Results

March 28, 2025

Contents

1 Learning

2 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 2.1 (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 2.2. 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 2.1. *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 2.3. 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 2.4. 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 2.1 (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* $n \in \mathbb{N}$.

Definition 2.5. 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 2.6. 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 2.2. *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 $1 > 0$.

Definition 2.7. 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 2.8 (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.*

Proposition 2.3. *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.*

Proposition 2.4. *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:*

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

Proposition 2.5 (The Supremum is the least upper bound)**.** *Let $S \subset \mathbb{R}$ be nonempty, and $L \in \mathbb{R} \cup \{\infty, -\infty\}$. Then*

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

Proposition 2.6. *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.

Proposition 2.7. *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.*

Proposition 2.8. *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 2.9. 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 2.10. 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 2.2. *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 2.9. *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 2.10. *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 2.11. 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 2.12. 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 2.11. *If $f : A \to B$ and $g : B \to C$ are bijective functions, then $f \circ g$ is bijective.*

Definition 2.13. 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 2.14. We write
$$|A| \le |B|$$
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 $|A| \le |B|$, but A and B do not have the same cardinality.
If $|A| \le |\mathbb{N}$ then we say that A is countable. If $|A| = |\mathbb{R}|$ then we say that A is uncountable.

Theorem 2.3. *If there exists a bijective function between two sets A and B, then we have that the cardanalities, 2.13, are equivalenet.*

Proposition 2.12. *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 2.1.*

Proposition 2.13. *The set $\mathbb{Z}$ is countable*

Proposition 2.14. *Every infinite subset of a countable set is also countable*

Proposition 2.15. *If A and B are countable, then $A \times B$ is countable*

Theorem 2.4. *The set $\mathbb{Q}$ is countable*

Theorem 2.5. *The open interval $(0, 1)$ of real numbers is uncountable.*

Theorem 2.6. $|(0, 1)| = |\mathbb{R}|$

Theorem 2.7. $|\mathcal{P}(A)| = |2^A|$

Lemma 2.1. *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 2.8. *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 2.9 (**Schröder-Bernstein Theorem**)**.** *If A and B are sets such that $|A| \le |B|$ and $|B| \le |A|$, then $|A| = |B|$.*

Theorem 2.10. $|\mathcal{P}(\mathbb{N})| = |\mathbb{R}|$

2.1 Metric Spaces

Definition 2.15. 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. $d(x, y) \ge 0$ (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 2.2. *(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 2.16. *Let (X, d) be a metric space and $Y \subset X$. Then the restriction $d|_{Y \times Y}$ is a metric on Y.*

Definition 2.16. 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 2.17. 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 2.18. 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 2.19. 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 2.17. *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 2.18. *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 2.19. *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 2.20. *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 2.21. *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 2.20. 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 2.22. *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 2.23. *A nonempty set $S \subset \mathbb{R}$ is connected if and only if S is an interval or a single point.*

Definition 2.21. 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 2.24. *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 2.25. *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 2.22. 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 2.26. *Let (X,d) be a metric space and $A \subset X$. Then A° is open and ∂A is closed.*

Proposition 2.27. *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 2.1. *Let (X,d) be a metric space and $A \subset X$. Then*

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

Proposition 2.28. *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 $x_n \in U$.*

A closed set contains the limits of its convergent sequences.

Proposition 2.29. *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 2.23. 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 2.30. *The space $\mathbb{R}^n$ with the standard metric is a complete metric space.*

Proposition 2.31. *The space of continuous functions $C([a, b], \mathbb{R})$ with the uniform norm as metric is a complete metric space.*

Definition 2.24. 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 2.32. *Let (X, d) be a metric space. If $K \subset X$ is compact, then K is closed and bounded.*

Lemma 2.3. *(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 2.11. *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 2.33. *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 2.12. *(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 2.25. 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 2.34. *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 2.4. *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 2.13. *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.*

Lemma 2.5. *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 2.14. *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.*

3 Algebra

Definition 3.1. 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 3.1 (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 $d \mid c_n$.

Remark 3.1. Since m/d is a solution, plug it into the polynomial. Then distributing your power of n and multiplying by d^n you will be able to rearrange to show that m divides c_0.
This result can be used to show that a number is a real number by letting $x =$ the number we want to show is a rational then rearrange to get a polynomial on one side and 0 on the other. Then using the result above we can see if the number we originally let $x =$ is a rational solution.

3.1 Divisibility in $\mathbb{Z}$

Definition 3.2 (Well Ordering Axiom)**.** Every nonempty subset of the set of nonnegative integers contains a smallest element.

Definition 3.3 ($\mathbb{Z}$)**.** 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 $b = 0$.

Theorem 3.2 (Division Algorithm)**.** *Let $a, b \in \mathbb{Z}$ with $b > 0$. Then there exist unique $q, r \in \mathbb{Z}$ such that*

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

Remark 3.2. Consider $S = \{a - bx \geq 0\}$. Start by showing S is nonempty by choosing $x = -|a|$, then rederive the form of S from this chosen x and show that it is greater than 0. Then by well ordering axiom, let r be the least positive element. Then show that $r < b$ and that r and q are unique. Remember to use absolute values for the uniqueness and recall $|r_2 - r_1| < b$.

Definition 3.4 (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 $c \leq d$.

Theorem 3.3 (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 3.3. Consider the set $S = \{am + bn\}$. Show nonempty and existence of nonnegative elements by letting $m = a$ and $n = b$. Let d be the minimum positive element by well ordering axiom. Show that d fits the definition of gcd, note that you will show $d \mid a$ by using the form given by division algo, you just need to show that $r = 0$. Once you set up 3.2, you can substitute your given expression of d because you then have 2 terms of a and a term of b, this is exactly the form of S. Thus $r \in S$ but $r < d$, since r, by the requirement of division alg, must be greater than or equal to 0, we have that $r = 0$. Then show by contradiction that d is in fact the least element.
Notice we found that the gcd is the smallest positive element. This means the gcd divides any linear combination of a and b.
Referenced in: 3.9

Proposition 3.1. *Let $a, b, c \in \mathbb{Z}$. If $a \mid bc$ and $\gcd(a, b) = 1$, then $a \mid c$.*

Remark 3.4. The gcd of 1 implies the form of Bezout. Then multiplying by c allows us to show that c is a multiple of a.
Notice the property of coprime, or just prime in general, integers here: When a doesnt share any factors with b, if a divides any multiple of b then we know a must divide the number multiplying b for the sole purpose of it having no factors to share with b.

Proposition 3.2. *Let $a, b, c \in \mathbb{Z}$. Suppose $\gcd(a, b) = 1$. If $a|c$ and $b|c$, then $ab|c$.*

Proposition 3.3. *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. $\gcd(ta, tb) = t \gcd(a, b) \quad$ for $t > 0$
3. $\gcd(a, c) = 1 \implies \gcd(ab, c) = \gcd(b, c)$

Remark 3.5. (3): Notice that the prime factorization of $d = \gcd(ab, c)$ must divide both ab and c. So if d shares any primes with a, then we have that those primes are also shared by c, so $\gcd(a, c) > 1$ which is a contradiction.
This is exactly the result we would expect for the exact same reason the proof worked: adding factors of a number that cannot share any factors with c tells us the $d =$ "the divisor that contains all divisors of both ab and c.
(2) If you let $d = \gcd(ta, tb)$, then you have $\frac{d}{t} \mid a$ and $\frac{d}{t} \mid b$. Then let $m = gcd(a, b)$ so we have that $\frac{d}{t} \mid m$ or $d \mid mt$. Then using 3.3, show that $mt \mid d$.
This seems like it wouldnt be true because the gcd of ta and tb obviously must include the greatest factor of t and any additional factors a and b might have. Where the RHS is only the factors of a and b. The only thing missing on the RHS is the contribution of every single factor of t.

Proposition 3.4. *A positive integer is divisible by 3 $\iff$ the sum of its digits is divisible by 3.*

Theorem 3.4. *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 3.6. Using 3.1, we see that p beign prime implies that property. The converse is obvious by contrapositive.

Theorem 3.5 (Fundamental Theoerem of Arithmetic)**.** *Every integer $n \neq 0, 1, -1$ has a unique prime factorization.*

Remark 3.7. Show every integer is either prime or has a prime factorization by contradiction. To show uniqueness of the factorization, show by contradiction that if two integers had the same factorization, then we would have something like $p_1 p_2 \cdots p_n = q_1 q_2 \cdots q_n$. But this means, $p_1(p_2 \cdots p_n) = q_1(q_2 \cdots q_n)$, so either p_1 divides q_1 or it divides the other integer. Since they are all prime, using 3.4 we can show they are equivalent.

Proposition 3.5. *If $n > 1$ has no positive prime faster less than or equal to $\sqrt{n}$, then s is prime.*

Remark 3.8. This is obvious, if any integer doesnt have a product where one element is less than its root, then nothing can divide it. Prove this by contradiction and show that if $p_1 p_2$ divide n then $n = p_1 p_2 k \geq p_1 p_2 > \sqrt{n}\sqrt{n} = n \implies n > n$.

Proposition 3.6. $a|b \iff a^n|b^n$

Remark 3.9. For the reverse direction, use the prime factorizations of a and b. Then you will have the same arguement as 3.5 where the primes divide primes so you will be able to show the primes divide.

3.2 Congruence and Congruence Classes

Definition 3.5 (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 $a \equiv b \pmod{n}$

Remark 3.10. Notice this means the integers have the same remainder when divided by n. To see this consider the definition above along with their form given by the division algo 3.2

Theorem 3.6 (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 $a \equiv c \pmod{n}$.*

Proposition 3.7 (Modulo Arithmetic)**.** *If $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$, then*

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

Definition 3.6 (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

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

Remark 3.11. So a congruence class is just sets of integers that all leave the same remainder when divided by n. See 3.1 and 3.8.

Theorem 3.7 (Congruence Class Equality)**.** *$a \equiv c \pmod{n}$ if and only if $[a] = [c]$.*

Remark 3.12. For the direction to the right, show that $[a] \subseteq [c]$ and $[c] \subseteq [a]$ by letting $x \in [a]$ and show that x must be congruent to c. For the reverse direction, by Reflexivity $a \in [a]$...
This is completely obvious because we already know, integers are only equivalent modulo n if they have the same remainder when divided by n.

Corollary 3.1. *Two congruence classes modulo n are either disjoint or identical.*

Remark 3.13. Prove this by contradiction

Proposition 3.8. *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, $[0], [1], [2], \ldots, [n-1]$.*

Remark 3.14. For (1) use the form given by the division algo.

Definition 3.7. 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.

Proposition 3.9. *If a, b are integers such that $a \equiv b \pmod{p}$ for every positive prime p, then $a = b$.*

Theorem 3.8. *If $[a] = [b]$ and $[c] = [d]$ in $\mathbb{Z}_n$, then*

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

Definition 3.8 (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 3.10. *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. $[a] \odot [1] = [a] = [1] \odot [a]$.

Theorem 3.9. *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 $c = 0$.*

Remark 3.15. For (1) $\implies$ (2), we have that $a \neq 0$ in $\mathbb{Z}_p$ implies p doesnt divide a. Since p is prime, this means $\gcd(a, p) = 1$. Then using 3.3 we can show (2).
For (2) $\implies$ (3), we note that if $b \neq 0$ then we have from (2) that there is a solution to $bx = 1$. Then we can say that $c = 1 \cdot c = bcx = 0$.
For (3) $\implies$ (1) we note 3.4.

Corollary 3.2. *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 3.9 (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 3.10 (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 $ac = 0$.

3.3 Rings

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

Definition 3.11 (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 3.12 (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 3.13 (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 3.14 (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 3.15 (Field)**.** A field is a commutative ring R with identity $1_R \neq 0_R$ that satisfies the axiom

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

Proposition 3.11. *Let R and S be rings. Define addition and multiplication on the Cartesian product $R \times S$ by*

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

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 3.10 (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*.*

Remark 3.16. To check that S is a subring, we need to show that S satisfies all the axioms of a ring.

Theorem 3.11. *For any element* a *in a ring* R*, the equation* $a + x = 0_R$ *has a unique solution.*

Remark 3.17. Use axioms of rings. Specifically, since any element of the ring summed with the zero element is itself, we can substitue for the zero element using a different expression of a with 0. Then using associativity, we can finish the proof.

Theorem 3.12. *If* $a + b = a + c$ *in a ring* R*, then* $b = c$*.*

Remark 3.18. Add $-a$ (the additive inverse from the axiom of rings 3.11) from both sides and use associativity.

Definition 3.16 (Subtraction)**.** Let R be a ring and $a \in R$. By 3.11, the equation $a + x = 0_R$ has a unique solution, call it $-a$. Then,

$$a + (-a) = 0_R = (-a) + a$$

Proposition 3.12. *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. $(-1_R)a = -a$*.*

Definition 3.17. Let $n, m \in \mathbb{N}$, if R is a ring with $a \in R$, then

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

Proposition 3.13. ***[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*.*

Definition 3.18. [Units] 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 3.9.

Definition 3.19. [Zero-Divisor] 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 3.10.

Definition 3.20. **[Idempotent]** An element $e \in R$ is called idempotent if it satisfies

$$e^2 = e.$$

That is, if multiplied by itself gives back itself.

Theorem 3.13. *R is an integral domain if and only if $a \neq 0_R$ and $ab = ac$ in R, imply $b = c$.*

Remark 3.19. Since subtraction is defined, we can subtract ac to the other side. Then if we *undistribute* a then since R is an integral domain, we have that one of them is equal to 0 but by hypothesis its not a. So you get that $b - c = 0$.
For the reverse direction, since cancellation holds, consider $ab = a0_R$.

Theorem 3.14. *Every field F is an integral domain.*

Remark 3.20. Suppose you had $ab = 0$ in a field. Since the multiplicative inverse exists in a field for nonzero elements, we can multiply $ab = 0$ by the multiplicative inverse and get that the other element is 0.

Theorem 3.15. *Every finite integral domain R is a field.*

Remark 3.21. Since R is finite, suppose it has n elements. Since R is an integral domain, it contains an identity, and when two nonzero elements are multiplied the product is nonzero. So if I take some element $a_t \in R$ and multiply it with every other element

$$a_t a_1, a_t a_2, a_t a_3, \ldots, a_t a_n$$

Then this in n elements. Since R only has n elements, and since one of these must be an identity, we have that for each a_t there exists a_j such that $a_t a_j = 1$. This proves that the finite integral domain is a field.
This really follows from the fact that multiplying each element still results in a finite number of elements, so their must be a one to one corresponance between the elements which means one of them is equal to 1. But since I can do that for any element, I have shown that $a_i x = 1$ has a solution.

Definition 3.21. **[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. $f(a + b) = f(a) + f(b)$ and $f(ab) = f(a)f(b)$ for all $a, b \in R$.

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

Remark 3.22. An isomorphism is perfect symmetry amongst two structures. Meaning, it is not only a bijection, where each element from one ring corresponds to a unique element from the other ring, and the correspondance (f) is closed under addition and multiplication. As in, f respects the operations that form algebraic structure *and* has a 1-1 correspondance between the rings.

Definition 3.22. **[Homomorphism]** Let R and S be rings. A function $f : R \to S$ is said to be a *homomorphism* if

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

Remark 3.23. This only shows that the operations that birth the structures have correspondance with eachother, but there is not correspondance amongst each element in each of the rings. (get better interpretation)

Theorem 3.16. *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* $f(u)^{-1} = f(u^{-1})$.

Remark 3.24. We only need to say that f is surjective for (4) and (5) because we are trying to use something about R and say that it implies something holds *for the entire set* S.
For example, to prove (4), we need to be able to let s be any element of S. Then since f is surjective, $\exists r \in R$ such that $f(r) = s$. Thus,

$$sf(1_R) = f(r)f(1_R) = f(r1_R) = f(r) = s$$

So we showed that $f(1_R)$ is the identity for the entire ring S.

Corollary 3.3. *If* $f : R \to S$ *is a homomorphism of rings, then the image of* f *is a subring of* S.

Remark 3.25. This is obvious intuitevely because the definition of isomorphism implies the image of f is equivalent to the codomain. So for anything less than a bijection, which is what we get with a homomorphism, we would expect that the image of f, $im(f)$, is a subring.
Recall, by 3.13, we need only show that f is closed under subtraction and multiplication. From 3.16, this is obvious.

Definition 3.23. [Polynomial Coefficients] A polynomial with coefficients in the ring R is an infinite sequence,

$$(a_0, a_1, a_2, a_3, \dots)$$

such that $a_i \in R$ and only finitely many of the a_i are nonzero. That is, for some index k, $a_i = 0_R$, $\forall i > k$. The elements $a_i \in R$ are called the coefficients of the polynomial.

Remark 3.26. Notice that a sequence also respects ordering. So these coefficients contain all the information we need to discuss a ring of polynomials, since almost all operations are amongst the coefficients.
For example, we say two polynomials are equal if they are equal sequences. That is, if $a_i = b_i$, $\forall i = j$.

Definition 3.24. [Polynomial Addition and Multiplication] Let $(a_0, a_1, \dots)$ and $(b_0, b_1, \dots)$ be polynomials with coefficients in R. Then define addition by,

$$(a_0, a_1, \dots) + (b_0, b_1, \dots) = (a_0 + b_0, a_1 + b_1, \dots)$$

and define multiplication by

$$(a_0, a_1, \dots)(b_0, b_1, \dots) = (c_0, c_1, \dots)$$

where,

$$\begin{aligned}
c_0 &= a_0 b_0 \\
c_1 &= a_0 b_1 + a_1 b_0 \\
c_2 &= a_0 b_2 + a_1 b_1 + a_2 b_0 \\
\vdots\ &\ \vdots \\
c_n &= a_0 b_n + a_1 b_{n-1} + \cdots + a_{n-1} b_1 + a_n b_0 \\
&= \sum_{i=0}^{n} a_i b_{n-i}
\end{aligned}$$

Theorem 3.17. *Let* R *be a ring with identity and let* P *be the set of polynomials with coefficients in* R*. Then* P *is a ring with identity. Also, if* R *is commutative, then so is* P.

Remark 3.27. So really you only have to show that since R is a ring with identity, this implies the ring of polynomials is a ring with identity. Upon realizing that addition and multiplication for polynomials is completely defined in terms of the coefficients, of which are elements of a ring, we see that all the axioms follow quite obviously. The only axiom that really requires checking is associativity of multiplication. Use the summation definition and show that each sum can be shown to be equavalent to $\sum_{u,v,w} a_u b_v c_w$ where the sum is taken over all $u \geq 0, v \geq 0$, and $w \geq 0$ such that $n = u + v + w$.

Theorem 3.18. *Let P be the ring of polynomials with coefficients in the ring R. Let R^* be the set of all polynomials in P of the form $(r, 0_R, 0_R, 0_R, 0_R, \dots)$, with $r \in R$. Then R^* is a subring of P and is isomorphic to R.*

Remark 3.28. This is rather obvious. We are just showing that the constants can be represented. This shows that the ring that the coefficients lie in is a subring of the ring of polynomials.

Remark 3.29. We will now switch to the more familiar notation, with the variable x equipped with n in x^n to denote its power and thus its position in the sequence. For example, we have

$$a + bx = (a, 0_R, 0_R, \dots) + (b, 0_R, 0_R, \dots)(0_R, 1_R, 0_R, 0_R, \dots)$$

or

$$x = (0_R, 1_R, 0_R, 0_R, \dots)$$

or

$$cx^n = (0_R, 0_R, 0_R, c, 0_R, 0_R, \dots)$$

Lemma 3.1. *Let P be the ring of polynomials with coefficients in the ring R and x the polynomial $(0_R, 1_R, 0_R, 0_R, \dots)$. Then for each element $a = (a, 0_R, 0_R, \dots)$ of R^* and each integer $n \geq 1$,*

1. *$x^n = (0_R, 0_R, \dots, 0_R, 1_R, 0_R, \dots)$, where 1_R is in position n.*

2. *$ax^n = (0_R, 0_R, \dots, 0_R, a, 0_R, \dots)$, where a is in position n.*

Remark 3.30. Show this by induction on n, in x^n. When you get to x^{n+1}, use the definition of multiplication given in 3.24.
The only point of this lemma is to show that our conception of writing polynomials in coefficient and variable form is actually valid with a pretty fundamental definition of polynomials.

Theorem 3.19. *Let P be the ring of polynomials with coefficients in the ring R. Then P contains an isomorphic copy R^* of R and an element x such that*

1. *$ax = xa$ for every $a \in R^*$.*

2. *Every element of P can be written in the form*

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

3. *If $a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = b_0 + b_1 x + b_2 x^2 + \cdots + b_m x^m$ with $n \leq m$, then $a_i = b_i$ for $i \leq n$ and $b_i = 0_R$ for $i > n$, in particular;*

4. *$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = 0_R \iff a_i = 0_R, \ \forall i \geq 0$*

Remark 3.31. After proving the lemmas and theorems proceeding this, this is a rather simple conclusion.
We now change notation. Since we have shown that R^* is isomorphic to R and that our way of writing polynomials is isomorphic to our rigorous definition, we can denote the space of polynomials as $R[x]$ and write R instead of R^*, since they mean the same thing.

Definition 3.25 (Polynomial)**.** Let R be a ring. A polynomial with coefficients in R is an expression of the form

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

where n is a nonnegative integer and $a_i \in R$. Note that the elements x could be in some larger ring.

Definition 3.26 (Polynomial Addition and Multiplication)**.** Let $R[x]$ be the ring of polynomials with coefficients in a ring R. The operations of polynomial addition and multiplication are defined as follows:

Polynomial Addition: For polynomials

$$(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) + (b_0 + b_1x + b_2x^2 + \cdots + b_mx^m),$$

addition is performed by adding corresponding coefficients:

$$(a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots + (a_n + b_n)x^n.$$

Polynomial Multiplication: For polynomials

$$(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n)(b_0 + b_1x + b_2x^2 + \cdots + b_mx^m),$$

multiplication is performed using the distributive property and collecting like powers of x:

$$a_0b_0 + (a_0b_1 + a_1b_0)x + (a_0b_2 + a_1b_1 + a_2b_0)x^2 + \cdots + a_nb_mx^{n+m}.$$

Coefficient of x^k in the product: For each $k \geq 0$, the coefficient of x^k in the product of two polynomials is given by:

$$a_0b_k + a_1b_{k-1} + a_2b_{k-2} + \cdots + a_kb_0 = \sum_{i=0}^{k} a_ib_{k-i},$$

where $a_i = 0_R$ if $i > n$ and $b_j = 0_R$ if $j > m$.

Properties: - If R is commutative, then $R[x]$ is also commutative. - If R has a multiplicative identity 1_R, then 1_R is also the multiplicative identity of $R[x]$.

Definition 3.27 (Degree and Leading Coefficient of a Polynomial)**.** Let

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

be a polynomial in $R[x]$ with $a_n \neq 0_R$. Then a_n is called the **leading coefficient** of $f(x)$.

The **degree** of $f(x)$ is the integer n; it is denoted as $\deg f(x)$. In other words, $\deg f(x)$ is the largest exponent of x that appears with a nonzero coefficient, and this coefficient is the leading coefficient.

Theorem 3.20. *If R is an integral domain and $f(x), g(x)$ are nonzero polynomials in $R[x]$, then*

$$\deg[f(x)g(x)] = \deg f(x) + \deg g(x).$$

Corollary 3.4. *If R is an integral domain, then so is $R[x]$.*

Corollary 3.5. *Let R be a ring. If $f(x), g(x)$, and $f(x)g(x)$ are nonzero in $R[x]$, then*

$$\deg[f(x)g(x)] \leq \deg f(x) + \deg g(x).$$

Corollary 3.6. *Let R be an integral domain and $f(x) \in R[x]$. Then $f(x)$ is a unit in $R[x]$ if and only if $f(x)$ is a constant polynomial that is a unit in R.*

In particular, if F is a field, the units in $F[x]$ are the nonzero constants in F.

Theorem 3.21 (The Division Algorithm in $F(x)$)**.** *Let F be a field and let $f(x), g(x) \in F[x]$ with $g(x) \neq 0_F$. Then there exist unique polynomials $q(x)$ and $r(x)$ such that*

$$f(x) = g(x)q(x) + r(x)$$

and either $r(x) = 0_F$ or $\deg r(x) < \deg g(x)$.

Definition 3.28 (Divisibility in $F(x)$)**.** Let F be a field and $a(x), b(x) \in F[x]$ with $b(x)$ nonzero. We say that $b(x)$ **divides** $a(x)$ (or that $b(x)$ is a **factor** of $a(x)$) and write $b(x) \mid a(x)$ if

$$a(x) = b(x)h(x)$$

for some $h(x) \in F[x]$.

Definition 3.29 (Greatest Common Divisor in $F(x)$)**.** Let F be a field and $a(x), b(x) \in F[x]$, not both zero. The *greatest common divisor* (gcd) of $a(x)$ and $b(x)$ is the monic polynomial of highest degree that divides both $a(x)$ and $b(x)$.

In other words, $d(x)$ is the gcd of $a(x)$ and $b(x)$ provided that $d(x)$ is monic and:

1. $d(x) \mid a(x)$ and $d(x) \mid b(x)$;
2. If $c(x) \mid a(x)$ and $c(x) \mid b(x)$, then $\deg c(x) \leq \deg d(x)$.

Remark 3.32. Note that by the next theorem REF, the gcd is any multiple of $d(x)$.

Theorem 3.22. *Let F be a field and $a(x), b(x) \in F[x]$ with $b(x)$ nonzero.*

1. *If $b(x)$ divides $a(x)$ then $cb(x)$ divides $a(x)$ for each nonzero $c \in F$.*
2. *Every divisor of $a(x)$ has degree less than or equal to deg$(a(x))$.*

Theorem 3.23. *Let F be a field, and $a(x), b(x) \in F[x]$, not both zero. Then there exists polynomials $u(x)$ and $v(x)$ such that $d(x) = a(x)u(x) + b(x)v(x)$.*

Remark 3.33. This is the exact extension of bezouts equality in the integers. The proof is almost identical.

Theorem 3.24. *Let F be a field and $a(x), b(x), c(x) \in F[x]$. If $a(x) \mid b(x)c(x)$ and $1_F = \gcd(a(x), b(x))$, then $a(x) \mid c(x)$.*

Definition 3.30 (Associates in $F(x)$)**.** Let R be a commutative ring with identity. An element $a \in R$ is said to be an *associate* of an element $b \in R$ if there exists a unit $u \in R$ such that

$$a = bu.$$

In this case, b is also an associate of a since u^{-1} is a unit and $b = au^{-1}$. If F is a field, then by Corollary 3.6, the units in $F[x]$ are the nonzero constants in F. Therefore, in $F[x]$,

$$f(x) \text{ is an associate of } g(x) \text{ if and only if } f(x) = cg(x) \text{ for some nonzero } c \in F.$$

Definition 3.31 (Irreducible and Reducible Polynomials)**.** Let F be a field. A nonconstant polynomial $p(x) \in F[x]$ is said to be *irreducible* if its only divisors are its associates and the nonzero constant polynomials (units).

A nonconstant polynomial that is not irreducible is said to be *reducible*.

Remark 3.34. So an associate is really just a special case where the two polynomials are a multiple together but the multiple is a unit (so its invertible). Then we are basically characterizing prime polynomials since we are finding polynomials that cannot be divided by polynomials that are not equivalent to it or a constant.

Theorem 3.25. *Let F be a field. A nonzero polynomial $f(x)$ is reducible in $F[x]$ if and only if $f(x)$ can be written as the product of two polynomials of lower degree.*

Theorem 3.26. *Let F be a field and $p(x)$ a nonconstant polynomial in $F[x]$. Then the following conditions are equivalent:*

1. *$p(x)$ is irreducible.*
2. *If $b(x)$ and $c(x)$ are any polynomials such that $p(x) \mid b(x)c(x)$, then $p(x) \mid b(x)$ or $p(x) \mid c(x)$.*
3. *If $r(x)$ and $s(x)$ are any polynomials such that $p(x) = r(x)s(x)$, then $r(x)$ or $s(x)$ is a nonzero constant polynomial.*

Remark 3.35. This is the analogous theorem to 3.9 where we looked at the equivalence classes of modulo primes. So here we see that every condition has an equivalent representation for polynomials.

Corollary 3.7. *Let F be a field and $p(x)$ an irreducible polynomial in $F[x]$. If $p(x) \mid a_1(x)a_2(x) \cdots a_n(x)$, then $p(x)$ divides at least one of the $a_i(x)$.*

Theorem 3.27. *Let F be a field. Every nonconstant polynomial $f(x)$ in $F[x]$ is a product of irreducible polynomials in $F[x]$. This factorization is unique in the following sense:*

If

$$f(x) = p_1(x)p_2(x)\cdots p_r(x) \quad \text{and} \quad f(x) = q_1(x)q_2(x)\cdots q_s(x),$$

with each $p_i(x)$ and $q_j(x)$ irreducible, then $r = s$ (that is, the number of irreducible factors is the same). After the $q_j(x)$ are reordered and relabeled if necessary,

$$p_i(x) \text{ is an associate of } q_i(x) \quad (i = 1, 2, 3, \ldots, r).$$

4 Linear Algebra

Definition 4.1. 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 4.2 (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 4.3 (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

$$x_1v_1 + \cdots + x_nv_n$$

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

Definition 4.4 (Dot Product)**.** Let $V = K^n$. Let $A, B \in K^n$ with $A = (a_1, \dots, a_n)$ and $B = (b_1, \dots, b_n)$. We define the *dot product* or *scalar product* as

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

Remark 4.1. Geometrically we say that A and B are orthogonal MORE HERE

Definition 4.5 (Linear Independence)**.** Let $v_1, \dots, v_n$ be vectors in a vector space. The set of vectors $\{v_1, \dots, v_n\}$ 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 4.6 (Basis)**.** Let V be a vector space. A set of vectors $\{v_1, \dots, v_n\}$ in V is called a *basis* of V if:

1. The vectors $v_1, \dots, v_n$ *generate* V, meaning that every vector in V can be written as a linear combination of $v_1, \dots, v_n$.
2. The vectors $v_1, \dots, v_n$ 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 $\{v_1, \dots, v_n\}$ *forms a basis* of V.

Theorem 4.1. *Let V be a vector space. Let $v_1, \dots, v_n$ be linearly independent elements of V. Let $x_1, \dots, x_n$ and $y_1, \dots, y_n$ 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 $i = 1, \dots, n$.

Theorem 4.2. *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 4.7 (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 4.3. *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 4.4. *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.*

Corollary 4.1. *Let V be a vector space and let W be a subspace. If $\dim W = \dim V$, then $V = W$.*

Corollary 4.2. *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 4.5. *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 $\leq n$.*

Definition 4.8. 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 4.6. *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 4.7. *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 4.8. *If V is a finite-dimensional vector space over K, and is the direct sum of subspaces U, W, then*

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

5 Analysis

Theorem 5.1 (Archimedean Property)**.** *If $x, y \in \mathbb{R}$ and $x > 0$, then there exists an $n \in \mathbb{N}$ such that*

$$nx > y.$$

Theorem 5.2 (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*

$$x < r < y.$$

5.1 Sequences

Definition 5.1 (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 5.2 (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 5.3 (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 5.4 (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 $x_n \longrightarrow x$.

Proposition 5.1. *A convergent sequence has a unique limit.*

Proposition 5.2. *Let (s_n) be a sequence of non-negative real numbers and suppose $s = \lim_{n\to\infty}$. Then*

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

Proposition 5.3. *Convergent sequences are bounded.*

Proposition 5.4 (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. $\lim_{n\to\infty}(x_n y_n) = \left(\lim_{n\to\infty} x_n\right)\left(\lim_{n\to\infty} y_n\right)$.
3. *If* $\lim_{n\to\infty} y_n \neq 0$ *and* $y_n \neq 0$ *for all* $n \in \mathbb{N}$*, then* $\lim_{n\to\infty} \frac{x_n}{y_n} = \frac{\lim_{n\to\infty} x_n}{\lim_{n\to\infty} y_n}$.

Lemma 5.1 (Squeeze lemma)**.** *Let $\{a_n\}_{n=1}^{\infty}$, $\{b_n\}_{n=1}^{\infty}$, and $\{x_n\}_{n=1}^{\infty}$ be sequences such that*

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

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

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

Definition 5.5. 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 5.3 (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

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

Proposition 5.5. *Let $n \in \mathbb{N}$ then,*

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

Proposition 5.6. *If $0 < c < 1$, then*

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

Proposition 5.7 (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.*

Remark 5.1. So we want to use that $\frac{|x_{n+1}|}{|x_n|}$ having a limit less than 1 implies we can find r such that $0 \leq L < r < 1$ and $\frac{|x_{n+1}|}{|x_n|} < r^n$. Since we wont ever be less than L, we need $1 > r > L$. Then there exists an N such that $\forall n \geq N$ we have that $\frac{|x_{n+1}|}{|x_n|} < r$ then write out each term of $\frac{|x_{n+1}|}{|x_n|}$ multiplying each term before it but stopping at x_N and show that this expression is bounded.

Proposition 5.8. *If* $(x_n)_{n=1}^{\infty}$ *is convergent and* $k \in \mathbb{N}$ *then*

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

Proposition 5.9. *If* $(x_n)_{n=1}^{\infty}$ *is a convergent sequence and* $x_n \geq 0$ *and* $k \in \mathbb{N}$ *then*

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

Definition 5.6. 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 5.10. *If* $\{x_n\}_{n=1}^{\infty}$ *is a convergent sequence, then every subsequence* $\{x_{n_i}\}_{i=1}^{\infty}$ *is also convergent, and*

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

Definition 5.7. 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 (5.3) since with each passing n, the value that is the supremum of all x_n to come, can only decrease.

Theorem 5.4. *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 5.2. Construct x_{n_k} inductively up to n_{k-1} then choose x_{n_k} such that $a_{n_{k-1}+1} - x_{n_k} < \frac{1}{k}$. The rest of the proof follows from here.

Proposition 5.11. *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*

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

Remark 5.3. This follows from squeeze theorem.

Proposition 5.12. *Every sequence has a monotone subsequence.*

Proposition 5.13. *Suppose $(x_n)_{n=1}^{\infty}$ is a bounded sequence and $(x_{n_k})_{k=1}^{\infty}$ is a subsequence. Then*

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

Proposition 5.14. *A sequence $(x_n)_{n=1}^{\infty}$ converges to $x \iff$ every subsequence $(x_{n_k})_{k=1}^{\infty}$ converges to x.*

Definition 5.8 (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 5.5 (Bolzano–Weierstrass)**.** *Suppose a sequence $\{x_n\}_{n=1}^{\infty}$ of real numbers is bounded. Then there exists a convergent subsequence $\{x_{n_i}\}_{i=1}^{\infty}$.*

Remark 5.4. Since x_n is bounded, we know $\forall n\ x_n \in [a, b]$ for some a and b as bounds. Then splitting this interval into two halves, we know that infinitely many x_{n_k} lie in one (or both) halves, so pick that side (suppose it was the top half), then set $a_1 = a, b_1 = b$, then since we choose the top half, pick $a_2 = \frac{a_1+b_1}{2}$, and $b_2 = b_1$. If we contiue to do this, we will have monotone sequences b_n and a_n such that $b_n - a_n = \frac{b_1-a_1}{2^{n-1}}$. So we essnetially use nested interval property to show that a convergent subsequence can be made just by the fact that

Proposition 5.15. *Let (s_n) be any sequence of nonzero real numbers. Then we have*

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

Definition 5.9 (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 $n \ge M$ and all $k \ge M$, we have

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

Lemma 5.2. *If a sequence is Cauchy, then it is bounded.*

Remark 5.5. Since x_n is cauchy, we can fix x_N so that $|x_n - x_N| < \varepsilon = 1$. Then apply reverse triangle inequality to obtain $|x_n| < 1 + |x_N|$.
This kinda hints that cauchy might imply convergence because this means for any ε, we can fix an x_N so that $|x_n - x_N| < \varepsilon$, but this is basically x_n converging to a limit of x_N.

Theorem 5.6 (Convergent $\iff$ Cauchy)**.** *A sequence of real numbers is Cauchy $\iff$ the sequence is convergent.*

Remark 5.6. Since x_n is cauchy, by 5.2, we know x_n is bounded. Since x_n is bounded, by 5.4 we know there exists subsequences convergent to lim sup and lim inf. Now we want to use 5.11 to show that x_n converges. Then using that x_n being cauchy applies to the subsequences too, we can show $|\limsup - \liminf| < \varepsilon$.

5.2 Series

Definition 5.10 (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 5.16 (Geometric Series)**.** *Suppose* $-1 < r < 1$*. Then the geometric series* $\sum_{n=0}^{\infty} r^n$ *converges, and*

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

Remark 5.7. Consider $S_n = \sum_{i=0}^{n} r^i$ and $S_n - rS_n$.

Proposition 5.17. *Let* $\sum_{n=1}^{\infty} x_n$ *be a series and let* $M \in \mathbb{N}$*. Then*

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

Remark 5.8.

Definition 5.11 (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 5.6.

Proposition 5.18. *If a series* $\sum_{n=1}^{\infty} x_n$ *converges, then* $\lim x_n = 0$.

Remark 5.9. Since $\sum x_n$ converges, we know it is cauchy. Since it is cauchy, for any ε there is an N such that $\forall n \geq N$, we have that $|\sum x_n - \sum x_{n-1}| < \varepsilon$ but this just means $x_n < \varepsilon$.

Proposition 5.19 (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*

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

Proposition 5.20. *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.*

Remark 5.10. This is analogous to monotone convergence theorem 5.3 since being told $x_n \geq 0 \; \forall n \in \mathbb{N}$ is the same as saying $S_m = \sum^m x_n$ is a monotone increasing sequence. So given that S_m is bounded, we can say it converges.

Definition 5.12 (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 5.21. *If the series* $\sum_{n=1}^{\infty} x_n$ *converges absolutely, then it converges.*

Remark 5.11. Since the series of absolute values converges, that series must be cauchy 5.11. So consider the partial sums of the cauchy series's, since in cauchy form you are just subtracting one series from the other, the residual will just be the terms leftover from the sequence that *went further out*. Then using triangle inequality we can show that $\left|\sum_{i=k+1}^{m} x_n\right| < \left|\sum_{i=k+1}^{m} |x_n|\right| = \sum_{i=k+1}^{m} |x_n|$.

Proposition 5.22 (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* $\sum_{n=1}^{\infty} y_n$*.*

Remark 5.12. Since we require by hypothesis that the terms are nonnegative, we have that both of the series are monotone increasing, so if $\sum y_n$ converges, then it is bounded, then $\sum x_n$ is bounded and monotone increasing, thus convergent 5.3.

Proposition 5.23 (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 $p > 1$.

Remark 5.13. Since the terms of the series are positive, the sequence of partial sums is monotone increasing. Thus we only need to show it is bounded 5.3. Since the sequence $\frac{1}{n^p}$ is monotone decreasing, we have that the x_nth term is greater than the x_{n+1}th term. So we could bound the sum $\sum_{n=1}^{\infty} \frac{1}{n^p}$ by a different sum that is constructed from *prior* terms of $\sum_{n=1}^{\infty} \frac{1}{n^p}$. To do this, we replace each term between the 2^kth one and the $2^{k+1} - 1$th one with the 2^kth term. Why? Because the 2^kth term is greater than all the terms after it so we can create an upper bound. Why 2^k specifically? Because we can easily make a geometric series out of this. Why stop each grouped sum at $2^{k+1} - 1$? Beacause we want every term of the next 2^k group to be replaced with the 2^{k+1} term, so we need that the 2^{k+1}th term has not been summed yet. Then we have the following

$$s_{2^k-1} = 1 + \sum_{i=1}^{k-1} \sum_{m=2^i}^{2^{i+1}-1} \frac{1}{m^p} < 1 + \sum_{i=1}^{k-1} \sum_{m=2^i}^{2^{i+1}-1} \frac{1}{(2^i)^p} = 1 + \sum_{i=1}^{k-1} \frac{2^i}{(2^i)^p}$$

where the last sum is geometric and hence converges. Where did the $k-1$ come from? Since we are summing up to

Proposition 5.24 (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.*

Remark 5.14. Assume the limit of the ratio is less than 1 ($0 < L < 1$). Then there is an r such that $0 \leq L < r < 1$ or $0 < r - L < 1$. Then there exists some value N where for any epsilon we can bound $\frac{|x_{n+1}|}{|x_n|} - L$. Thus let $\varepsilon = r - L$ then show that this implies $\frac{x_{n+1}}{x_n} < r$. Thus, using the arguement in 5.7 we can bound each value past some N, thus we can bound the sum by a geometric series of r.

Proposition 5.25 (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.*

Remark 5.15. If we suppose that $L < 1$ then there exists an r such that $0 \leq L < r < 1$. Then there exists an N where $\forall n \geq N$, $\sup\{|x_k^{1/k} \mid k \geq n\} < r$. So we also have by default that $|x_n|^{1/n} < r \implies |x_n| < r^n$. Thus we can write the sum $\sum_{n=1}^{\infty} |x_n|$ as first the sum from $n = 1$ up to $n = N - 1$ plus the sum from $n = N$ to ∞, and since we have found that $|x_n|$ is bounded for every term after the Nth one, we have that one of the series is just a real number and the other is the tail of a geometric series. Thus $\sum_{n=1}^{\infty} |x_n|$ is monotone increasing (since every term is positive) and is bounded above (what we just showed).

Proposition 5.26 (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.

Proof. If you write out the terms of the sequence while thinking about the fact that x_n is monotone decreasing of positive real numbers, it becomes clear that the series can be grouped to be the sum of a bunch of negative numbers. That is,

$$S_{2k} = (-x_1 + x_2) + (-x_3 + x_4) + \cdots + (-x_{2k-1} + x_{2k}) = \sum_{l=1}^{k} (-x_{2l-1} + x_{2l})$$

But if I instead was summing by ending each group with an odd I would have $(x_{2l} - x_{2l+1}) \geq 0$, again because it is monotone decreasing. But this observation of an alternative grouping suggests,

$$\begin{aligned} &(x_2 - x_3) + (x_4 - x_5) + \cdots + (x_{2l} - x_{2l+1}) \geq 0 \\ \implies &-x_1 + (x_2 - x_3) + (x_4 - x_5) + \cdots + (x_{2l} - x_{2l+1}) \geq -x_1 \end{aligned}$$

and this tells us that S_{2k} is bounded below, and since it is monotone decreasing, by 5.3 we have that the subsequence S_{2k} is convergent. Now we want to choose M so that S_{2k} converges for all $k \geq M$. Then we want to choose some value N so that for $m \geq N$, regardless of whether m *lands on* an even indexed term or an odd indexed term, S_m converges. So suppose that m is of the form $m = 2k$, then since S_{2k} converges for some $k \geq M$, we have that S_m converges if $m \geq N$, or equivalently, $2k \geq N$, so we need $N \geq 2M$ since $2k \geq 2M \implies k \geq M$. Then if m is of the form $m = 2k+1$ then choosing $m \geq 2M + l$ gives us that $2k+1 \geq 2M+1 \implies k \geq M$ which only tells us that S_{2k} is convergent. But we have S_{2k+1}, but we can write that as $S_{2k+1} = S_{2k} - x_{2k+1}$. So if we also choose N so that $|x_{2k+1}| < \varepsilon$ then we will altogether have the existence of an N such that for all $m \geq N$ and for any $\varepsilon > 0$

$$|S_{2k} - L| < \varepsilon, \quad |x_{2k+1}| < \varepsilon, \quad \text{and} \quad S_{2k+1} = S_{2k} - x_{2k+1}$$

so using triangle inequality you can show that S_m converges using the above.

□

Definition 5.13. Consider a series $\sum_{n=1}^{\infty} x_n$. Given a bijective function $\sigma : \mathbb{N} \to \mathbb{N}$ the corresponding rearrangement is the series

$$\sum_{k=1}^{\infty} x_{\sigma(k)}$$

We simply sum the series in a different order.

Proposition 5.27. *Let $\sum_{n=1}^{\infty} x_n$ be an absolutely convergent series converging to x. Let $\sigma : \mathbb{N} \to \mathbb{N}$ be a bijection. Then $\sum_{n=1}^{\infty} x_{\sigma(n)}$ is absolutely convergent and converges to x.*

Theorem 5.7 (Mertens Theorem)**.** *Suppose $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty} b_n$ are two convergent series, converging to A and B respectively. Suppose at least one of the series converges absolutely. Then $\sum_{n=0}^{\infty} c_n$ converges to AB, where*

$$c_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0 = \sum_{i=0}^{n} a_i b_{n-i}$$

5.2.1 Power Series

Definition 5.14. Fix $x_0 \in \mathbb{R}$. A power series about x_0 is a series of the form

$$\sum_{n=0}^{\infty} a_n (x - x_0)^n$$

A power series is really a function of x. A power series is said to be convergent if there is at least one $x \neq x_0$ that makes the series converge. If $x = x_0$ every term except the first is 0 so it alwasy converges.

Proposition 5.28. *Let $\sum_{n=0}^{\infty} a_n (x - x_0)^n$ be a power series. If the series is convergent, then either it converges absolutely at all $x \in \mathbb{R}$, or there exists a number ρ, such that the series converges absolutely on the interval $(x_0 - \rho, x_0 + \rho)$ and diverges when $x < x_0 - \rho$ or $x > x_0 + \rho$.*

The number ρ is called the radius of convergence of the power series. We write $\rho = \infty$ if the series converges for all x, and we write $\rho = 0$ if the series is divergent.

Remark 5.16. So we want to find when $\sum_{n=0}^{\infty} a_n(x-x_0)^n$ converges. Since it is raised to the power of n, we use the root test 5.25

$$L = \limsup |a_n(x-x_0)^n|^{1/n} = |x-x_0| \limsup |a_n|^{1/n}$$

So when $L < 1$ we have that the power series will converge. Thus the convergence of the power series is characterized by the result of $\limsup |a_n|^{1/n}$. So let $R = \limsup |a_n|^{1/n}$. Then we have that if $R = \infty$ then $L = \infty$ for every x not equal to x_0 (since it trivially conveges there), so the series diverges. If $0 < R < \infty$ then if $L < 1$ then $R|x-x_0| < 1 \implies |x-x_0| < 1/R$. And if instead for $0 < R < \infty$ we had $L > 1$, then $R|x-x_0| > 1 \implies |x-x_0| > 1/R$.
So all together we have that if $R = \infty$ then the power series is divergent. If $R = 0$ then the power series converges everywhere. Otherwise, the power series converges in the interval of its radius of convergence $\rho = 1/R$.

5.6

5.2.2 Sequences and Series Key Examples

Example 5.1. Find the exact value of the sum of the series

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

5.3 Continuity

Functions are really just expressions (in a specific language) of relationships between things which can be specified in whatever which way the function requires. Continuity is one type of relationship that can be made, that is, a continuous function. Continuity is useful to describe motion, fluid relationships, or anything that just morally should not be expressed in discrete nature. We begin the discussion with an aside on how to get arbitrarily close to points in a set.

Definition 5.15 (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.

Remark 5.17. This is really just an extremely small open set around a point, but not including the point. Think about an open ball with an open center ball (visually).

Proposition 5.29. *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 $\lim_{n\to\infty} x_n = x$.*

Remark 5.18. If x is a cluster point, then we can construct a sequence such that for any n we pick $x_n \in (x - \frac{1}{n}, x + \frac{1}{n})$ so clearly x converges.

Conversely, if x_n converges then we can show that the set given in 5.15 can always be shown to be nonempty (a sequence can be made).

We would expect this based on the definition of a cluster point 5.15. Since the definition implies there are always elements in the domain that are close enough where they are within any distance of the point. So a sequence is really just an ordered list of numbers where after some threshold, the sequence will only contain the same values that satisfy the cluster point definition. This is the same reason why given the sequence converges to c we automatically know elements in the domain exist that satisfy the cluster point definition.

Definition 5.16. 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.

Remark 5.19. So as I near an element, that doesnt have to be in the domain of the function, that is, since elements of the funcitons domain can get within any distance of the limit point, it implies that if we want the function to be within any small distance of some value it converges towards (but doesnt have to equal) then we will always be able to find elements of our domain that are close enough to the limiting element.

Proposition 5.30. *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 5.3. *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.*

Proposition 5.31. *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 5.32. *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 5.33. *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*

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

Remark 5.20. Let x_n be a sequence convergent to c. Then we can use the sequential limit definition which makes $(f(x_n))_{n=1}^{\infty}$ and $(g(x_n))_{n=1}^{\infty}$ sequences so we can use 5.4.

Proposition 5.34. *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 5.17. 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 5.35. *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 5.36. *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 5.18. 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 5.37. *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 $f(c)$.*

Remark 5.21. (1) So if we have that c is not a cluster point of S. Then there is a δ distance of c such that the only value in the domain making the set nonempty is c itself. Thus we pick that δ for any ε then use that $x = c$.
(2) Just let $L = f(c)$ in the definition of the limit.
(3) ($\implies$) Since f is continuous, by definition 2.25, $\forall \varepsilon > 0$, $\exists \delta > 0$ such that if $x \in S$ and $|x - c| < \delta$, then $|f(x) - f(c)| < \varepsilon$. Let $x_n \in S$ be a sequence convergent to c. Now find an $M \in \mathbb{N}$ so that $\forall n \geq M$, $|x_n - c| < \delta \implies |f(x_n) - f(c)| < \varepsilon$.
($\impliedby$) For this direction, we prove the contrapositive, since otherwise you would have to show that a discrete convergence implies a continuous one, when the contrapositive is much more direct. It ends up being close to the statement above, only you are negating the definition of a limit.

Proposition 5.38. *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.

Remark 5.22. Let x_n be a sequence of real numbers that converges to c. Then we can use the limit algebra we found in 5.4 5.8 since we can now treat the function as a sequence.

Proposition 5.39. *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 5.40. *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.*

Remark 5.23. Use sequential definition again, we get that the input of $f(g(x_n))$ is just another sequence, which makes $f(g(x_n))$ just another sequence.

Proposition 5.41. *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 5.4. *A continuous function $f : [a, b] \to \mathbb{R}$ is bounded.*

Proof. We prove this by contrapositive. Suppose f is not bounded. Then for any sequence in the domain of f, we have $|f(x_n)| \geq M \quad \forall M \in \mathbb{N}$. That just comes from the definition of unbounded. Now since the sequence x_n is in a bounded set, by 5.5 there exists a convergent subsequence. Note the limit of the sequence is in the interval since the interval is closed. So since f is unbounded, we have that as the subsequence approaches c (its limit), f does not approach $f(c)$. If instead f was approaching $f(c)$, you would see a contradiction because f would then be bounded on $[a, b]$.

We would expect a function to be bounded if it is continuous on a closed and bounded interval. Since how else would continuity be satisfied, if f wasnt bounded on a closed interval then there would exist elements in the domain where no matter how close we get f will always have a discontuity there. □

Theorem 5.8 (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 $[a, b]$.*

Proof. Suppose f is continuous on $[a, b]$. Then since by 5.4 the set representing the image of the function is bounded. Thus there exists sequences $f(x_n)$ and $f(y_n)$ such that they converge to the supremum and the infimum of the image of f. Since x_n and y_n need not converge, we need to use that there are subsequences that converge (this holds because x_n and y_n are in a bounded set so by 5.5 there exists convergent subsequences) to find exactly what x and y actually are. That is, we need to find what values in the domain will map to our supremum and infimum. Since the limit of $f(x_n)$ is the same as the limit of $f(x_{n_i})$ (because we only found the convergent subsequence since x_n didnt have to converge, even though $f(x_n)$ *did* converge to the sup, we couldve picked a bad x_n that did everything around the value mapping to the sup other than converge for every every possible δ. So if x_n did in fact converge, then x_{n_i} will converge to the same value anyway). So $f(x_{n_i})$ converges to $\inf[a, b]$ and $f(y_{n_i})$ converges to $\sup[a, b]$ □

Lemma 5.5. *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 $f(c) = 0$.*

Proof. We will construct a sequence similar to how we did in 5.5. So let a_n and b_n be sequences such that $a_1 = a$ and $b_1 = b$. Then by induction, define a_n and b_n by,

$$\text{If } f\left(\frac{a_n + b_n}{2}\right) > 0 \text{ then let } a_{n+1} = a_n \text{ and } b_{n+1} = \frac{a_n + b_n}{2}$$
$$\text{If } f\left(\frac{a_n + b_n}{2}\right) \leq 0 \text{ then let } a_{n+1} = \frac{a_n + b_n}{2} \text{ and } b_{n+1} = b_n$$

Then we show inductively that $a_n < b_n$ for all n.
Since a_n and b_n are monotone and are each bounded above and below, these sequences converge, so let c and d be their respective limits. We want to show that $c = d$. So we want $d - c = 0$ or $\lim(b_n - a_n) = 0$.
Notice, $b_{n+1} - a_{n+1} = \frac{b_n - a_n}{2}$. So we have $d - c = \lim(b_n - a_n) = \frac{b-a}{2^{n-1}}$.
So since f is continuous at all points of its domain which is $[a, b]$, f is continuous at c and also from the above we have $f(c) \geq 0$ and $f(c) \leq 0$. So we conclude that $f(c) = 0$.
This proof sort of shows us how the element c is found, we find it by repeated narrowing down, using the average of the last values as the step taking us narrower. □

Theorem 5.9 (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 $f(c) = y$.*

Remark 5.24. Now apply 5.5 to $g(x) = f(x) - y$ so that $f(a) < 0$ and $f(b) > 0$. Then we will find $g(c)$ and thus $f(c)$ by using the lemma.

Corollary 5.1. *If $f : [a, b] \to \mathbb{R}$ is continuous, then the direct image $f([a, b])$ is a closed and bounded interval or a single number.*

Proof. From 5.4 we have that f is bounded. And from 5.8 we have that the inf and sup of the image of f is contained in the image of f (since the proof of EVT showed that there are subsequences that converge to the values that map to the inf and sup of the image, and since these subsequences were in a closed and bounded set $[a, b]$, we have that the set $[a, b]$ contains all the limit points.) □

Definition 5.19. Let $S \subset \mathbb{R}$, and let $f : S \to \mathbb{R}$ be a function. Suppose $\forall \epsilon > 0$, $\exists \delta > 0$ such that whenever $x, c \in S$ and $|x - c| < \delta$, then $|f(x) - f(c)| < \epsilon$. Then we say f is *uniformly continuous.*

Remark 5.25. So we are now choosing δ so that for any $c \in S$, the function is continuous. So δ only depends on ε and the domain the function is defined over (since we need to find a delta so that given any epsilon, we have that any x and c in the domain that are less than this delta will make the function continuous at the point).
Note that if a function is uniform continuous then it is obviously continuous since uniform continuity is stronger (or harder to acheive) then continuity.
Also, all of our results of continuity have been proved assuming a closed and bounded interval for the function, if this was not the case, then suppose the domain is open. Then we would have that as a sequence in the domain approaches an endpoint, the function approaches some number but since the domain is open, the function does not have to equal its value at the point that the sequence is converging. Why do we need the domain to be bounded? We need this becuase if the domain was unbounded then we would never be able to say that the function contains a max and a min in its image (consider $f(x) = x$).
So recall f is continuous if

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

In the above, we see that our choice of δ depends on our choice of ε and x. If we want our choice of δ to depend only on ε, then we have

$$\text{Uniform Continuity: } \forall \varepsilon > 0, \exists \delta > 0, \text{ such that } x, y \in S \text{ and } |x - y| < \delta \implies |f(x) - f(y)| < \varepsilon$$

Theorem 5.10. *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then f is uniformly continuous.*

Proof. Note that the theorem actually makes sense, since, as discussed in the remark of 5.19, we have a continuous function that is defined on a closed and bounded interval, this being the domain is key.
We prove this by contrapositive, so we suppose that f is not uniformally continuous.
Now we have that no matter how close the inputs of f are, the difference of f evaluated at those inputs is always unbounded $\forall \varepsilon > 0$. So for sequences x_n and y_n, we have

$$|x_n - y_n| < 1/n \implies |f(x_n) - f(y_n)| \geq \varepsilon$$

Then to show that f is continuous, we need to show that as a sequence approaches some value c, the sequence of the function does not converge to $f(c)$. So we need convergent sequences. Since $[a, b]$ is closed and bounded, there exists convergent subsequences x_{n_k} and y_{n_k}. So suppose $x_{n_k} \to c$. Then to show that y_{n_k} also converges to the same value (which is essentially by construction), we have

$$|y_{n_k} - c| \leq |y_{n_k} - x_{n_k}| + |x_{n_k} - c| < 1/n_k + |x_{n_k} - c|$$

Note that we have $|y_{n_k} - x_{n_k}| < 1/n_k$ because if the sequences x_n and y_n are always within $1/n$ of eachother, then we must have that $|y_{n_k} - x_{n_k}|$ are within $1/n_k$ of eachother.
So now we have two subsequences that converge to the same value and are within any small distance of eachother. So we want to show that as these subsequences converge to c, the function evaluated at the subsequences does not converge to $f(c)$. So consider,

$$|f(x_{n_k}) - f(c)| \leq |f(x_{n_k}) - f(y_{n_k})| + |f(y_{n_k}) - f(c)|$$

So the first term on the RHS is unbounded for any epsilon greater than 0, but we dont get much about $|f(x_{n_k}) - f(c)|$ or $|f(y_{n_k}) - f(c)|$ being unbounded. So instead consider the reverse triangle inequality, then we have

$$\varepsilon \leq |f(x_{n_k}) - f(c)| + |f(y_{n_k}) - f(c)|$$

So one or both of these terms must diverge, thus we have that one of x_{n_k} or y_{n_k} converges to c while the function does not converge to $f(c)$. This means f is not continuous at c.
Why did we need two sequences x_n and y_n? This came from negating uniform continuity, since we wanted to show that these sequences could get as close as we would like. Then having two sequences also helped with triangle inequality to show that f is not continuous. $\square$

Definition 5.20 (Lipschitz Continuity)**.** A function $f : S \to \mathbb{R}$ is *Lipschitz continuous**, if there exists a $K \in \mathbb{R}$ such that

$$|f(x) - f(y)| \leq K|x - y| \quad \forall x, y \in S.$$

Remark 5.26. To interpret this definition, notice it is saying that if there exists a real number K so that for any x and y in the domain of f, the slope of the secant line between $(x, f(x))$ and $(y, f(y))$ is less than K. That is, $\exists K \in \mathbb{R}$ such that

$$\left|\frac{f(x) - f(y)}{x - y}\right| \leq K \quad \forall x, y \in S$$

That is, this K holds for every x, y. So lipschitz continuity is broken if there exists two points where the slope between them is unbounded. This makes sense.

Proposition 5.42. *A Lipschitz continuous function is uniformly continuous.*

Proof. From the definition you have

$$|f(x) - f(y)| \leq K|x - y| \quad \forall x, y \in S$$

so for any $\varepsilon > 0$, we want to choose $\delta = \varepsilon/K$ □

Lemma 5.6. *Let $S \subset \mathbb{R}$ and let $f : S \to \mathbb{R}$ be a uniformly continuous function. Let $\{x_n\}_{n=1}^{\infty}$ be a Cauchy sequence in S. Then $\{f(x_n)\}_{n=1}^{\infty}$ is Cauchy.*

Remark 5.27. This almost doesnt even need a proof since the sequential definition of uniform continuity already looks so similar to the definition of cauchy sequences.

Proposition 5.43. *A function $f : (a, b) \to \mathbb{R}$ is uniformly continuous if and only if the function $\tilde{f} : [a, b] \to \mathbb{R}$ is continuous, that is, the limits*

$$L_a := \lim_{x \to a} f(x) \qquad \textit{and} \qquad L_b := \lim_{x \to b} f(x)$$

exist and the function $\tilde{f} : [a, b] \to \mathbb{R}$ defined by

$$\tilde{f}(x) := \begin{cases} f(x) & \textit{if } x \in (a, b), \\ L_a & \textit{if } x = a, \\ L_b & \textit{if } x = b \end{cases}$$

is continuous.

Proof. This theorem is saying that if f is unif continuous on an open set, then we can extend the function to be continuous on a closed and bounded set, where the closed and bounded set is made by extending the open set to include its limit points. It is also saying that a continuous function on a closed and bounded set is uniformally continuous when it is restricted to an open set that is a subset of $[a, b]$.
We start by assuming $\tilde{f}(x)$ is continuous on $[a, b]$. Then by 5.10 we have that $\tilde{f}$ is uniformally continuous on $[a, b]$, thus it is uniformally continuous on any subset of $[a, b]$, namely (a, b), so we have that f is uniformally continuous.
Conversely, if we assume f is uniformally continuous on (a, b), then to show that it is continuous on $[a, b]$, we need to show that the limits $L_a = \lim_{x \to a} f(x)$ and $L_b = \lim_{x \to b} f(x)$ exist. So take two sequences x_n and y_n in (a, b) that are both convergent to a. Since they converge they are cauchy, thus we have that f is cauchy and thus convergent. So let $f(x_n) \to L_1$ and $f(y_n) \to L_2$. We want to show that $L_1 = L_2$ which shows that L_a exists since we have that for all sequences convergent to a, all the subsequential limits are equal so the sequence has the same limit and converges 5.10 ("all sequences

convergent to a" because x_n and y_n are general).
I am about to omit lots of details. So for any epsilon, there exists a delta such that $|f(x)-f(y)| < \varepsilon/3$, and for all $n \geq M$ we have $|x_n - y_n|$, $|f(x_n) - L_1| < \varepsilon/3$, and $|f(y_n) - L_2| < \varepsilon/3$, then we have

$$|L_1 - L_2| \leq |L_1 - f(x_n)| + |f(x_n) - f(y_n)| + |f(y_n) - L_2| < \varepsilon$$

So we have that the limit of the function defined on (a, b) as a (and b) are approached exist and are equivalent for any sequence in (a, b) that converges to a (or b). We have that $\tilde{f}$ is continuous because it is continuous at a and at b and it is continuous for any value $c \in (a, b)$. ☐

Definition 5.21. Let $S \subset \mathbb{R}$. We say $f : S \to \mathbb{R}$ is *increasing* if $x, y \in S$ with $x < y$ implies $f(x) < f(y)$. We define *decreasing* in the same way, switching inequalities.

Proposition 5.44. *Let $S \subset \mathbb{R}$, $c \in \mathbb{R}$, $f : S \to \mathbb{R}$ be increasing, and $g : S \to \mathbb{R}$ be decreasing. If c is a cluster point of $S \cap (-\infty, c)$, then*

$$\lim_{x \to c^-} f(x) = \sup\{f(x) : x < c,\ x \in S\} \quad \textit{and} \quad \lim_{x \to c^-} g(x) = \inf\{g(x) : x < c,\ x \in S\}.$$

If c is a cluster point of $S \cap (c, \infty)$, then

$$\lim_{x \to c^+} f(x) = \inf\{f(x) : x > c,\ x \in S\} \quad \textit{and} \quad \lim_{x \to c^+} g(x) = \sup\{g(x) : x > c,\ x \in S\}.$$

If ∞ is a cluster point of S, then

$$\lim_{x \to \infty} f(x) = \sup\{f(x) : x \in S\} \quad \textit{and} \quad \lim_{x \to \infty} g(x) = \inf\{g(x) : x \in S\}.$$

If $-\infty$ is a cluster point of S, then

$$\lim_{x \to -\infty} f(x) = \inf\{f(x) : x \in S\} \quad \textit{and} \quad \lim_{x \to -\infty} g(x) = \sup\{g(x) : x \in S\}.$$

Namely, all the one-sided limits exist whenever they make sense. For monotone functions therefore, when we say the left-hand limit $x \to c^-$ exists, we mean that c is a cluster point of $S \cap (-\infty, c)$, and same for the right-hand limit.

Corollary 5.2. *If $I \subset \mathbb{R}$ is an interval and $f : I \to \mathbb{R}$ is monotone and not constant, then $f(I)$ is an interval if and only if f is continuous.*

Proof. Firstly note what this is really saying; is the function is monotone and defined on an interval of real numbers, then the range of f being an interval is equivalent to it being continuous. This makes sense intuitvely since if we were to suppose that $f(I)$ is an interval subset of real numbers. Then we can get ε close to any $f(c)$ in the range of f, and since the domain is ... ☐

Corollary 5.3. *Let $I \subset \mathbb{R}$ be an interval and $f : I \to \mathbb{R}$ be monotone. Then f has at most countably many discontinuities.*

Proposition 5.45. *Let f be an injective, continuous function on an interval I. Then f is strictly monotonic.*

Proposition 5.46. *If f is a strictly monotonic function, then it is bijective and so has an inverse defined on its range.*

Proof. Suppose $x_1 \neq x_2$ then either $f(x_1) < f(x_2)$ or $f(x_2) < f(x_1)$, in either case, the function values are not equal so it is injective. By default, the function is surjective since the inverse is constrained to have its domain as the range of f, and by definition of the range, every value gets mapped to. ☐

Proposition 5.47. *If $I \subset \mathbb{R}$ is an interval and $f : I \to \mathbb{R}$ is strictly monotone, then the inverse $f^{-1} : f(I) \to I$ is continuous.*

6 Combinatorics

Theorem 6.1. *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 6.2. *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 6.3. *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 6.4. *The number of ways a set of n distinct objects can be partitioned into k subsets with n_k objects in the kth subset is*

$$\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1! n_2! \cdots n_k!}$$

Theorem 6.5. *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

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

Theorem 6.6. *Let n, k, and j be nonnegative integers with $0j \leq k \leq n$. Then for a set with n distinct elements, all of the following hold*

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

2. $$\binom{n}{k} = \binom{n}{n-k}$$

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

4. $$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$$

5.

$$2^n = \sum_{k=0}^{n} \binom{n}{k}$$

6.

$$\binom{k}{k} + \binom{k+1}{k} + \binom{k+2}{k} + \cdots + \binom{n}{k} = \binom{n+1}{k+1}$$

7 Probability

To define probability rigorously, we need to be have a system of sets defined so that the elements of these sets can be considered as probabilities over sets of outcomes. In the below, the σ-algebra is the system of sets that can generate these probability spaces.

Definition 7.1 (Event Space (Algebra and σ-algebra))**.** Let Ω be an abstract space (sample 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 $\bigcap_{i=1}^\infty A_i \in \mathcal{A}$.

The below result shows us that we can generate the sigma algebras by considering complements, unions, and intersections of open, closed, and clopen intervals. This means that we can generate any sigma-algebra by intervals of the form $(-\infty, a]$ where $a \in \mathbb{Q}$, which means (since the rationals are countable) that all sigma algebras can be made by a countably infinite number of unions, intersections, and/or complements.

Theorem 7.1 (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 $a \in \mathbb{Q}$.*

Definition 7.2 (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. $P(A) \geq 0$
2. $P(\Omega) = 1$
3. 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
$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Theorem 7.2. *Let $A_1, A_2, \ldots, A_n$ be events, then*

$$\begin{aligned} P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{1\leq i_1 \leq i_2 \leq n} P(A_{i_1} \cap A_{i_2}) \\ + \sum_{1\leq i_1\leq i_2\leq i_3\leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \sum_{1\leq i_1 < i_2 \leq i_3 < i_4 \leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3} \cap A_{i_4}) \\ + \cdots + (-1)^{n+1} P(A_1 \cap \cdots \cap A_n) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1\leq i_1 < \cdots < i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k}) \end{aligned}$$

Definition 7.3. (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 $\lim_{n\to\infty} 1_{A_n}(\omega) = 1_A(\omega)$ $\forall \omega \in \Omega$.

Definition 7.4 (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 7.1. *Let $\mathcal{A}$ be a σ-algebra and $(A_n)_{n\geq 1}^{\infty}$ be a sequence of sets in $\mathcal{A}$. Then,*

$$\liminf_{n\to\infty} A_n \in \mathcal{A}, \quad \limsup_{n\to\infty} A_n \in \mathcal{A}, \quad \text{and } \liminf_{n\to\infty} A_n \subseteq \limsup_{n\to\infty} A_n$$

Lemma 7.2. *Let $\mathcal{A}$ be a σ-algebra and $(A_n)_{n\geq 1}^{\infty}$ be a sequence of sets in $\mathcal{A}$. Then,*

$$\lim_{n\to\infty} A_n = A \iff \limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n = A$$

Theorem 7.3 (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 $\lim_{n\to\infty} P(A_n) = P(A)$.*

Definition 7.5 (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 7.4. *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 (7.2)*
2. *$A_n \downarrow A \implies P(A_n) \downarrow P(A)$.*
3. *$A_n \uparrow A \implies P(A_n) \uparrow P(A)$*

Proposition 7.1. *Let $A_i \in \mathcal{A}$ be a sequence of events. Then,*

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

7.1 Conditional Probability and Independence

Definition 7.6. 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

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

Proposition 7.2 (Conditional Probability Measure). *The conditional probability is a probability measure (7.2).*

Definition 7.7. 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.

Proposition 7.3. *If A and B are independent, so also are A and B^c, A^c and B, and A^c and B^C.*

Proposition 7.4 (Multiplication Rule). *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 7.8 (Partition). 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 7.5 (Total Probability). *Suppose that $A_1, \ldots, A_n$ is a partition of Ω with $P(A_i) > 0$ for $i = 1, \ldots, n$. Then for any event E we have*

$$P(E) = \sum_{i=1}^{n} P(E \cap A_i) = \sum_{i=1}^{n} P(E \mid A_i)P(A_i)$$

Theorem 7.5 (Bayes Theorem). *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 7.9. 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 $1 \leq i_1 < i_2 < \cdots < i_k \leq n$,

$$P(A_{i_1} A_{i_2} \ldots A_{i_k} \mid B) = P(A_{i_1} \mid B)P(A_{i_2} \mid B) \cdots P(A_{i_k} \mid B).$$

7.2 Random Variables

Definition 7.10 (Random Variable)**.** Let $(\Omega, \mathcal{A}, P)$ be a probablity space. A random variable is a measurable function $X : \Omega \to \mathbb{R}$ such that for every Borel set $B \in \mathcal{B}(\mathbb{R})$, the preimage

$$X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\} \in \mathcal{A}$$

Equivalently, X is a random variable if $\forall x \in \mathbb{R}$, the set

$$\{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{A}$$

Definition 7.11 (Cumulative Distribution Function)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be a real-valued random variable. The distribution function, or the cumulative distribution function (CDF) of X, denoted $F_X : \mathbb{R} \to [0, 1]$, is defined by

$$F_X(x) = P(X \leq x), \quad \forall x \in \mathbb{R}.$$

The function $F_X(x)$ satisfies the following properties

1. (Monotonicity) $F_X(x)$ is monotone increasing
2. (Right Continuity) $F_X(x)$ is right continuous
$$\lim_{h \to 0^+} F_X(x + h) = F_X(x)$$
3. (Limits at Infinity)
$$\lim_{x \to -\infty} F_X(x) = 0, \quad \lim_{x \to \infty} F_X(x) = 1.$$
4. (Jumps) If $F_X(x)$ has a jump at x, then
$$P(X = x) = F_X(x) - \lim_{y \to x^-} F_X(y),$$
then X has positive probability at x
5. (Absolute Continuity) If $F_X(x)$ is absolutely continuous, then there exists a probability density function (PDF) $f_X(x)$ such that
$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt.$$

Definition 7.12 (Expected Value)**.** Let X be a real-valued random variable on a countable space Ω. The expectation X, denoted $E(X)$, is defined to be

$$E(X) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}), \text{ if continuous then } E(X) = \int_{\Omega} X(\omega) dP(\omega)$$

If X is a discrete random variable where $p_k = P(X = x_k)$ then

$$E(X) = \sum_{k} x_k p_k$$

If X is a continuous random variable, then

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

Definition 7.13 (nth Moment)**.** The nth moment of the random variable X is the expectation $E(X^n)$.

$$E(X^n) = \sum_{\omega} X^n(\omega)P(X = \omega) \quad \text{or} \quad \int_{-\infty}^{\infty} x^n dF(x) = \int_{-\infty}^{\infty} x^n f_X(x)dx$$

Definition 7.14 (Variance)**.** 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. Specifically

$$\sum_k (x_k - \mu_X)^2 p_X(x_k) \text{ (discrete)} \qquad \int_{-\infty}^{\infty} x^n f_X(x)dx \text{ (continuous)}$$

Definition 7.15. Let g be a real valued function defined on the range of a random variable X. If X is a discrete random variable then

$$E[g(X)] = \sum_k g(k)P(X = k)$$

while if X is continuous random variable with density function f, then

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

Definition 7.16. The *median* of a random variable X is any real value m that satisfies

$$P(X \geq m) \geq \frac{1}{2} \quad \text{and} \quad P(X \leq m) \geq \frac{1}{2}.$$

Definition 7.17. For $0 < p < 1$, the pth quantile of a random variable X is any real value x satisfying

$$P(X \geq x) \geq 1 - p \quad \text{and} \quad P(X \leq x) \geq p$$

Theorem 7.6. *Let $h : \mathbb{R} \to [0, \infty)$ be a nonnegative function and let X be a real valued random variable. Then*

$$P(\{\omega \mid h(X(\omega)) \geq a\}) \leq \frac{E(h(X))}{a}, \quad \forall a > 0.$$

Proof. Let $Y = h(X(\omega))$ be a random variable. Then we want to show

$$aP(Y \geq a) \leq E(Y)$$

So we investigate $E(Y)$,

$$E(Y) = \int_\Omega Y_\omega dP_\omega$$

Since this function is monotone increasing, any smaller subset will be less than it, thus

$$E(Y) = \int_\Omega Y_\omega dP_\omega \geq \int_{Y \geq a} Y_\omega dP_\omega \leq a \int_{Y \geq a} 1 dP_\omega = aP(Y \geq a)$$

Where the last inequality comes from the definition of the set $(\{Y = h(X(\omega)) \geq a\})$. Then we have that we have summed over all of the probabilites in that set and a is a constant to it just multiplies. This is our result. $\square$

Corollary 7.1 (Markovs Inequality)**.**

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

Proposition 7.6 (Affine Equivariance)**.** *Let X be a random variable and a and b be real numbers. Then,*

$$E(aX + b) = aE(X) + b$$
$$Var(aX + b) = a^2 Var(X)$$

Corollary 7.2 (**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. $P(\{|X - E(X)| \geq a\}) \leq \frac{\sigma_X^2}{a^2}$

7.3 Distributions

We will now look at the different distributions associtated with a random variable and we will discuss the motivations for using each one.
The most obvious distribution of all is the one where each point has equivalent probability.

Definition 7.18 (Uniform Distribution)**.** Let $[a, b] \subset \mathbb{R}$. A random variable X has a *uniform distribution on the interval* $[a, b]$ if X has the density function

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & x \notin [a, b] \end{cases}$$

The variance and the mean are given by

$$E(X) = \frac{b + a}{2} \quad \text{and} \quad Var(X) = \frac{(a - b)^2}{12}$$

Proof. Use the continuous expectation that utilizes the pdf of the random variable. This is given by,

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

So we have

$$\int_a^b \frac{x}{b - a} dx$$

The rest follows from there. Note that the result being the average of the interval is intuitive since the probability density is uniform, it is essentially leaving the real number line (our measuring system/device) with the same uniform (all the same) probabilistic density.
To find variance, we use $Var(X) = E(X^2) - (E(X))^2$. □

Definition 7.19 (Binomial Distribution)**.** Let n be a positive integer and $0 \leq p \leq 1$. A random variable X has the *binomial distribution* 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 $X\ Bin(n, p)$.
The variance and the mean are given by

$$E(X) = np \quad Var(X) = np(1 - p)$$

Proof. To find the mean, we can use linearity of expectation to acheive the quick solution of

$$E(X = x_i) = E\left(\sum_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} E(x_i) = \sum_{i=1}^{n} p = np$$

Or we can go through the calculation,

$$\begin{aligned}\sum_{k=0}^{n} kP(X = k) &= \sum_{k=1}^{n} k\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} \\ &= np\sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}(1-p)^{n-k} \\ &= np\sum_{j=0}^{n-1} \frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}(1-p)^{n-k}\end{aligned}$$

□

Remark 7.1. A binomial random variable is associated with experiments in which n independent trials where each trial is associated with a success with probability p and a failure with probability $1 - p$ and X represents the number of success that occur in the n trials. Recall the binomial coefficient is counting the number of subsets of size k that is contained in a set of size n.

Remark 7.2. Now consider you are trying to figure out how many flips of a coin till you get a head. Let the probability of a head be p. Then the probability of obtaining heads on the 10th flip is $(1-p)^9p$. This motivates the below definition

Definition 7.20 (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,

$$P(X = k) = (1-p)^{k-1}p, \quad k = 1, 2, 3, \dots$$

The mean and variance are given by

$$E(X) = \frac{1}{p} \quad \text{and} \quad Var(X) = \frac{1-p}{p^2}$$

Remark 7.3. Notice for a geometric random variable, we have the following property

$$P(X > i + j \mid X > i) = P(X > j), \qquad \forall i, j \geq 1$$

this is called the memoryless property. It says that if we flipped a coin i times and no head has turned up yet, then the probability for no head to turn up in the next j flips is exactly the same as the probability of no head for the first j flips of the coin.

Definition 7.21 (Negative Binomial Distribution)**.** A random variable X is called a *negative binomial* random variable with parameters p and k if its pmf is given by

$$P_X(x) = P(X = x) = \binom{x-1}{k-1}p^k(1-p)^{x-k} \qquad x = k, k+1, \dots$$

where $0 \leq p \leq 1$.
The mean and variance of the negative binomial random variable are given by

$$E(X) = \frac{k}{p} \quad \text{and} \quad Var(X) = \frac{k(1-p)}{p^2}$$

Remark 7.4. A negative binomial random variable X is associated with a sequence of independent Bernoulli trials (this is binomial rv with $n = 1$) with probability of success p, and X represents the number of trials until the kth success is obtained. For example, if we were flipping a coin and if $X = x$, then it must be true that there were exactly $k-1$ heads thrown in the first $x-1$ flips, and a head must have been thrown on the xth flip. There are $\binom{x-1}{k-1}$ sequences of length x with these properties, and each one has the same probability $p^{k-1}(1-p)^{x-k}$. Note that when $k = 1$, X is just a geometric random variable. success is obtained

Remark 7.5. Now imagine you have a finite population with N objects. Then suppose K objects are of type 1 and $N-K$ objects are of type 2. We draw a sample of n without replacement from the N total and want to know the probability of getting exactly k objects of type 1. This is the exact same as the binomial distribution only now we are not replacing. So for some $P(X = k)$ we have $\binom{K}{k}$ ways of choosing the k type 1 objects from the total K amount of them. Then we have $\binom{N-K}{n-k}$ ways to select the remaining $n-k$ objects of type 2 from the total $N-K$ of type 2. Then $\binom{N}{n}$ is the total number of ways to select the n objects.

Definition 7.22 (**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 7.23 (**Poisson Distribution**)**.** A random variable X follows a Poisson distribution with parameter $\lambda > 0$ if its pmf is given by

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots.$$

The mean and the variance is given by

$$E(X) = \lambda \quad \text{and} \quad Var(X) = \lambda$$

Remark 7.6. The Poisson has a range of applications, mostly due to its ability to approximate the binomial distrbution whenever np is "large enough", that is, with small p and large n. Some examples of a poisson r.v. include the number of telephone calls arriving during various intervals of time, the number of misprints on a page of a book, and the number of customers entering somewhere in a given span of time. So really it models the number of events occuring in a fixed interval of time or space, under the assumption that the events occur independently and at a constant average rate λ.
So notice as the rate λ of the event occuring increases, the probability of observing this event at each kth interval increases exponentially with k. Additionally, the probability depends inversely on the factorial of k and also depends inversely on e^{λ}.

Definition 7.24 (Exponential Distribution)**.** A random variable X is called an exponential r.v. with parameter $\lambda > 0$ if its pdf is given by

$$f_X(x) = \begin{cases} \frac{1}{\theta}e^{-\frac{x}{\theta}} & x, \theta > 0 \\ 0 & x < 0 \end{cases}$$

The mean and variance of the exponential distribution are given by,

$$E(X) = \frac{1}{\lambda} \qquad Var(X) = \frac{1}{\lambda^2}$$

Remark 7.7. The exponential distribution has the same memoryless property that the geometric random variable has, namely

$$P(X > s + t \mid X > s) = P(X > t) \qquad s, t \geq 0$$

This is the continuous form of the poisson distribution where for the poisson distribution, λ is the mean number of events in an interval, and θ is the mean waiting time until the first event occurs. Thus, $\theta = \frac{1}{\lambda}$

Definition 7.25. A random variable Z has the standard normal distribution if Z has density function

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$$

on the real line.
The mean and variance are given by

$$E(Z) = 0 \qquad Var(Z) = 1.$$

Definition 7.26 (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

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

where the mean and variance are given by

$$E(X) = \mu \qquad Var(X) = \sigma^2$$

Remark 7.8. Note that if X is defined as above, then $Z = \frac{X-\mu}{\sigma}$ is a standard normal random variable.

Definition 7.27 (Gamma Distribution)**.** A random variable X is called a gamma r.v. with parameters $\alpha > 0$, and $\lambda > 0$ if its pdf is given by

$$f_X(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\gamma(\alpha)} & x > 0 \\ 0 & x < 0 \end{cases}$$

where $\gamma(\alpha)$ is the gamma function defined by

$$\gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} dx \qquad \alpha > 0$$

The mean and variance of the gamma r.v. are given by

$$E(X) = \frac{\alpha}{\lambda} \qquad Var(X) = \frac{\alpha}{\lambda^2}$$

The mean and the variance of the gamma r.v. is given by

$$E(X) = \alpha\theta \qquad Var(X) = \alpha\theta^2$$

Remark 7.9. Where we use the poisson to calculate the average number of events in a given span of time, we use exponential to calculate the average wait time till an event, now we use the gamma distribution to find the waiting time until the αth event.

Definition 7.28. A random variable X is called a chi-squared r.v. with n degrees of freedom if the pdf is given by

$$f_X(x) = \frac{(1/2)e^{-x/2}(x/2)^n}{\gamma(n/2)}$$

where the mean and the variance are given by

$$E(X) = n \qquad Var(X) = \alpha\theta$$

Remark 7.10. This is obtained from the gamma distribution by letting $\alpha = n/2$, $\lambda = 1/2$.