c is a scalar, then cv is an element of W.

3. The element O of V is also an element of W.

Then W itself is a vector space. Indeed, properties **VS1** through **VS8**, being satisfied for all elements of V, are satisfied *a fortiori* for the elements of W.

Definition 3.3 (Linear Combination)**.** Let V be an arbitrary vector space, and let $v_1, \ldots, v_n$ be elements of V. Let $x_1, \ldots, x_n$ be scalars. An expression of the form

$$x_1 v_1 + \cdots + x_n v_n$$

is called a *linear combination* of $v_1, \ldots, v_n$.

Definition 3.4 (Dot Product)**.** Let $V = K^n$. Let $A, B \in K^n$ with $A = (a_1, \ldots, a_n)$ and $B = (b_1, \ldots, b_n)$. We define the *dot product* or *scalar product* as

$$A \cdot B = a_1 b_1 + \cdots + a_n b_n.$$

Remark 3.1. Geometrically we say that A and B are orthogonal MORE HERE

Definition 3.5 (Linear Independence)**.** Let $v_1, \ldots, v_n$ be vectors in a vector space. The set of vectors $\{v_1, \ldots, v_n\}$ is said to be *linearly independent* if the only solution to the equation

$$a_1 v_1 + \cdots + a_n v_n = O$$

is $a_1 = a_2 = \cdots = a_n = 0$. That is, the vectors are linearly independent if no nontrivial linear combination of them results in the zero vector.

Definition 3.6 (Basis)**.** Let V be a vector space. A set of vectors $\{v_1, \ldots, v_n\}$ in V is called a *basis* of V if:

1. The vectors $v_1, \ldots, v_n$ *generate* V, meaning that every vector in V can be written as a linear combination of $v_1, \ldots, v_n$.

2. The vectors $v_1, \ldots, v_n$ are *linearly independent*, meaning that the only solution to

$$a_1 v_1 + \cdots + a_n v_n = O$$

   is $a_1 = a_2 = \cdots = a_n = 0$.

If these conditions are satisfied, we say that $\{v_1, \ldots, v_n\}$ *forms a basis* of V.

Theorem 3.1. *Let V be a vector space. Let $v_1, \ldots, v_n$ be linearly independent elements of V. Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be scalars. Suppose that*

$$x_1 v_1 + \cdots + x_n v_n = y_1 v_1 + \cdots + y_n v_n.$$

Then $x_i = y_i$ for all $i = 1, \ldots, n$.

Theorem 3.2. *Let $\{v_1, \dots, v_n\}$ be a set of generators of a vector space V. Let $\{v_1, \dots, v_r\}$ be a maximal subset of linearly independent elements. Then $\{v_1, \dots, v_r\}$ is a basis of V.*

Definition 3.7 (Dimension of a Vector Space)**.** Let V be a vector space having a basis consisting of n elements. We define n to be the *dimension* of V. If V consists only of the zero vector O, then V does not have a basis, and we define the dimension of V to be 0.

Theorem 3.3. *Let V be a vector space, and $\{v_1, \dots, v_n\}$ a maximal set of linearly independent elements of V. Then $\{v_1, \dots, v_n\}$ is a basis of V.*

Theorem 3.4. *Let V be a vector space of dimension n, and let $v_1, \dots, v_n$ be linearly independent elements of V. Then $v_1, \dots, v_n$ constitute a basis of V.*

Proof. According to Theorem 3.3, $\{v_1, \dots, v_n\}$ is a maximal set of linearly independent elements of V. Hence it is a basis by Theorem 3.3. □

Corollary 3.1. *Let V be a vector space and let W be a subspace. If $\dim W = \dim V$, then $V = W$.*

Proof. A basis for W must also be a basis for V by Theorem 3.4. □

Corollary 3.2. *Let V be a vector space of dimension n. Let r be a positive integer with $r < n$, and let $v_1, \dots, v_r$ be linearly independent elements of V. Then one can find elements $v_{r+1}, \dots, v_n$ such that*

$$\{v_1, \dots, v_n\}$$

is a basis of V.

Theorem 3.5. *Let V be a vector space having a basis consisting of n elements. Let W be a subspace which does not consist of O alone. Then W has a basis, and the dimension of W is $\leq n$.*

Proof. Let w_1 be a nonzero element of W. If $\{w_1\}$ is not a maximal set of linearly independent elements of W, we can find an element w_2 of W such that w_1, w_2 are linearly independent. Proceeding in this manner, one element at a time, there must be an integer $m \leq n$ such that we can find linearly independent elements $w_1, w_2, \dots, w_m$, and such that

$$\{w_1, \dots, w_m\}$$

is a maximal set of linearly independent elements of W (by Theorem 3.3, we cannot go on indefinitely finding linearly independent elements, and the number of such elements is at most n). If we now use Theorem 3.3, we conclude that $\{w_1, \dots, w_m\}$ is a basis for W. □

Definition 3.8. Let V be a vector space over the field K. Let U, W be subspaces of V. We define the *sum* of U and W to be the subset of V consisting of all sums $u + w$ with $u \in U$ and $w \in W$. We denote this sum by $U + W$. It is a subspace of V. Indeed, if $u_1, u_2 \in U$ and $w_1, w_2 \in W$ then

$$(u_1 + w_1) + (u_2 + w_2) = u_1 + u_2 + w_1 + w_2 \in U + W.$$

If $c \in K$, then

$$c(u_1 + w_1) = cu_1 + cw_1 \in U + W.$$

Finally, $O + O \in W$. This proves that $U + W$ is a subspace.

We shall say that V is a *direct sum* of U and W if for every element v of V there exist *unique* elements $u \in U$ and $w \in W$ such that $v = u + w$.

Theorem 3.6. *Let V be a vector space over the field K, and let U, W be subspaces. If $U + W = V$, and if $U \cap W = \{O\}$, then V is the direct sum of U and W.*

Theorem 3.7. *Let V be a finite-dimensional vector space over the field K. Let W be a subspace. Then there exists a subspace U such that V is the direct sum of W and U.*

Theorem 3.8. *If V is a finite-dimensional vector space over K, and is the direct sum of subspaces U, W, then*

$$\dim V = \dim U + \dim W.$$

4 Analysis

Theorem 4.1 (Archimedean Property)**.** *If $x, y \in \mathbb{R}$ and $x > 0$, then there exists an $n \in \mathbb{N}$ such that*

$$nx > y.$$

Proof. Notice that $nx > y \implies n > y/x$. So if this didn't hold, we would have that $\mathbb{N}$ is bounded above. Suppose by contradiction, we have

$$\exists t \in \mathbb{R}, \forall n \in \mathbb{N}, \quad n \leq t$$

Thus there must exist a least upper bound, call it $m \in \mathbb{R}$. Then

$$\exists n \text{ such that } m - 1 \leq n \leq m \leq t \implies m \leq n.$$

This contradicts that $\exists y, x$ so that $n \leq y/x \quad \forall n \in \mathbb{N}$. Hence, the Archimedean property holds. □

Theorem 4.2 (Density of $\mathbb{Q}$ in $\mathbb{R}$)**.** *If $x, y \in \mathbb{R}$ and $x < y$, then there exists an $r \in \mathbb{Q}$ such that*

$$x < r < y.$$

Proof. Let $r = \frac{m}{n}$ and $m, n \in \mathbb{Z}$ such that $n \neq 0$ and $gcd(m, n) = 1$. Then we want to show the existence of m and n such that for any x and y,

$$x < \frac{m}{n} < y \quad \implies \quad 0 < n(y - x).$$

Then by 4.1, we have that $\exists n \in \mathbb{N}$ such that

$$1 < n(y - x) \quad \text{or} \quad \frac{1}{n} < y - x \quad \text{or} \quad nx + 1 < ny.$$

So we have that the n *scaled difference* of y and x is greater than 1, this tells me I can fit an integer m between nx and ny. To pick this m, let $S = \{k \in \mathbb{Z} \mid k > nx\}$. By 4.1, we know S is nonempty, then by the Well Ordering Axiom, we have that there exists a least element, call it m. Then $m \in S$ so $nx < m$ or $x < m/n$. Now it remains to show that $m < ny$. Since m is the least element of S, we must have $m - 1 \notin S$. Thus

$$m - 1 < nx \quad \implies \quad m < nx + 1 < ny.$$

This gives us, $m/n < y$ which proves the statement.

□

4.1 Sequences

Definition 4.1 (Sequence)**.** A *sequence* (of real numbers) is a function $x : \mathbb{N} \to \mathbb{R}$. Instead of $x(n)$, we usually denote the nth element in the sequence by x_n. To denote a sequence we write

$$\{x_n\}_{n=1}^{\infty}$$

Definition 4.2 (Bounded Sequence)**.** A sequence $\{x_n\}_{n=1}^{\infty}$ is *bounded* if there exists $M \in \mathbb{R}$ such that

$$|x_n| \leq M \quad \text{for all } n \in \mathbb{N}.$$

That is, the sequence x_n is bounded whenever the set $\{x_n \mid n \in \mathbb{N}\}$ is bounded.

Definition 4.3 (Monotone Sequence)**.** A sequence $\{x_n\}_{n=1}^{\infty}$ is *monotone increasing* if $x_n \leq x_{n+1}$ for all $n \in \mathbb{N}$. A sequence $\{x_n\}_{n=1}^{\infty}$ is *monotone decreasing* if $x_n \geq x_{n+1}$ for all $n \in \mathbb{N}$. If a sequence is either monotone increasing or monotone decreasing, we can simply say the sequence is *monotone*.

Definition 4.4 (Convergent Sequence)**.** A sequence x_n is said to *converge* to a number $x \in \mathbb{R}$ if

$$\forall \varepsilon > 0, \exists N \in \mathbb{N} \text{ such that } \forall n \geq N, |x_n - x| < \varepsilon.$$

Note that this is equivalently written $\lim_{n\to\infty} x_n = x$ or $x_n \longrightarrow x$.

Remark 4.1. The definition of a convergence sequence seems as though it does not lend itself easily to application, but a change in perspective of the definition allows you to see the usefulness. Think of it as, me and some other guy are both looking at x_n, he chooses $\varepsilon > 0$, this determines how precise our limit must be. So I then choose an $N \in \mathbb{N}$ such that x_n is always within ε of x for all n *after* the N which we specifically found given ε.

Proposition 4.1. *A convergent sequence has a unique limit.*

Proof. Suppose x_n converges to both x and y. Then by definition 4.4, we have $\forall \varepsilon > 0$, $\exists N_1 \in \mathbb{N}$ such that $\forall n \geq N_1$, $|x_n - x| < \varepsilon/2$, and for the same ε, $\exists N_2 \in \mathbb{N}$ such that $\forall n \geq N_2$, $|x_n - y| < \varepsilon/2$. Thus if we choose $N = max(N_1, N_2)$ we obtain,

$$|x - y| = |x - x_n + x_n - y| \leq |x - x_n| + |x_n - y| < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

Since $|y - x| < \varepsilon$, $\forall \varepsilon > 0$, is equivalent to $y = x$, this proves that if the limit exists, it is unique. $\square$

Exercise 4.1. *Claim: The sequence* $\{\frac{1}{n}\}_{n=1}^{\infty}$ *is convergent and converges to* 0.
To apply the definition of convergence we would need to show that for any $\varepsilon > 0$, there exists some value $N \in \mathbb{N}$ such that x_n is bounded by ε for all n after that N. In other words, we would that $\forall \varepsilon > 0, \exists N \in \mathbb{N}$ such that $\forall n \geq N$, we would have $|\frac{1}{n}| < \varepsilon \implies n > \frac{1}{\varepsilon}$. Notice this n exists by 4.1. This is how we find the N value that we use in our proof most of the time.

Exercise 4.2. *Let* (s_n) *be a sequence of non-negative real numbers and suppose* $s = \lim_{n\to\infty}$. *Then*

$$\lim_{n\to\infty} \sqrt{s_n} = \sqrt{\lim_{n\to\infty} s_n}$$

Proof. From the definition of convergence, we need to bound the magnitude of the difference of $\sqrt{s_n} - \sqrt{s}$. So we massage the expression that we are supposed to be concluding with to see if we find some bound.

$$\left|\sqrt{s_n} - \sqrt{s}\right| \implies \left|\frac{(\sqrt{s_n} - \sqrt{s})(\sqrt{s_n} + \sqrt{s})}{\sqrt{s_n} + \sqrt{s}}\right| = \left|\frac{s_n - s}{\sqrt{s_n} + \sqrt{s}}\right|$$

Since $\sqrt{s_n} \geq 0$, we have that $\left|\frac{s_n - s}{\sqrt{s_n}+\sqrt{s}}\right| \leq \left|\frac{s_n - s}{\sqrt{s}}\right|$. This is the type of expression we want, we have that $s_n - s$ along with other elements, of which we can bound, are greater than the expression we are trying to bound by ε. So we choose $N \in \mathbb{N}$ such that

$$\mid s_n - s \mid < \sqrt{s}\varepsilon \implies \frac{|s_n - s|}{\sqrt{s}} < \varepsilon \implies \left|\frac{s_n - s}{\sqrt{s_n} + \sqrt{s}}\right| < \varepsilon \implies \left|\sqrt{s_n} - \sqrt{s}\right| < \varepsilon.$$

This proves the statement. $\square$

Proposition 4.2. *Convergent sequences are bounded.*

Proof. Suppose $x_n \longrightarrow x$. Then there exists an $N \in \mathbb{N}$ such that $\forall n > N$ we have $|x_n - x| < 1$. Then for $n > N$,

$$|x_n| = |x_n - x + x| \leq |x_n - x| + |x| < 1 + |x|.$$

Now consider the set

$$M = \{|x_1|, |x_2|, \ldots, |x_{N-1}|, 1 + |x|\}.$$

Observe that M is finite. Then let

$$B = \max(\{|x_1|, |x_2|, \ldots, |x_{N-1}|, 1 + |x|\}.$$

Then for all $n \in \mathbb{N}$,

$$|x_n| \leq B.$$

This satisfies definition 4.2. $\square$

Proposition 4.3 (Algebra of Limits)**.** *Let* $\{x_n\}_{n=1}^{\infty}$ *and* $\{y_n\}_{n=1}^{\infty}$ *be convergent sequences.*

1. $\lim_{n\to\infty}(x_n + y_n) = \lim_{n\to\infty} x_n + \lim_{n\to\infty} y_n$.

2. $\lim_{n\to\infty}(x_n y_n) = (\lim_{n\to\infty} x_n)(\lim_{n\to\infty} y_n)$.

3. If $\lim_{n\to\infty} y_n \neq 0$ *and* $y_n \neq 0$ *for all* $n \in \mathbb{N}$, *then* $\lim_{n\to\infty} \frac{x_n}{y_n} = \frac{\lim_{n\to\infty} x_n}{\lim_{n\to\infty} y_n}$.

Proof. We start with (i). Suppose $\{x_n\}_{n=1}^{\infty}$ and $\{y_n\}_{n=1}^{\infty}$ are convergent sequences and write $z_n := x_n + y_n$. Let $x := \lim_{n\to\infty} x_n$, $y := \lim_{n\to\infty} y_n$, and $z := x + y$.

Let $\epsilon > 0$ be given. Find an M_1 such that for all $n \geq M_1$, we have $|x_n - x| < \epsilon/2$. Find an M_2 such that for all $n \geq M_2$, we have $|y_n - y| < \epsilon/2$. Take $M := \max\{M_1, M_2\}$. For all $n \geq M$, we have

$$|z_n - z| = |(x_n + y_n) - (x + y)|$$

$$= |x_n - x + y_n - y|$$

$$\leq |x_n - x| + |y_n - y|$$

$$< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Therefore (i) is proved. Proof of (ii) is almost identical and is left as an exercise.

Let us tackle (iii). Suppose again that $\{x_n\}_{n=1}^{\infty}$ and $\{y_n\}_{n=1}^{\infty}$ are convergent sequences and write $z_n := x_n y_n$. Let $x := \lim_{n\to\infty} x_n$, $y := \lim_{n\to\infty} y_n$, and $z := xy$.

Let $\epsilon > 0$ be given. Let $K := \max\{|x|, |y|, \epsilon/3, 1\}$. Find an M_1 such that for all $n \geq M_1$, we have $|x_n - x| < \epsilon/3K$. Find an M_2 such that for all $n \geq M_2$, we have $|y_n - y| < \epsilon/3K$. Take $M := \max\{M_1, M_2\}$. For all $n \geq M$, we have

$$|z_n - z| = |(x_n y_n) - (xy)|$$

$$= |(x_n - x + x)(y_n - y + y) - xy|$$

$$= |(x_n - x)y + x(y_n - y) + (x_n - x)(y_n - y)|$$

$$\leq |(x_n - x)y| + |x(y_n - y)| + |(x_n - x)(y_n - y)|$$

$$= |x_n - x||y| + |x||y_n - y| + |x_n - x||y_n - y|$$

$$< \frac{\epsilon}{3K}K + \frac{\epsilon}{3K}K + \frac{\epsilon}{3K}K \quad \text{(now notice that } \frac{\epsilon}{3K} \leq 1 \text{ and } K \geq 1\text{)}$$

$$\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.$$

Finally, we examine (iv). Instead of proving (iv) directly, we prove the following simpler claim:

Claim: If $\{y_n\}_{n=1}^{\infty}$ is a convergent sequence such that $\lim_{n\to\infty} y_n \neq 0$ and $y_n \neq 0$ for all $n \in \mathbb{N}$, then $\{1/y_n\}_{n=1}^{\infty}$ converges and

$$\lim_{n\to\infty} \frac{1}{y_n} = \frac{1}{\lim_{n\to\infty} y_n}.$$

Once the claim is proved, we take the sequence $\{1/y_n\}_{n=1}^{\infty}$, multiply it by the sequence $\{x_n\}_{n=1}^{\infty}$ and apply item (iii).

Proof of claim: Let $\epsilon > 0$ be given. Let $y := \lim_{n\to\infty} y_n$. As $|y| \neq 0$, then

$$\min\left\{\frac{|y|^2\epsilon}{2}, \frac{|y|}{2}\right\} > 0.$$

Find an M such that for all $n \geq M$, we have

$$|y_n - y| < \min\left\{\frac{|y|^2\epsilon}{2}, \frac{|y|}{2}\right\}.$$

For all $n \geq M$, we have $|y - y_n| < |y|/2$, and so

$$|y| = |y - y_n + y_n| \leq |y - y_n| + |y_n| < \frac{|y|}{2} + |y_n|.$$

Subtracting $|y|/2$ from both sides we obtain $|y|/2 < |y_n|$, or in other words,

$$\frac{1}{|y_n|} < \frac{2}{|y|}.$$

We finish the proof of the claim:

$$\begin{aligned}\left|\frac{1}{y_n} - \frac{1}{y}\right| &= \left|\frac{y - y_n}{y y_n}\right| \\ &= \frac{|y - y_n|}{|y||y_n|} \\ &\leq \frac{|y - y_n|}{|y|} \cdot \frac{2}{|y|} \\ &< \frac{|y|^2 \epsilon}{2} \cdot \frac{2}{|y|} \\ &= \epsilon.\end{aligned}$$

And we are done. □

Lemma 4.1 (Squeeze lemma)**.** *Let $\{a_n\}_{n=1}^{\infty}$, $\{b_n\}_{n=1}^{\infty}$, and $\{x_n\}_{n=1}^{\infty}$ be sequences such that*

$$a_n \leq x_n \leq b_n \quad \text{for all } n \in \mathbb{N}.$$

Suppose $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ converge and

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n.$$

Then $\{x_n\}_{n=1}^{\infty}$ converges and

$$\lim_{n\to\infty} x_n = \lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n.$$

Proof. Let $x := \lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n$. Let $\varepsilon > 0$ be given. Find an M_1 such that for all $n \geq M_1$, we have $|a_n - x| < \varepsilon$, and an M_2 such that for all $n \geq M_2$, we have $|b_n - x| < \varepsilon$. Set $M := \max\{M_1, M_2\}$. Suppose $n \geq M$. In particular, $x - a_n < \varepsilon$, or $x - \varepsilon < a_n$. Similarly, $b_n < x + \varepsilon$. Putting everything together, we find

$$x - \varepsilon < a_n \leq x_n \leq b_n < x + \varepsilon.$$

In other words, $-\varepsilon < x_n - x < \varepsilon$ or $|x_n - x| < \varepsilon$. So $\{x_n\}_{n=1}^{\infty}$ converges to x. □

We can also formally define divergent sequences even though we really already know from our definition of convergence.

Definition 4.5. We say x_n *diverges to infinity* if

$$\forall K \in \mathbb{R}, \exists M \in \mathbb{N}, \text{ such that } \exists n \geq M \text{ where} x_n > K.$$

This is written

$$\lim_{n\to\infty} x_n = \infty$$

Theorem 4.3 (Monotone Convergence Theorem)**.** *A monotone sequence $\{x_n\}_{n=1}^{\infty}$ is bounded if and only if it is convergent.*

Furthermore, if $\{x_n\}_{n=1}^{\infty}$ is monotone increasing and bounded, then

$$\lim_{n\to\infty} x_n = \sup\{x_n : n \in \mathbb{N}\}.$$

If $\{x_n\}_{n=1}^{\infty}$ is monotone decreasing and bounded, then

$$\lim_{n\to\infty} x_n = \inf\{x_n : n \in \mathbb{N}\}.$$

Proof. If we assume x_n is convergent, then by (4.2) we have that x_n is bounded.

Conversely, suppose x_n is monotone increasing and bounded above. Since x_n is a sequence of real numbers, by (1.23 or 1.3), or the completeness property, the least upper bound x exists. Thus for any $\varepsilon > 0$ $\exists N \in \mathbb{N}$ such that $\forall n \geq N$, $x_N \leq x - \varepsilon < x_n \leq x < x + \varepsilon \implies |x_n - x| < \varepsilon$. □

Exercise 4.3. *Let $n \in \mathbb{N}$ then,*

$$\lim_{n\to\infty} n^{1/n} = 1.$$

Proof. We want $x_n = 1 - n^{1/n}$ to converge to 0. Firstly, observe that $n^{1/n}$ is bounded below by 1. To see this, by contradiction suppose we had $n^{1/n} < 1 \implies n < 1$ which is not true for all n. Thus

$$|n^{1/n} - 1| = n^{1/n} - 1$$

This implies that we need to find n such that

$$n^{1/n} - 1 < \varepsilon \implies n < (\varepsilon + 1)^n.$$

In search of a bound, if we consider the REF binomial expansion of $(1+\varepsilon)^n$,

$$(1+\varepsilon)^n = \sum_{k=0}^{n} \binom{n}{k} \varepsilon^k = 1 + n\varepsilon + \frac{1}{2}n(n-1)\varepsilon^2 + \cdots$$

Since we only need that $n < (\varepsilon+1)^n$ and since we have $\frac{1}{2}n(n-1)\varepsilon^2 \leq (1+\varepsilon)^n$, it suffices to show that $n < \frac{1}{2}n(n-1)\varepsilon^2 \implies n > \frac{2}{\varepsilon^2} + 1$. Thus $\forall \varepsilon > 0$ choosing $N = \frac{2}{\varepsilon^2} + 2$, we have that $\forall n \geq N$

$$n > \frac{2}{\varepsilon^2} + 1 \implies n < \frac{1}{2}n(n-1)\varepsilon^2 \leq (1+\varepsilon)^n \implies n^{1/n} - 1 < \varepsilon.$$

This concludes the proof.

□

Exercise 4.4. *If $0 < c < 1$, then*

$$\lim_{n\to\infty} c^n = 0.$$

Proof. Let $L = \lim c^n$. Then $c^{n+1} = cc^n \implies L = cL \implies 0 = L(1-c)$. Since the real numbers are an integral domain REF and $c \neq 1$, we have $L = 0$. □

Remark 4.2. The idea of the proof in the next exercise uses the result of exercise 4.4. Notice if $L < 1$, then each term (since it's in absolute values) is less than the other by a ratio. But this only happens after we get to our limit, so its for all n after whatever M makes us convergent. But how exactly would I show that this sequence is a ratio (like a $(1/c)^n$) type)? This is where you are going to have to get weird. Break the sequence (mentally) into two parts, before M (meaning, before the terms are a ratio of each other) and after M (once the terms are a ratio of each other). So we could potentially express x_n using this.

Exercise 4.5 (Ratio Test for Sequences)**.** *Let $(x_n)_{n=1}^{\infty}$ be a sequence such that $x_n \neq 0$ $\forall n \in \mathbb{N}$ and such that the limit*

$$L = \lim_{n\to\infty} \frac{|x_{n+1}|}{|x_n|}$$

exists.

1. *If $L < 1$, then $\lim\limits_{n\to\infty} x_n = 0$.*

2. *If $L > 1$, then $\{x_n\}_{n=1}^{\infty}$ is unbounded.*

Proof. (1) Suppose $L < 1$. Since $\frac{|x_{n+1}|}{|x_n|} \geq 0$ for all n, we have that $L \geq 0$. Choose an $r \in \mathbb{R}$ such that $L < r < 1$. Since $r - L > 0$ we can treat $r - L$ like an ε such that, $\exists M \in \mathbb{N}$ such that $\forall n \geq M$, we have

$$\left| \frac{|x_{n+1}|}{|x_n|} - L \right| < r - L.$$

Therefore, for $n \geq M$,

$$\frac{|x_{n+1}|}{|x_n|} - L < r - L \quad \text{or} \quad \frac{|x_{n+1}|}{|x_n|} < r.$$

For $n > M$, use that each term is a multiple in $(0,1)$ of the terms before it, so we write

$$|x_n| = |x_M| \frac{|x_{M+1}|}{|x_M|} \frac{|x_{M+2}|}{|x_{M+1}|} \cdots \frac{|x_n|}{|x_{n-1}|} < |x_M| r r \cdots r = |x_M| r^{n-M} = (|x_M| r^{-M}) r^n.$$

The sequence $\{r^n\}_{n=1}^{\infty}$ converges to zero and hence $|x_M| r^{-M} r^n$ converges to zero. Since $\{x_n\}_{n=M+1}^{\infty}$ converges to zero, we have that $\{x_n\}_{n=1}^{\infty}$ converges to zero.

Now suppose $L > 1$. Pick r such that $1 < r < L$. As $L - r > 0$, there exists an $M \in \mathbb{N}$ such that for all $n \geq M$,

$$\left| \frac{|x_{n+1}|}{|x_n|} - L \right| < L - r.$$

Therefore,

$$\frac{|x_{n+1}|}{|x_n|} > r.$$

Again, for $n > M$, write

$$|x_n| = |x_M| \frac{|x_{M+1}|}{|x_M|} \frac{|x_{M+2}|}{|x_{M+1}|} \cdots \frac{|x_n|}{|x_{n-1}|} > |x_M| r r \cdots r = |x_M| r^{n-M} = (|x_M| r^{-M}) r^n.$$

The sequence $\{r^n\}_{n=1}^{\infty}$ is unbounded (since $r > 1$), and so $\{x_n\}_{n=1}^{\infty}$ cannot be bounded. Consequently, $\{x_n\}_{n=1}^{\infty}$ cannot converge. $\square$

Exercise 4.6. *If $(x_n)_{n=1}^{\infty}$ is convergent and $k \in \mathbb{N}$ then*

$$\lim_{n \to \infty} x_n^k = \left(\lim_{n \to \infty} x_n \right)^k$$

Proof. Let $\lim_{n \to \infty} x_n = x$. We aim to show that $x_n^k \to x^k$. By definition of limit, for every $\varepsilon > 0$, we must find $N \in \mathbb{N}$ such that for all $n \geq N$,

$$\left| x_n^k - x^k \right| < \varepsilon.$$

For $k \geq 1$, one can factor the difference of powers as

$$x_n^k - x^k \ = \ (x_n - x) \left(x_n^{k-1} \ + \ x_n^{k-2} \, x \ + \ \cdots + \ x_n \, x^{k-2} \ + \ x^{k-1} \right).$$

Hence

$$\left| x_n^k - x^k \right| \ \leq \ |x_n - x| \left(\left| x_n^{k-1} \right| + \left| x_n^{k-2} x \right| + \cdots + \left| x_n x^{k-2} \right| + \left| x^{k-1} \right| \right).$$

Since $x_n \to x$, there exists N_1 such that for all $n \geq N_1$, we have $\left| x_n \right| < |x| + 1$. Then each term $\left| x_n^{k-j} x^{j-1} \right|$ is at most $(|x| + 1)^{k-j} \, |x|^{j-1}$. Consequently, for $n \geq N_1$,

$$\left| x_n^{k-1} \right| + \left| x_n^{k-2} x \right| + \cdots + \left| x_n x^{k-2} \right| + \left| x^{k-1} \right| \ \leq \ k \, (|x| + 1)^{k-1}.$$

Therefore,

$$\left| x_n^k - x^k \right| \ \leq \ |x_n - x| \, k \, (|x| + 1)^{k-1} \quad \text{for all } n \geq N_1.$$

Since $x_n \to x$, there exists N_2 such that for all $n \geq N_2$, we have $\left| x_n - x \right| < \frac{\varepsilon}{k \, (|x|+1)^{k-1}}$. Setting $N = \max(N_1, N_2)$, it follows that for all $n \geq N$,

$$\left| x_n^k - x^k \right| \ \leq \ |x_n - x| \, k \, (|x| + 1)^{k-1} \ < \ \frac{\varepsilon}{k \, (|x| + 1)^{k-1}} \, k \, (|x| + 1)^{k-1} \ = \ \varepsilon.$$

Hence $x_n^k \to x^k$. $\square$

Exercise 4.7. *If $(x_n)_{n=1}^{\infty}$ is a convergent sequence and $x_n \geq 0$ and $k \in \mathbb{N}$ then*

$$\lim_{n\to\infty} x_n^{1/k} = \left(\lim_{n\to\infty} x_n\right)^{1/k}$$

Proof. Let $\lim_{n\to\infty} x_n = x$ with each $x_n \geq 0$. We wish to show $x_n^{1/k} \to x^{1/k}$. By definition of the limit, for each $\varepsilon > 0$, we must find N such that for all $n \geq N$,

$$\left| x_n^{1/k} - x^{1/k} \right| < \varepsilon.$$

For $a, b \geq 0$ and $k \geq 1$, we have

$$a^{1/k} - b^{1/k} = \frac{a-b}{a^{(k-1)/k} + a^{(k-2)/k} b^{1/k} + \cdots + b^{(k-1)/k}}.$$

Applying this with $a = x_n$ and $b = x$, we get

$$x_n^{1/k} - x^{1/k} = \frac{x_n - x}{x_n^{(k-1)/k} + x_n^{(k-2)/k} x^{1/k} + \cdots + x^{(k-1)/k}}.$$

Thus

$$\left| x_n^{1/k} - x^{1/k} \right| \leq \frac{|\, x_n - x \,|}{\min\limits_{z \in S_n} z},$$

where S_n is the set of all terms $x_n^{(k-j)/k} x^{(j-1)/k}$ that appear in the denominator. Since $x_n \to x > 0$, for large n, both x_n and x are positive and close to each other. In particular, there exists N_1 such that for $n \geq N_1$, x_n is bounded below by, say, $\frac{x}{2}$ (assuming $x > 0$). Consequently, each term in the denominator is at least $\left(\frac{x}{2}\right)^{(k-j)/k} x^{(j-1)/k}$, which is a positive constant (depending on x and k, but not on n). Denote

$$m = \min_{0 \leq j \leq k-1} \left\{ \left(\tfrac{x}{2}\right)^{\frac{k-j}{k}} x^{\frac{j-1}{k}} \right\} > 0.$$

Then for $n \geq N_1$,

$$x_n^{(k-1)/k} + x_n^{(k-2)/k} x^{1/k} + \cdots + x^{(k-1)/k} \geq k\, m.$$

Since $x_n \to x$, we also have $\left|x_n - x\right| \to 0$. Choose N_2 so that for $n \geq N_2$, $\left|x_n - x\right| < \varepsilon\, m$. Setting $N = \max(N_1, N_2)$, for $n \geq N$ we get

$$\left| x_n^{1/k} - x^{1/k} \right| \leq \frac{|x_n - x|}{k\, m} < \frac{\varepsilon\, m}{k\, m} = \frac{\varepsilon}{k}.$$

Thus $x_n^{1/k} \to x^{1/k}$.

□

Definition 4.6. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence. Let $\{n_i\}_{i=1}^{\infty}$ be a strictly increasing sequence of natural numbers, that is, $n_i < n_{i+1}$ for all $i \in \mathbb{N}$ (in other words $n_1 < n_2 < n_3 < \cdots$). The sequence

$$\{x_{n_i}\}_{i=1}^{\infty}$$

is called a *subsequence* of $\{x_n\}_{n=1}^{\infty}$.

Proposition 4.4. *If $\{x_n\}_{n=1}^{\infty}$ is a convergent sequence, then every subsequence $\{x_{n_i}\}_{i=1}^{\infty}$ is also convergent, and*

$$\lim_{n\to\infty} x_n = \lim_{i\to\infty} x_{n_i}.$$

Proof. By the definition of a subsequence (4.6), we have that $i \leq n_i$ in x_{n_i} and x_n. Then $\forall \varepsilon > 0$ $\exists N \in \mathbb{N}$ such that

$$|x_n - x| < \varepsilon \implies |x_{n_i} - x| \leq |x_n - x| < \varepsilon.$$

This concludes the proof. □

Definition 4.7. Let $\{x_n\}_{n=1}^\infty$ be a bounded sequence. Define the sequences $\{a_n\}_{n=1}^\infty$ and $\{b_n\}_{n=1}^\infty$ by

$$a_n := \sup\{x_k : k \geq n\}, \quad b_n := \inf\{x_k : k \geq n\}.$$

Define, if the limits exist,

$$\limsup_{n\to\infty} x_n := \lim_{n\to\infty} a_n, \quad \liminf_{n\to\infty} x_n := \lim_{n\to\infty} b_n.$$

In words, the supremum of a sequence x_n is the supremum of all x_n's after the nth value that we are currently on. So the limit of the supremum is the supremum of all terms to come. Notice that the sequence a_n is monotone decreasing (4.3) since with each passing n, the value that is the supremum of all x_n to come, can only decrease.

Theorem 4.4. *If $\{x_n\}_{n=1}^\infty$ is a bounded sequence, then there exists a subsequence $\{x_{n_k}\}_{k=1}^\infty$ such that*

$$\lim_{k\to\infty} x_{n_k} = \limsup_{n\to\infty} x_n.$$

Similarly, there exists a (perhaps different) subsequence $\{x_{m_k}\}_{k=1}^\infty$ such that

$$\lim_{k\to\infty} x_{m_k} = \liminf_{n\to\infty} x_n.$$

Remark 4.3. In the below proof, we are trying to find an x_{n_i} that converges to the same limit as the supremum. So we want the

Proof. Define $a_n = \sup\{x_k : k \geq n\}$. Let $x := \limsup_{n\to\infty} x_n = \lim_{n\to\infty} a_n$. We define the subsequence inductively. Let $n_1 = 1$, meaning $x_{n_1} = x_n$, and suppose $n_1, n_2, \ldots, n_{k-1}$ are defined for some $k \geq 2$. Since the subsequences index $(n_k)_{k=1}^\infty$ is strictly increasing, $n_k \geq n_{k-1}+1$, pick an $m \geq n_{k-1}+1$ such that

$$a_{n_k+1} - x_m < \frac{1}{k}.$$

Such an m exists as a_{n_k+1} is a supremum of the set $\{x_\ell : \ell \geq n_{k-1}+1\}$ and hence there are elements of the sequence arbitrarily close (or even possibly equal) to the supremum. Set $n_k = m$. The subsequence $\{x_{n_k}\}_{k=1}^\infty$ is defined. Next, we must prove that it converges to x. For all $k \geq 2$, we have $a_{n_k+1} \geq a_{n_k}$ (why?) and $a_{n_k} \geq x_{n_k}$. Therefore, for every $k \geq 2$,

$$|a_{n_k} - x_{n_k}| = a_{n_k} - x_{n_k} \leq a_{n_k+1} - x_{n_k} < \frac{1}{k}.$$

Let us show that $\{x_{n_k}\}_{k=1}^\infty$ converges to x. Note that the subsequence need not be monotone. Let $\epsilon > 0$ be given. As $\{a_n\}_{n=1}^\infty$ converges to x, the subsequence $\{a_{n_k}\}_{k=1}^\infty$ converges to x. Thus, there exists an $M_1 \in \mathbb{N}$ such that for all $k \geq M_1$, we have

$$|a_{n_k} - x| < \frac{\epsilon}{2}.$$

Find an $M_2 \in \mathbb{N}$ such that

$$\frac{1}{M_2} \leq \frac{\epsilon}{2}.$$

Take $M := \max\{M_1, M_2\}$. For all $k \geq M$,

$$|x - x_{n_k}| = |a_{n_k} - x_{n_k} + x - a_{n_k}| \leq |a_{n_k} - x_{n_k}| + |x - a_{n_k}| \leq \frac{1}{M_2} + \frac{\epsilon}{2} \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

$\square$

Exercise 4.8. *Let $S \subset \mathbb{R}$ be a nonempty bounded set. Then there exist monotone sequences $\{x_n\}_{n=1}^\infty$ and $\{y_n\}_{n=1}^\infty$ such that $x_n, y_n \in S$ and*

$$\sup S = \lim_{n\to\infty} x_n \quad \textit{and} \quad \inf S = \lim_{n\to\infty} y_n.$$

Proposition 4.5. *Let $\{x_n\}_{n=1}^{\infty}$ be a bounded sequence. Then $\{x_n\}_{n=1}^{\infty}$ converges if and only if*

$$\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n.$$

Furthermore, if $\{x_n\}_{n=1}^{\infty}$ converges, then

$$\lim_{n\to\infty} x_n = \liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n.$$

Proof. Let a_n and b_n be as in definition (4.7). In particular, for all $n \in \mathbb{N}$,

$$b_n \le x_n \le a_n.$$

First suppose $\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n$. Then $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ both converge to the same limit. By the squeeze lemma (4.1), $\{x_n\}_{n=1}^{\infty}$ converges and

$$\lim_{n\to\infty} b_n = \lim_{n\to\infty} x_n = \lim_{n\to\infty} a_n.$$

Now suppose $\{x_n\}_{n=1}^{\infty}$ converges to x. By (4.4), there exists a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ converging to $\limsup_{n\to\infty} x_n$. As $\{x_n\}_{n=1}^{\infty}$ converges to x, every subsequence converges to x and so $\limsup_{n\to\infty} x_n = \lim_{k\to\infty} x_{n_k} = x$. Similarly, $\liminf_{n\to\infty} x_n = x$. □

Exercise 4.9. *Suppose $(x_n)_{n=1}^{\infty}$ is a bounded sequence and $(x_{n_k})_{k=1}^{\infty}$ is a subsequence. Then*

$$\liminf_{n\to\infty} x_n \le \liminf_{k\to\infty} x_{n_k} \le \limsup_{k\to\infty} x_{n_k} \le \limsup_{n\to\infty} x_n$$

Proof. We want to prove that $\limsup_{k\to\infty} x_{n_k} \le \limsup_{n\to\infty} x_n$. Define $a_n := \sup\{x_k : k \ge n\}$ as usual. Also define $c_n := \sup\{x_{n_k} : k \ge n\}$. It is not true that $\{c_n\}_{n=1}^{\infty}$ is necessarily a subsequence of $\{a_n\}_{n=1}^{\infty}$. However, as $n_k \ge k$ for all k, we have $\{x_{n_k} : k \ge n\} \subset \{x_k : k \ge n\}$. A supremum of a subset is less than or equal to the supremum of the set, and therefore

$$c_n \le a_n \quad \text{for all } n \implies \lim_{n\to\infty} c_n \le \lim_{n\to\infty} a_n,$$

which is the desired conclusion. □

Exercise 4.10. *A bounded sequence $(x_n)_{n=1}^{\infty}$ converges to x $\iff$ every subsequence $(x_{n_k})_{k=1}^{\infty}$ converges to x.*

Proof. Suppose $x_n \to x$. Let $\{x_{n_k}\}$ be any subsequence of $\{x_n\}$. By definition of a subsequence, the terms x_{n_k} are simply some of the terms x_n, but in the same order. Hence, for every $\varepsilon > 0$, if we can make $|x_n - x| < \varepsilon$ for all n large enough, the same is true for the subsequence terms x_{n_k}. Given $\varepsilon > 0$, there exists N such that for all $n \ge N$, $|x_n - x| < \varepsilon$. Since $n_k \ge k$ (in fact $n_k \to \infty$ as $k \to \infty$), eventually $n_k \ge N$. Thus $\left|x_{n_k} - x\right| < \varepsilon$ for all k sufficiently large. This shows $x_{n_k} \to x$. Therefore, every convergent subsequence converges to the same limit x.

For the other direction, we prove the contrapositive. Assume $\{x_n\}$ does not converge to x. By the definition of the limit, this means there exists some $\varepsilon_0 > 0$ such that for **every** $N \in \mathbb{N}$, there is an $n \ge N$ with

$$\left|x_n - x\right| \;\ge\; \varepsilon_0.$$

Intuitively, no matter how far out in the sequence you go, there are terms that stay at least ε_0-away from x. We build a subsequence $\{x_{n_k}\}$ that remains outside the ε_0-neighborhood of x. Start by letting n_1 be an index such that $\left|x_{n_1} - x\right| \ge \varepsilon_0$. Having chosen n_k, pick $n_{k+1} > n_k$ such that $\left|x_{n_{k+1}} - x\right| \ge \varepsilon_0$. This is always possible by the non-convergence assumption.

Thus, for each k,

$$\left|x_{n_k} - x\right| \;\ge\; \varepsilon_0.$$

Therefore, the subsequence $\{x_{n_k}\}$ does not converge to x. Since $\{x_n\}$ is given to be bounded, any subsequence (including $\{x_{n_k}\}$) is also bounded. By the Bolzano–Weierstrass Theorem (or any equivalent result about bounded sequences in $\mathbb{R}$), the subsequence $\{x_{n_k}\}$ has a further convergent subsequence, say $\{x_{n_{k_j}}\}$. Let

$$\lim_{j\to\infty} x_{n_{k_j}} \;=\; y$$

for some $y \in \mathbb{R}$. By hypothesis, every convergent subsequence of $\{x_n\}$ must converge to x. Thus we must have $y = x$. On one hand, we deduce $x_{n_{k_j}} \to x$. On the other hand, from the construction we know each $x_{n_{k_j}}$ satisfies $\left|x_{n_{k_j}} - x\right| \geq \varepsilon_0$. No sequence whose terms stay at a fixed distance $\varepsilon_0 > 0$ from x can converge to x. This contradiction arises from assuming $\{x_n\}$ does not converge to x.

Hence, it must be that $\{x_n\}$ does converge to x. This completes the proof of the contrapositive and thus the original statement. $\square$

Definition 4.8 (Subsequential Limit)**.** Let $(x_n)_{n=1}^{\infty}$ be a sequence. A *subsequential limit* is any extended real number that is the limit of some subsequence of $(x_n)_{n=1}^{\infty}$.

Theorem 4.5 (Bolzano–Weierstrass)**.** *Suppose a sequence $\{x_n\}_{n=1}^{\infty}$ of real numbers is bounded. Then there exists a convergent subsequence $\{x_{n_i}\}_{i=1}^{\infty}$.*

Proof. As the sequence is bounded, then there exist two numbers $a_1 < b_1$ such that $a_1 \leq x_n \leq b_1$ for all $n \in \mathbb{N}$. We will define a subsequence $\{x_{n_i}\}_{i=1}^{\infty}$ and two sequences $\{a_i\}_{i=1}^{\infty}$ and $\{b_i\}_{i=1}^{\infty}$ such that $\{a_i\}_{i=1}^{\infty}$ is monotone increasing, $\{b_i\}_{i=1}^{\infty}$ is monotone decreasing, $a_i \leq x_{n_i} \leq b_i$ and such that $\lim_{i\to\infty} a_i = \lim_{i\to\infty} b_i$. That x_{n_i} converges then follows by the squeeze lemma (4.1).

We define the sequences inductively. We will define the sequences so that for all i, we have $a_i < b_i$, and that $x_n \in [a_i, b_i]$ for infinitely many $n \in \mathbb{N}$. We have already defined a_1 and b_1. We take $n_1 := 1$, that is $x_{n_1} = x_1$. Suppose that up to some $k \in \mathbb{N}$, we have defined the subsequence $x_{n_1}, x_{n_2}, \ldots, x_{n_k}$, and the sequences $a_1, a_2, \ldots, a_k$ and $b_1, b_2, \ldots, b_k$. Let

$$y := \frac{a_k + b_k}{2}.$$

Clearly $a_k < y < b_k$. If there exist infinitely many $j \in \mathbb{N}$ such that $x_j \in [a_k, y]$, then set $a_{k+1} := a_k$, $b_{k+1} := y$, and pick $n_{k+1} > n_k$ such that $x_{n_{k+1}} \in [a_k, y]$. If there are not infinitely many j such that $x_j \in [a_k, y]$, then it must be true that there are infinitely many $j \in \mathbb{N}$ such that $x_j \in [y, b_k]$. In this case pick $a_{k+1} := y$, $b_{k+1} := b_k$, and pick $n_{k+1} > n_k$ such that $x_{n_{k+1}} \in [y, b_k]$.

We now have the sequences defined. What is left to prove is that $\lim_{i\to\infty} a_i = \lim_{i\to\infty} b_i$. The limits exist as the sequences are monotone. In the construction, $b_i - a_i$ is cut in half in each step. Therefore,

$$b_{i+1} - a_{i+1} = \frac{b_i - a_i}{2}.$$

By induction,

$$b_i - a_i = \frac{b_1 - a_1}{2^{i-1}}.$$

Let $x := \lim_{i\to\infty} a_i$. As $\{a_i\}_{i=1}^{\infty}$ is monotone,

$$x = \sup\{a_i : i \in \mathbb{N}\}.$$

Let $y := \lim_{i\to\infty} b_i = \inf\{b_i : i \in \mathbb{N}\}$. Since $a_i < b_i$ for all i, then $x \leq y$. As the sequences are monotone, then for all i, we have

$$y - x \leq b_i - a_i = \frac{b_1 - a_1}{2^{i-1}}.$$

Because $\frac{b_1 - a_1}{2^{i-1}}$ is arbitrarily small and $y - x \geq 0$, we have $y - x = 0$. By squeeze lemma (4.1), this concludes the proof. $\square$

Exercise 4.11. *Let (s_n) be any sequence of nonzero real numbers. Then we have*

$$\liminf \left|\frac{s_{n+1}}{s_n}\right| \leq \liminf |s_n|^{1/n} \leq \limsup |s_n|^{1/n} \leq \limsup \left|\frac{s_{n+1}}{s_n}\right|.$$

Exercise 4.12. *If $\lim \left|\frac{s_{n+1}}{s_n}\right|$ exists and equals L then $\lim |s_n|^{1/n}$ exists and equals L.*

Definition 4.9 (Cauchy Sequence)**.** A sequence $\{x_n\}_{n=1}^{\infty}$ is a *Cauchy sequence* if for every $\varepsilon > 0$, there exists an $M \in \mathbb{N}$ such that for all $n \geq M$ and all $k \geq M$, we have

$$|x_n - x_k| < \varepsilon.$$

Lemma 4.2. *If a sequence is Cauchy, then it is bounded.*

Theorem 4.6 (Convergent $\iff$ Cauchy)**.** *A sequence of real numbers is Cauchy $\iff$ the sequence is convergent.*

4.2 Series

So we have built a good understanding of sequences, to make sense of what is about to come, consider the following example. Suppose you have an infinite number of people, each of them representing a number (like their age or something), if we give a calculator to the first person and tell them to put their age in then tell the next person to put their age in and tell the same person after them to do so. At any moment if we stop this process, say at person k, then the number on the calculator is the kth value of our sequence, where the sequence represents the sum of a sequence of numbers.

Definition 4.10 (Series)**.** Given a sequence $(x_n)_{n=1}^{\infty}$, we define

$$\sum_{n=1}^{\infty} x_n$$

as a *series*. A series *converges* if the sequence $(s_k)_{k=1}^{\infty}$, called the partial sums, and defined by

$$s_k = \sum_{n=1}^{k} x_n = x_1 + x_2 + \cdots + x_k$$

converges. So a series converges if

$$\sum_{n=1}^{\infty} x_n = \lim_{k\to\infty} \sum_{n=1}^{k} x_n.$$

Proposition 4.6 (Geometric Series)**.** *Suppose $-1 < r < 1$. Then the geometric series $\sum_{n=0}^{\infty} r^n$ converges, and*

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}$$

Exercise 4.13. *Let $\sum_{n=1}^{\infty} x_n$ be a series and let $M \in \mathbb{N}$. Then*

$$\sum_{n=1}^{\infty} x_n \text{ converges} \iff \sum_{n=M}^{\infty} x_n \text{ converges.}$$

Definition 4.11 (Cauchy Series)**.** A series $\sum_{n=1}^{\infty} x_n$ is said to be *Cauchy* if the sequence of the partial sums $(s_n)_{n=1}^{\infty}$ is a Cauchy sequence.
Note that a series is convergent if and only if it is Cauchy 4.6.

Exercise 4.14. *If a series $\sum_{n=1}^{\infty} x_n$ converges, then $\lim x_n = 0$.*

Proposition 4.7 (Linearity of Series)**.** *Let $\alpha \in \mathbb{R}$ and $\sum_{n=1}^{\infty} x_n$ and $\sum_{n=1}^{\infty} y_n$ be convergent series. Then*

1. *$\sum_{n=1}^{\infty} \alpha x_n$ is a convergent series and*

$$\sum_{n=1}^{\infty} \alpha x_n = \alpha \sum_{n=1}^{\infty} x_n.$$

2. $\sum_{n=1}^{\infty}(x_n + y_n)$ *is a convergent series and*

$$\sum_{n=1}^{\infty}(x_n + y_n) = \left(\sum_{n=1}^{\infty} x_n\right) + \left(\sum_{n=1}^{\infty} y_n\right).$$

Proof. For the first item, we simply write the kth partial sum

$$\sum_{n=1}^{k} \alpha x_n = \alpha \left(\sum_{n=1}^{k} x_n\right).$$

We look at the right-hand side and note that the constant multiple of a convergent sequence is convergent. Hence, we take the limit of both sides to obtain the result.

For the second item, we also look at the kth partial sum

$$\sum_{n=1}^{k}(x_n + y_n) = \left(\sum_{n=1}^{k} x_n\right) + \left(\sum_{n=1}^{k} y_n\right).$$

We look at the right-hand side and note that the sum of convergent sequences is convergent. Hence, we take the limit of both sides to obtain the proposition. □

Proposition 4.8. *If $x_n \geq 0$ for all n, then $\sum_{n=1}^{\infty} x_n$ converges if and only if the sequence of partial sums is bounded above.*

Definition 4.12 (Absolute Convergence)**.** A series $\sum_{n=1}^{\infty} x_n$ *converges absolutely* if the series $\sum_{n=1}^{\infty} |x_n|$ converges. If a series converges, but does not converge absolutely, we say it *converges conditionally*

Proposition 4.9. *If the series $\sum_{n=1}^{\infty} x_n$ converges absolutely, then it converges.*

Proposition 4.10 (Comparison Test)**.** *Let $\sum_{n=1}^{\infty} x_n$ and $\sum_{n=1}^{\infty} y_n$ be series such that $0 \leq x_n \leq y_n$ for all $n \in \mathbb{N}$.*

1. *If $\sum_{n=1}^{\infty} y_n$ converges, then so does $\sum_{n=1}^{\infty} x_n$.*
2. *If $\sum_{n=1}^{\infty} x_n$ diverges, then so does $\sum_{n=1}^{\infty} y_n$.*

Proposition 4.11 (P-Series)**.** *(p-series or the p-test). For $p \in \mathbb{R}$, the series*

$$\sum_{n=1}^{\infty} \frac{1}{n^p}$$

converges if and only if $p > 1$.

Proposition 4.12 (Root Test)**.** *Let $\sum_{n=1}^{\infty} x_n$ be a series and let*

$$L = \limsup_{n\to\infty} |x_n|^{1/n}.$$

1. *If $L < 1$, then $\sum_{n=1}^{\infty} x_n$ converges absolutely.*
2. *If $L > 1$, then $\sum_{n=1}^{\infty} x_n$ diverges.*

Proposition 4.13 (Ratio Test)**.** *Let $\sum_{n=1}^{\infty} x_n$ be a series, $x_n \neq 0$ for all n, and such that*

1. *If $\limsup_{n\to\infty} \left|\frac{x_{n+1}}{x_n}\right| = L < 1$, then $\sum_{n=1}^{\infty} x_n$ converges absolutely.*
2. *If $\liminf_{n\to\infty} \left|\frac{x_{n+1}}{x_n}\right| = L > 1$, then $\sum_{n=1}^{\infty} x_n$ diverges.*

Proposition 4.14 (Alternating Series Test)**.** *Let $\{x_n\}_{n=1}^{\infty}$ be a monotone decreasing sequence of positive real numbers such that $\lim_{n\to\infty} x_n = 0$. Then the alternating series*

$$\sum_{n=1}^{\infty}(-1)^n x_n$$

converges.

4.3 Continuity

Remark 4.4. Now we will generalize the results up to now so we can apply it to mappings between sets.

Definition 4.13 (Cluster Point)**.** A number $x \in \mathbb{R}$ is called a cluster point of a set $S \subset \mathbb{R}$ if for every $\epsilon > 0$, the set

$$(x - \epsilon, x + \epsilon) \cap (S \setminus \{x\})$$

is nonempty.

Equivalently, x is a cluster point of S if for every $\epsilon > 0$, there exists some $y \in S$ such that $y \neq x$ and $|x - y| < \epsilon$.

A cluster point of S need not belong to S.

Proposition 4.15. *Let $S \subset \mathbb{R}$. Then $x \in \mathbb{R}$ is a cluster point of S if and only if there exists a convergent sequence of numbers $\{x_n\}_{n=1}^{\infty}$ such that $x_n \neq x$ and $x_n \in S$ for all n, and $\lim_{n\to\infty} x_n = x$.*

Definition 4.14. Let $f : S \to \mathbb{R}$ be a function and c a cluster point of $S \subset \mathbb{R}$. Suppose there exists an $L \in \mathbb{R}$ and for every $\epsilon > 0$, there exists a $\delta > 0$ such that whenever $x \in S \setminus \{c\}$ and $|x - c| < \delta$, we have

$$|f(x) - L| < \epsilon.$$

We then say $f(x)$ *converges* to L as x goes to c, and we write

$$f(x) \to L \quad \text{as} \quad x \to c.$$

We say L is a *limit* of $f(x)$ as x goes to c, and if L is unique (it is), we write

$$\lim_{x\to c} f(x) := L.$$

If no such L exists, then we say that the limit does not exist or that f *diverges* at c.

Proposition 4.16. *Let c be a cluster point of $S \subset \mathbb{R}$ and let $f : S \to \mathbb{R}$ be a function such that $f(x)$ converges as x goes to c. Then the limit of $f(x)$ as x goes to c is unique.*

Lemma 4.3. *Let $S \subset \mathbb{R}$, let c be a cluster point of S, let $f : S \to \mathbb{R}$ be a function, and let $L \in \mathbb{R}$. Then $f(x) \to L$ as $x \to c$ if and only if for every sequence $\{x_n\}_{n=1}^{\infty}$ such that $x_n \in S \setminus \{c\}$ for all n, and such that $\lim_{n\to\infty} x_n = c$, we have that the sequence $\{f(x_n)\}_{n=1}^{\infty}$ converges to L.*

Proof. Suppose $f(x) \to L$ as $x \to c$, and $\{x_n\}_{n=1}^{\infty}$ is a sequence such that $x_n \in S\setminus\{c\}$ and $\lim_{n\to\infty} x_n = c$. We wish to show that $\{f(x_n)\}_{n=1}^{\infty}$ converges to L. Let $\epsilon > 0$ be given. Find a $\delta > 0$ such that if $x \in S \setminus \{c\}$ and $|x - c| < \delta$, then $|f(x) - L| < \epsilon$. As $\{x_n\}_{n=1}^{\infty}$ converges to c, find an M such that for $n \geq M$, we have that $|x_n - c| < \delta$. Therefore, for $n \geq M$,

$$|f(x_n) - L| < \epsilon.$$

Thus $\{f(x_n)\}_{n=1}^{\infty}$ converges to L.

For the other direction, we use proof by contrapositive. Suppose it is not true that $f(x) \to L$ as $x \to c$. The negation of the definition is that there exists an $\epsilon > 0$ such that for every $\delta > 0$ there exists an $x \in S \setminus \{c\}$, where $|x - c| < \delta$ and $|f(x) - L| \geq \epsilon$.

Let us use $1/n$ for δ in the statement above to construct a sequence $\{x_n\}_{n=1}^{\infty}$. We have that there exists an $\epsilon > 0$ such that for every n, there exists a point $x_n \in S \setminus \{c\}$, where $|x_n - c| < 1/n$ and $|f(x_n) - L| \geq \epsilon$. The sequence $\{x_n\}_{n=1}^{\infty}$ just constructed converges to c, but the sequence $\{f(x_n)\}_{n=1}^{\infty}$ does not converge to L. And we are done. □

Proposition 4.17. *Let $S \subset \mathbb{R}$ and let c be a cluster point of S. Suppose $f : S \to \mathbb{R}$ and $g : S \to \mathbb{R}$ are functions such that the limits of $f(x)$ and $g(x)$ as x goes to c both exist, and*

$$f(x) \leq g(x) \quad \textit{for all } x \in S \setminus \{c\}.$$

Then

$$\lim_{x\to c} f(x) \leq \lim_{x\to c} g(x).$$

Proposition 4.18. *Let $S \subset \mathbb{R}$ and let c be a cluster point of S. Suppose $f : S \to \mathbb{R}$, $g : S \to \mathbb{R}$, and $h : S \to \mathbb{R}$ are functions such that*

$$f(x) \leq g(x) \leq h(x) \quad \textit{for all } x \in S \setminus \{c\}.$$

Suppose the limits of $f(x)$ and $h(x)$ as x goes to c both exist, and

$$\lim_{x\to c} f(x) = \lim_{x\to c} h(x).$$

Then the limit of $g(x)$ as x goes to c exists and

$$\lim_{x\to c} g(x) = \lim_{x\to c} f(x) = \lim_{x\to c} h(x).$$

Proposition 4.19. *Let $S \subset \mathbb{R}$ and let c be a cluster point of S. Suppose $f : S \to \mathbb{R}$ and $g : S \to \mathbb{R}$ are functions such that the limits of $f(x)$ and $g(x)$ as x goes to c both exist. Then*

1. $\lim_{x\to c}(f(x) + g(x)) = (\lim_{x\to c} f(x)) + (\lim_{x\to c} g(x))$.
2. $\lim_{x\to c}(f(x) - g(x)) = \lim_{x\to c} f(x) - \lim_{x\to c} g(x)$.
3. $\lim_{x\to c}(f(x)g(x)) = (\lim_{x\to c} f(x)) (\lim_{x\to c} g(x))$.
4. *If $\lim_{x\to c} g(x) \neq 0$ and $g(x) \neq 0$ for all $x \in S \setminus \{c\}$, then*
$$\lim_{x\to c} \frac{f(x)}{g(x)} = \frac{\lim_{x\to c} f(x)}{\lim_{x\to c} g(x)}.$$

Proposition 4.20. *Let $S \subset \mathbb{R}$ and let c be a cluster point of S. Suppose $f : S \to \mathbb{R}$ is a function such that the limit of $f(x)$ as x goes to c exists. Then*

$$\lim_{x\to c} |f(x)| = \left|\lim_{x\to c} f(x)\right|.$$

Definition 4.15. Let $f : S \to \mathbb{R}$ be a function and $A \subset S$. Define the function $f|_A : A \to \mathbb{R}$ by

$$f|_A(x) := f(x) \quad \text{for } x \in A.$$

We call $f|_A$ the *restriction* of f to A.

Proposition 4.21. *Let $S \subset \mathbb{R}$, $c \in \mathbb{R}$, and let $f : S \to \mathbb{R}$ be a function. Suppose $A \subset S$ is such that there is some $\alpha > 0$ such that*

$$(A \setminus \{c\}) \cap (c - \alpha, c + \alpha) = (S \setminus \{c\}) \cap (c - \alpha, c + \alpha).$$

1. *The point c is a cluster point of A if and only if c is a cluster point of S.*
2. *Supposing c is a cluster point of S, then $f(x) \to L$ as $x \to c$ if and only if $f|_A(x) \to L$ as $x \to c$.*

Proposition 4.22. *Let $S \subset \mathbb{R}$ be such that c is a cluster point of both $S \cap (-\infty, c)$ and $S \cap (c, \infty)$, let $f : S \to \mathbb{R}$ be a function, and let $L \in \mathbb{R}$. Then c is a cluster point of S and*

$$\lim_{x\to c} f(x) = L \quad \textit{if and only if} \quad \lim_{x\to c^-} f(x) = \lim_{x\to c^+} f(x) = L.$$

Definition 4.16. Suppose $S \subset \mathbb{R}$ and $c \in S$. We say $f : S \to \mathbb{R}$ is *continuous* at c if for every $\epsilon > 0$ there is a $\delta > 0$ such that whenever $x \in S$ and $|x - c| < \delta$, we have $|f(x) - f(c)| < \epsilon$.

When $f : S \to \mathbb{R}$ is continuous at all $c \in S$, then we simply say f is a *continuous function*.

Proposition 4.23. *Consider a function $f : S \to \mathbb{R}$ defined on a set $S \subset \mathbb{R}$ and let $c \in S$. Then:*

1. *If c is not a cluster point of S, then f is continuous at c.*
2. *If c is a cluster point of S, then f is continuous at c if and only if the limit of $f(x)$ as $x \to c$ exists and*
$$\lim_{x\to c} f(x) = f(c).$$

3. *The function f is continuous at c if and only if for every sequence $\{x_n\}_{n=1}^{\infty}$ where $x_n \in S$ and $\lim_{n\to\infty} x_n = c$, the sequence $\{f(x_n)\}_{n=1}^{\infty}$ converges to $f(c)$.*

Proposition 4.24. *Let $f : \mathbb{R} \to \mathbb{R}$ be a polynomial. That is,*

$$f(x) = a_d x^d + a_{d-1}x^{d-1} + \cdots + a_1 x + a_0,$$

for some constants $a_0, a_1, \ldots, a_d$. Then f is continuous.

Proposition 4.25. *Let $f : S \to \mathbb{R}$ and $g : S \to \mathbb{R}$ be functions continuous at $c \in S$.*

1. *The function $h : S \to \mathbb{R}$ defined by $h(x) := f(x) + g(x)$ is continuous at c.*
2. *The function $h : S \to \mathbb{R}$ defined by $h(x) := f(x) - g(x)$ is continuous at c.*
3. *The function $h : S \to \mathbb{R}$ defined by $h(x) := f(x)g(x)$ is continuous at c.*
4. *If $g(x) \neq 0$ for all $x \in S$, the function $h : S \to \mathbb{R}$ given by $h(x) := \frac{f(x)}{g(x)}$ is continuous at c.*

Proposition 4.26. *Let $A, B \subset \mathbb{R}$ and $f : B \to \mathbb{R}$ and $g : A \to B$ be functions. If g is continuous at $c \in A$ and f is continuous at $g(c)$, then $f \circ g : A \to \mathbb{R}$ is continuous at c.*

Proposition 4.27. *Let $f : S \to \mathbb{R}$ be a function and $c \in S$. Suppose there exists a sequence $\{x_n\}_{n=1}^{\infty}$, where $x_n \in S$ for all n, and $\lim_{n\to\infty} x_n = c$ such that $\{f(x_n)\}_{n=1}^{\infty}$ does not converge to $f(c)$. Then f is discontinuous at c.*

Lemma 4.4. *A continuous function $f : [a, b] \to \mathbb{R}$ is bounded.*

Theorem 4.7 (Minimum-maximum theorem / Extreme value theorem)**.** *A continuous function $f : [a, b] \to \mathbb{R}$ achieves both an absolute minimum and an absolute maximum on $[a, b]$.*

Lemma 4.5. *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Suppose $f(a) < 0$ and $f(b) > 0$. Then there exists a number $c \in (a, b)$ such that $f(c) = 0$.*

Theorem 4.8 (Bolzano's Intermediate Value Theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Suppose $y \in \mathbb{R}$ is such that $f(a) < y < f(b)$ or $f(a) > y > f(b)$. Then there exists a $c \in (a, b)$ such that $f(c) = y$.*

5 Probability

Remark 5.1. To build a mathematical model of uncertainty and randomness we start by thinking what it is we will be measuring. We want to be able to measure the likelihood of some event happening. So we will measure events. What's an event though? It is an occurrence of something and this something is comprised of smaller events that the event we are concerned with is comprised of. These smaller events we will call outcomes, or singleton event. So for example, lets say I want to know the probability that my plane crashes. The event of a plane crash is composed of infinitely many outcomes... the event maybe consists of an outcome where a mechanic overlooked something, then something wobbled in just the right way, then the plane took a certain turn which caused some screw to loosen, ..., then the plane crashed. So the screw loosening is one outcome, the plane turning is another, etc. all these outcomes make up the event where a plane crashes. So we will consider sets of outcomes, we will call these events.

Remark 5.2. The below theorems will show us how to build these events, how to count the number of outcomes in the events.

Theorem 5.1. *Let $X_1, X_2, \ldots, X_n$ be finite sets with cardinalities $|X_1|, |X_2|, \ldots, |X_n|$. If a process consists of making sequential choices such that:*

- *The first choice is made from X_1,*
- *The second choice is made from X_2,*
- $\ldots$,

- *The nth choice is made from X_n,*

where the number of choices at each stage is independent of previous choices, then the total number of ways to complete the process is:

$$|X_1| \cdot |X_2| \cdots |X_n| = \prod_{i=1}^{n} |X_i|.$$

Theorem 5.2. *Let n and k be nonnegative integers with $0 \leq k \leq n$. The number of distinct subsets of size k that a set of size n has is given by the binomial coefficient*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Theorem 5.3. *For any integer $n \geq 0$ and any real or complex numbers a,b,*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k},$$

Theorem 5.4. *The number of ways to arrange n distinct objects in a sequence is*

$$P(n) = n! = n(n-1)(n-2)\cdots 2 \cdot 1$$

The number of ways to select and arrange k objects from n distinct objects is

$$P(n,k) = \frac{n!}{(n-k)!}.$$

5.1 Axioms of Probability

Remark 5.3. So how will we define the abstract space we will be working in so that we can effectively measure the likelihood of events? So consider some event A that you want to know the likelihood of. If the event A is possible, then it should also be possible that A^c, meaning, we should be able to measure both of these. So lets call the space $\mathcal{A}$, we will add sets to $\mathcal{A}$ that we think should be possible to measure if it is to be possible to measure A. So we have $A \in \mathcal{A}$ and $A^c \in \mathcal{A}$. If $A \in \mathcal{A}$ and $B \in \mathcal{A}$ then we should be able to measure $A \cup B$ and $A \cap B$, so we include all intersections and unions of possible events in $\mathcal{A}$.

Definition 5.1 (Algebra and σ-algebra)**.** Let Ω be an abstract space. Let 2^Ω denote all subsets of Ω. With $\mathcal{A}$ being a subset of 2^Ω. Then $\mathcal{A}$ is an algebra if it satisfies $(1), (2)$, and (3). $\mathcal{A}$ is a σ-algebra if it satisfies $(1), (2)$, and (4).

1. $\emptyset \in \mathcal{A}$ and $\Omega \in \mathcal{A}$
2. If $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$.
3. If the finite sequence of events $A_1, A_2, \ldots, A_n \in \mathcal{A}$ then $\bigcup_{i=1}^n A_i \in \mathcal{A}$ and $\bigcap_{i=1}^n A_i \in \mathcal{A}$.
4. If the countable sequence of events $A_1, A_2, \cdots \in \mathcal{A}$ then $\bigcup_{i=1}^\infty A_i \in \mathcal{A}$ and $\bigcap_{i=1}^\infty A_i \in \mathcal{A}$.

Remark 5.4. If $C \subset 2^\Omega$, then the σ-algebra generated by C, denoted $\sigma(C)$, is the smallest σ-algebra containing C.

We choose b_n to be strictly increasing so that the $, b_n]$ part of the interval $(a_n, b_n]$ converges to $, b_n)$. This is what allows us to have that $(a,b) = \bigcup_{n=1}^\infty (a_n, b_n]$.

Theorem 5.5 (Borel σ-algebra)**.** *If $\Omega = \mathbb{R}$, the Borel σ-algebra is the σ-algebra generated by open sets (or equivalently closed sets). Then the Borel σ-algebra can be generated by intervals of the form $(-\infty, a]$, where $a \in \mathbb{Q}$.*

Proof. Let C denote all open intervals. Since every open set in $\mathbb{R}$ is the countable union of open intervals, we have $\sigma(C) =$ the Borel σ-algebra of $\mathbb{R}$. Let D denote all intervals of the form $(-\infty, a]$, where $a \in \mathbb{Q}$. Let $(a, b) \in C$, and let $(a_n)_{n\geq 1}^{\infty}$ be a sequence of rationals decreasing to a and $(b_n)_{n\geq 1}^{\infty}$ be a sequence of rationals strictly increasing to b. Then

$$(a,b) = \bigcup_{n=1}^{\infty} (a_n, b_n] = \bigcup_{n=1}^{\infty} ((-\infty, b_n] \cap (a_n, \infty)) = \bigcup_{n=1}^{\infty} ((-\infty, b_n] \cap (-\infty, a_n]^c)$$

Since the right most expression is of the form of D and since we have that any element of C is equivalent to an element of D, we have $C \subset \sigma(D)$, hence $\sigma(C) \subset \sigma(D)$. However since $(-\infty, a]$ contains all its limit points, we know each element of D is a closed set, since closed sets are Borel sets, we have that $\sigma(D)$ is contained in the Borel sets $\mathcal{B}$. Thus we have

$$\mathcal{B} = \sigma(C) \subset \sigma(D) \subset \mathcal{B},$$

and hence $\sigma(D) = \mathcal{B}$. □

Remark 5.5. So the theorem above shows that when our sample space is the real numbers, or any space with the proper topology, we can generate the σ-algebra we define the probability measure on by using intervals of the form $(-\infty, a]$, where $a \in \mathbb{Q}$. Since $a \in \mathbb{Q}$, we have made the σ-algebra from countable sets. Which is what we needed since before we knew that open sets $\mathcal{C}$ could cover any set, we had to show that the countable collection of $(-\infty, a]$ could also cover (and thus measure) any set.

Remark 5.6. For our actual probability measure, we need an event that is guaranteed and an event that is impossible, so we include $\emptyset \in \mathcal{A}$ and $\Omega \in \mathcal{A}$ since we want $P(\emptyset) = 0$ and $P(\Omega) = 1$. We also would want that if we measure two events A and B where A and B share no outcomes $A \cap B = \emptyset$, then we want that $P(A \cup B) = P(A) + P(B)$.

Definition 5.2 (Probability Measure)**.** A probability measure defined on a σ-algebra $\mathcal{A}$ of Ω is a function $P : \mathcal{A} \to [0, 1]$ that satisfies

1. $P(\Omega) = 1$
2. For every pairwise disjoint ($A_n \cap A_m = \emptyset$ whenever $n \neq m$) countable sequence $(A_n)_{n\geq 1}$ of elements of $\mathcal{A}$, we have
$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Definition 5.3. (Indicator Function) If $A \in 2^{\Omega}$, then the indicator function $1_A(\omega)$ be given by

$$1_A(\omega) = \begin{cases} 1 & \textit{if } \omega \in A, \\ 0 & \textit{if } \omega \notin A. \end{cases}$$

We say $A_n \in \mathcal{A}$ converges to A if $\lim_{n\to\infty} 1_{A_n}(\omega) = 1_A(\omega)\ \forall \omega \in \Omega$.

Remark 5.7. A few comments and clarifications about the definition above. Since 2^{Ω} is not necessarily all measurable, meaning, the σ-algebra $\mathcal{A}$ may not include every subset of Ω. So for any singleton outcome $\omega \in \Omega$, we have if $\forall \omega \in A$, $\omega \in A_n$ as $n \to \infty$ then A_n converges to A. Note this is precisely what $\lim_{n\to\infty} 1_{A_n}(\omega) = 1_A(\omega)\ \forall \omega \in \Omega$ is stating.

Remark 5.8. So if we can define convergence of a sequence of sets, then we must have some conception of convergene of supremum and infimum. How will we define these? We want the supremum of a set to be the elements (outcomes) that are in **at least one** of the *infinite events*, that is, for all events events past some nth event, I want to know what is in **any** of the events that come after this one, then letting $n \to \infty$ we see that the elements remaining are in **at least one** of the *infinite events*. Whereas, we want the infimum to be the *smaller* set, when compared to the supremum. So instead of considering all elements that are in **any** event past the nth (we will again let $n \to \infty$) event, we will consider the elements that are in **every single** event past this nth one. From this, it is easy to see that the infimum is a subset of the supremum, which is what we wanted. We also want that when these are equivalent, the sequence of events converges.

Definition 5.4 (Supremum and Infimum of Sequence of Sets)**.** Let A_n be a sequence of sets. If $A_n \in \mathcal{A}$ $\forall n \in \mathbb{N}$ then define

$$\limsup_{n\to\infty} A_n = \cap_{n=1}^{\infty} \cup_{m\geq n} A_m$$
$$\liminf_{n\to\infty} A_n = \cup_{n=1}^{\infty} \cap_{m\geq n} A_m.$$

Lemma 5.1. *Let $\mathcal{A}$ be a σ-algebra and $(A_n)_{n\geq 1}^{\infty}$ be a sequence of sets in $\mathcal{A}$. Then,*

$$\liminf_{n\to\infty} A_n \in \mathcal{A}, \quad \limsup_{n\to\infty} A_n \in \mathcal{A}, \quad \text{and } \liminf_{n\to\infty} A_n \subseteq \limsup_{n\to\infty} A_n$$

Proof. By definition (5.1), the σ-algebra $\mathcal{A}$ is closed under countable unions and intersections. since $A_n \in \mathcal{A}$, we have that for any fixed n, $\cap_{k\geq n} A_k \in \mathcal{A}$. Then countably infinite many unions of this is also in $\mathcal{A}$. That is,

$$\liminf_{n\to\infty} A_n = \cup_{n=1}^{\infty} \cap_{k\geq n} A_k \in \mathcal{A}.$$

Similarly,

$$\limsup_{n\to\infty} A_n = \cup_{n=1}^{\infty} \cap_{k\geq n} A_k \in \mathcal{A}.$$

Now suppose $x \in \liminf_{n\to\infty} A_n = \cup_{n=1}^{\infty} \cap_{k\geq n} A_k$, then for some $N \in \mathbb{N}$, $x \in A_k$, $\forall k \geq n$. Thus $x \in \limsup_{n\to\infty} A_n = \cup_{n=1}^{\infty} \cap_{k\geq n} A_k$ since x is in all such A_k where $k \geq n$ and $\limsup A_n$ only requires that x be in at least one. Therefore,

$$\liminf_{n\to\infty} A_n \subseteq \limsup_{n\to\infty} A_n.$$

□

Lemma 5.2. *Let $\mathcal{A}$ be a σ-algebra and $(A_n)_{n\geq 1}^{\infty}$ be a sequence of sets in $\mathcal{A}$. Then,*

$$\lim_{n\to\infty} A_n = A \iff \limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n = A$$

Proof. Suppose $x \in A = \limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n$. Then,

$$x \in \cup_{n=1}^{\infty} \cap_{k\geq n} A_k \implies \exists N \in \mathbb{N} \text{ such that } x \in A_k,\ \forall k \geq N.$$
$$\text{Then if we had } x \notin \cup_{n=1}^{\infty} \cap_{k\geq n} A_k \implies \exists N \in \mathbb{N} \text{ such that } x \notin A_k,\ \forall k \geq N.$$

□

Theorem 5.6 (Continuity of Probability Measure)**.** *Let P be a probability measure, and let A_n be a sequence of events in the σ-algebra $\mathcal{A}$ which converges to A. Then $A \in \mathcal{A}$ and $\lim_{n\to\infty} P(A_n) = P(A)$.*

Proof. Define $\limsup A_n$ and $\liminf A_n$ as definition (5.4). By lemma 5.1, we have $\limsup_{n\to\infty} A_n \in \mathcal{A}$ and $\liminf_{n\to\infty} A_n \in \mathcal{A}$. So by hypothesis, A_n converges to A, then from lemma 5.2,

$$\lim_{n\to\infty} 1_{A_n} = 1_A, \quad \forall\omega \iff A = \limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n$$

□

Theorem 5.7. *Let $A_1, A_2, \ldots, A_n$ be events, then*

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{1\leq i_1\leq i_2\leq n} P(A_{i_1} \cap A_{i_2})$$
$$+ \sum_{1\leq i_1\leq i_2\leq i_3\leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \sum_{1\leq i_1<i_2\leq i_3<i_4\leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3} \cap A_{i_4})$$
$$+ \cdots + (-1)^{n+1} P(A_1 \cap \cdots \cap A_n) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1\leq i_1<\cdots<i_k\leq n} P(A_{i_1} \cap \cdots \cap A_{i_k})$$

Definition 5.5 (Monotone Sequence of Sets)**.** A sequence of events $(A_n)_{n\geq 1}^{\infty}$ is said to be an *monotone increasing* sequence of sets if

$$A_1 \subseteq A_2 \subseteq \cdots \subseteq A_k \subseteq A_{k+1} \subseteq \cdots$$

Similarly, a sequence of sets $(A_n)_{n\geq 1}^{\infty}$ is said to be a *monotone decreasing* sequence if

$$A_1 \supseteq A_2 \supseteq \cdots \supseteq A_k \supseteq A_{k+1} \supseteq \cdots$$

Further, if an increasing sequence $(A_n)_{n\geq 1}^{\infty}$ converges to some event A, then we write $A_n \uparrow A$ and we have $A = \cup_{n\geq 1}^{\infty} A_n$. Similarly, if $(A_n)_{n\geq 1}^{\infty}$ decreases to A then we write $A_n \downarrow A$, with $A = \cap_{n\geq 1}^{\infty} A_n$.

Theorem 5.8. *Let $\mathcal{A}$ be a σ-algebra and let $(A_n)_{n\geq 1}^{\infty} \in \mathcal{A}$ be a sequence of sets. Suppose $P : \mathcal{A} \to [0,1]$ is a probability measure. Then the following are equivalent,*

1. *Axiom (2) of definition (5.2)*
2. $A_n \downarrow A \implies P(A_n) \downarrow P(A)$.
3. $A_n \uparrow A \implies P(A_n) \uparrow P(A)$

Remark 5.9. In the (3) $\implies$ (1) proof, note that we assume A_n is pairwise disjoint because that is what needs to be satisfied, by the definition of the probability measure.

Proof. (2) $\iff$ (3): Suppose $A_n \uparrow A$ and $P(A_n) \uparrow P(A)$. Then $A_n^c \downarrow A^c$ and $P(A_n^c) \downarrow P(A^c)$. But since $P(A_n^c) = 1 - P(A_n)$, proving (3) $\iff$ (2) suffices.
(3) $\implies$ (1): Suppose $A_n \uparrow A$ and $P(A_n) \uparrow P(A)$. Also, assume A_n is *pairwise disjoint*, meaning $\forall i, j \in [n]$, where $i \neq j$, we have $A_i \cap A_j$. Let $B_n = \cup_{p\geq 1}^{n} A_p$ and let $B = \cup_{n\geq 1}^{\infty} A_n$. Then by axiom (2) of the probability measure (5.2), we have $P(B_n) = \sum_{p=1}^{n} A_p$. Then as $n \to \infty$, we have that $P(B_n) \uparrow P(B)$, so $P(B_n)$ is increasing sequence, increasing to $P(B)$ since $P(A_n) \uparrow P(A)$ so $P(B_n) = P(\cup_{n\geq 1}^{\infty} A_n)$.
(1) $\iff$ (3): Suppose A_n is a sequence increasing to A. Define the sequence $(B_n)_{n\geq 1}^{\infty}$

$$\begin{aligned} B_1 &= A_1 \\ B_2 &= A_2 \setminus A_1 \\ &\vdots \\ B_k &= A_k \setminus A_{k-1} \\ &\vdots \end{aligned}$$

Remark 5.10. Since A_n is increasing, we have that the A_{k-1} set contains every set before it, so letting $B_k = A_k \setminus A_{k-1}$ for every k ensures each B_k contains only the elements that A_k provided. Then since probabilities are nonnegative, we see that B_n is monotone increasing.

Then we have $A = \cup_{i=1}^{\infty} B_i$ and $B_n \cap B_m = \emptyset$ whenever $m \neq n$, meaning B_n is pairwise disjoint. Thus from (1),

$$P(A) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i)$$

But since we also have

$$P(A_n) = \sum_{i=1}^{n} P(B_i)$$

thus we have $P(A_n) \uparrow P(A)$. □

Exercise 5.1. *Let $A_i \in \mathcal{A}$ be a sequence of events. Show that*

$$P(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$$

Remark 5.11. Suppose we wanted to determine the probability of some event but we want to update this probability given

Definition 5.6. Let B be an event in the sample space Ω such that $P(B) > 0$. Then for all events A the *conditional probability* of A given B is defined as

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

Definition 5.7. A collection of events $(A_i)_{i \in I}$ is an independent collection if for every finite subset J of I, one has

$$P(\cap_{i \in J} A_i) = \prod_{i \in J} P(A_i).$$

If the above condition is satisfied for for the whole collection, we say the collection $(A_i)_{i \in I}$ is mutually independent. Also, if A_i and A_j are independent $\forall i, j$ with $i \neq j$, that is if any two events you pick from the collection $(A_i)_{i \in I}$ are independent, then the collection is pariwise independent.

Exercise 5.2. *If A and B are independent, so also are A and B^c, A^c and B, and A^c and B^C.*

Proposition 5.1. *If $A_1, A_2, \ldots, A_n \in \mathcal{A}$ and if $P(A_1 \cap \cdots \cap A_{n-1}) > 0$, then*

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \ldots P(A_n \mid A_1 \cap \cdots \cap A_n).$$

Definition 5.8. A countable collection of events $B_1, \ldots, B_n$ are a *partition* of Ω if the sets B_i are pairwise disjoint and together they make up Ω. That is, for all i and j, $B_i \cap B_j = \emptyset$ whenever $i \neq j$ and $\bigcup_{i=1}^{n} B_i = \Omega$

Proposition 5.2. *Suppose that $B_1, \ldots, B_n$ is a partition of Ω with $P(B_i) > 0$ for $i = 1, \ldots, n$. Then for any event A we have*

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A \mid B_i)P(B_i).$$

Theorem 5.9. *Let $B_1, B_2, \ldots, B_n$ be a partition of the sample space Ω such that each $P(B_i) > 0$. Then for any event A with $P(A) > 0$, and for any $k = 1, \ldots, n$, we have:*

$$P(B_k \mid A) = \frac{P(AB_k)}{P(A)} = \frac{P(A \mid B_k)P(B_k)}{\sum_{i=1}^{n} P(A \mid B_i)P(B_i)}.$$

Definition 5.9. Let $A_1, A_2, \ldots, A_n$ and B be events with $P(B) > 0$. Then $A_1, A_2, \ldots, A_n$ are *conditionally independent, given B*, if the following condition holds:

For any $k \in \{2, \ldots, n\}$ and indices $1 \leq i_1 < i_2 < \cdots < i_k \leq n$,

$$P(A_{i_1} A_{i_2} \ldots A_{i_k} \mid B) = P(A_{i_1} \mid B)P(A_{i_2} \mid B) \cdots P(A_{i_k} \mid B).$$

Theorem 5.10. *(a) A probability on the countable set Ω is characterized by its values on the atoms:*

$$p_\omega = P(\omega), \quad \omega \in \Omega.$$

(b) Let $(p_\omega)_{\omega \in \Omega}$ be a family of real numbers indexed by Ω. Then there exists a unique probability P such that $P(\omega) = p_\omega$ if and only if $p_\omega \geq 0$ and $\sum_{\omega \in \Omega} p_\omega = 1$

Definition 5.10. A random variable is a measurble function $X : \Omega \to \mathbb{R}$ such that for all Borel measurable sets $B \subseteq \mathbb{R}$, the preimage of B is an event in $\mathcal{A}$, that is

$$X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\} \in \mathcal{F}.$$

This means that X is $\mathcal{A}$-measurable, ensuring that we can compute probabilities of the form $P(X \in B)$

Remark 5.12. So a random variable inputs events or outcomes and outputs a real number, then the probability measure will assign probabilities in $[0, 1]$ to the values of X. We can then define the distribution of X by

$$P^X(A) = P(\omega \mid X(w) \in A) = P(X^{-1}(A)) = P(X \in A)$$

This is completely determined by the following

$$p_j^X = P(X = j) = \sum_{\omega \mid X(w) = j} \quad \text{and } P_X(A) = \sum_{j \in A} p_j^X$$

Definition 5.11. Let X be a real-valued random variable on a countable space Ω. The expectation of X, denoted $E(X)$, is defined to be

$$E(X) = \sum_{\omega} X(\omega)p_{\omega} \quad \text{or} \quad \int_{-\infty}^{\infty} X(\omega)p_{\omega}d\omega??$$

provided this sum converges. Notice that if the random variable is discrete we use the finite sum, if it is continuous, we use the continuous sum.

Definition 5.12. The nth moment of the random variable X is the expectation $E(X^n)$.

$$E(X^n) = \sum_{\omega} X^n(\omega)p_{\omega} \quad \text{or} \quad \int_{-\infty}^{\infty} X^n(\omega)P(X(\omega))d\omega$$

Theorem 5.11. *Let $h : \mathbb{R} \to [0, \infty)$ be a nonnegative function and let X be a real valued random variable. Then*

$$P(\{\omega \mid h(X(\omega)) \geq a\}) \leq \frac{E(H(X))}{a}, \quad \forall a > 0.$$

Corollary 5.1 (Markovs Inequality)**.**

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

Definition 5.13. Let X be a real valued random variable with $X^2 \in \mathcal{L}^1$ where $\mathcal{L}^1$ is the space of real valued random variables on $(\Omega, \mathcal{A}, P)$. The variance of X is defined to be

$$\sigma^2 = \sigma_X^2 = E((X - E(X))^2) = E(X^2) - (E(X))^2$$

The standard deviation of X, σ_X, is the nonnegative square root of the variance.

Corollary 5.2 (**Chebyshev's Inequality**)**.** *If X^2 is in $\mathcal{L}^1$, then for $a > 0$ we have*

1. $P(\{|X| \geq a\}) \leq \frac{EX^2}{a^2}$
2. $P(\{|X - E(X)| \geq a\}) \leq \frac{\sigma_X^2}{a^2}$

Definition 5.14 (**Binomial Distribution**)**.** Let n be a positive integer and $0 \leq p \leq 1$. A random variable X has the *binomialdistribution* with parameters n and p if the possible values of X are $\{0, 1, \ldots, n\}$ and the probabilities are

$$P(\{X = k\}) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, \ldots n.$$

This is denoted $X\ Bin(n, p)$.

Definition 5.15 (**Geometric Distribution**)**.** A random variable X follows a Geometric distribution with parameter p (success probability per trial) if the probability of k independent trials till a success on the kth trial is given by,

$$P(X = k) = (1-p)^{k-1}p, \quad k = 1, 2, 3, \ldots$$

Definition 5.16 (**Hypergeometric Distribution**)**.** A hypergeometric random variable represents the number of successes of size n, drawn without replacement from a population of size N that contains K successes. The PMF is given by

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \quad \max(0, n - (N - K)) \leq k \leq \min(n, K).$$

Definition 5.17 (**Poisson Distribution**)**.** A Poisson random variable models the number of events occurring in a fixed interval of time or space, under the assumption that events occur independently and at a constant average rate λ. A random variable X follows a Poisson distribution with rate parameter $\lambda > 0$ if

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots.$$

Definition 5.18 (Normal Distribution). A random variable X follows a Normal distribution with mean μ and variance σ^2, written as $X \sim \mathcal{N}(\mu, \sigma^2)$, if its probability density function (PDF) is

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

6 Advanced Risk and Portfolio Management

6.1 Data Science

6.1.1 Probabilistic Framework

6.1.2 Mean-Covariance Framework

In this framework, we model randomness by measuring only two characteristics of the random variable. We consider only the mean $E(\mathbf{X})$ and the covarince $Cv(\mathbf{X})$. The expectation gives us the location of our random variable in the multidimensional environment we model it in, and the covariance gives us the amount of dispersion in this random variable with each of the dimensions we define the space to be. Perhaps a better way of seeing this, is to notice that the first and second order terms of the taylor expansion of the characteristic function are fully characterized by the mean and covariance.

We will class random variables based on their first two mooments, μ and σ^2. We will then consider affine (linear) transformations of the reference variable for a given class. For example, suppose we have a random variable $\mathbf{X}$ and we transform it into $\mathbf{Y} = \mathbf{a} + \mathbf{b}\mathbf{X}$ which amounts to a rotation, scaling, and translation of $\mathbf{X}$. Since the expectation is linear, this gives us the handy property seen below, the expectation will only act on $\mathbf{X}$,

$$\underbrace{\begin{pmatrix} \mathbb{E}\{Y_1\} \\ \vdots \\ \mathbb{E}\{Y_{\bar{k}}\} \end{pmatrix}}_{\mathbb{E}\{\boldsymbol{Y}\}} = \underbrace{\begin{pmatrix} a_1 \\ \vdots \\ a_{\bar{k}} \end{pmatrix}}_{\boldsymbol{a}} + \underbrace{\begin{pmatrix} b_{1,1} & \cdots & b_{1,\bar{n}} \\ \vdots & \ddots & \vdots \\ b_{\bar{k},1} & \cdots & b_{\bar{k},\bar{n}} \end{pmatrix}}_{\boldsymbol{b}} \underbrace{\begin{pmatrix} \mathbb{E}\{X_1\} \\ \vdots \\ \mathbb{E}\{X_{\bar{n}}\} \end{pmatrix}}_{\mathbb{E}\{\boldsymbol{X}\}}. \tag{1}$$

So we give the following definitions

Definition 6.1. Given a probability space $(\Omega, \mathcal{F}, P)$ (**??**) with a random variable X (**??**), integration with respect to the probability measure (**??**), yields the expectation

$$\mathbb{E}\{X\} \equiv \int_\Omega X(\omega) d\mathbb{P}\{\omega\}. \tag{2}$$

More specifically, the expectation applied to the indicator function $1_\mathcal{E}$ for a given set $\mathcal{E}$ is the probability of the event $\mathcal{E}$ itself

$$1_{\boldsymbol{x}\in\mathcal{E}} \equiv 1_\mathcal{E}(\boldsymbol{x}) \equiv \begin{cases} 0 & \text{if } \boldsymbol{x} \notin \mathcal{D} \\ 1 & \text{if } \boldsymbol{x} \in \mathcal{E} \end{cases} \implies \mathbb{E}\{1_\mathcal{E}\} = \mathbb{P}\{\mathcal{E}\}.$$

Note that the indicator $1_\mathcal{E}$ is a special type of random variable, and thus the expectation is well defined. We commonly use the disitrbution of a random variable (**??**) to calculate the expectation as

$$\mathbb{E}\{X\} = \int_{-\infty}^{+\infty} x dF_X(x). \tag{3}$$

Where $dF_X(x)$ is the pdf (**??**) of the random variable.
The mean vector, given below, is the weighted average of all possible outcomes where the weights are the likelihoods. Better said, it is the center of mass of the distribution. Each $\mathbb{E}\{\mathbf{X}_n\}$ is the mean of the nth marginal variable $\mathbf{X}_n$

$$\mathbb{E}\{\boldsymbol{X}\} \equiv \begin{pmatrix} \mathbb{E}\{X_1\} \\ \vdots \\ \mathbb{E}\{X_n\} \\ \vdots \\ \mathbb{E}\{X_{\bar{n}}\} \end{pmatrix}, \tag{4}$$

The mean vector is a functional (**??**) of the distribution $F_X(x)$ since it inputs the distribution and outputs a vector.

$$\mathbb{C}v\{\boldsymbol{X}\} \equiv \begin{pmatrix} \mathbb{V}\{X_1\} & \mathbb{C}v\{X_1, X_2\} & \cdots & \mathbb{C}v\{X_1, X_{\bar{n}}\} \\ \mathbb{C}v\{X_2, X_1\} & \mathbb{V}\{X_2\} & \cdots & \mathbb{C}v\{X_2, X_{\bar{n}}\} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{C}v\{X_{\bar{n}}, X_1\} & \mathbb{C}v\{X_{\bar{n}}, X_2\} & \cdots & \mathbb{V}\{X_{\bar{n}}\} \end{pmatrix}. \tag{5}$$

6.1.3 Linear Models

6.1.4 Machine Learning

6.1.5 Estimation

6.1.6 Inference

6.1.7 Sequential Decisions

6.2 Quantiative Finance

6.2.1 Financial Engineering

6.2.2 Risk Management

6.2.3 Portfolio Management