

Math 375: Discrete Mathematics

Contents

1	Module 1: Basics of Rigorous Mathematics	5
1.1	Lecture 1. Sets and Functions	5
1.2	Lecture 2. Basic Logic.	12
1.3	Lecture 3. Induction.	18
2	Counting	25
2.1	Lecture 4. Binomial Theorem and some basics in counting.	25
2.2	Lecture 5. More on sets and functions.	31
2.3	Lecture 6. More on counting	38
2.4	Lecture 7. Recursion and Generating Functions	45
3	An Excursion into Probability Theory	49
3.1	Lecture 8. Basic probability theory	49
3.2	Lecture 9. More probability theory	56
4	An Excursion into Algebra	63

CONTENTS

4.1	Lecture 10. Arithmetic	63
4.2	Lecture 11. More on congruence relations	68
4.3	Lecture 12. The Chinese Remainder Theorem	73
5	Graph Theory	75
5.1	Lecture 13. Definitions and Examples	75
5.2	Lecture 14. (Two lectures) Matchings	79
5.3	Lecture 15. Ramsey Theory	82

Chapter 1

Module 1: Basics of Rigorous Mathematics

In this first module, we will cover some basics for doing rigorous mathematics.

1.1 Lecture 1. Sets and Functions

The foundations of mathematics are built on logic and though logic is used in everyday language, it can be a bit abstract and imposing when first exposed to it. To alleviate some of the unavoidable confusion, we will start somewhat informally.

Sets. Sets are one of the basic objects of mathematics. Informally, a set S is a collection of elements or objects. The objects in the set S can be essentially anything. You have encountered sets repeatedly throughout your exposure to mathematics but as they play such an important role in rigorous mathematics, we will belabor the point with some simple examples.

Example: The Natural Numbers. The natural numbers are the set of positive counting numbers:

$$\mathbf{N} = \{1, 2, 3, 4, \dots\}$$

We refer to say 17 as an element of \mathbf{N} and write $17 \in \mathbf{N}$. Note that -17 is not an element of \mathbf{N} and we denote that by $-17 \notin \mathbf{N}$. There seems to be some inconsistency in mathematics as to

1.1. LECTURE 1. SETS AND FUNCTIONS

whether or not we should include 0 in the natural numbers and it is a matter, perhaps, of taste. People often denote the natural numbers using what is called "blackboard bold" font and in that font, we have \mathbb{N} . One typically uses this font on the board but I prefer to use regular bold in printed writing such as this present document.

Example: The Integers. The **integers** are the set of all negative and positive counting along with 0:

$$\mathbf{Z} = \{0, \pm 1, \pm 2, \pm 3, \pm 4, \dots\}$$

As above, the integers are also denoted by \mathbb{Z} .

Example: The Rational Numbers. The **rational numbers** are all numbers the numbers of the form $\frac{m}{n}$ where $m, n \in \mathbf{Z}$. It is convenient to establish some short hand notation for defining sets and we will use the following scheme for the rationals:

$$\mathbf{Q} = \left\{ \frac{m}{n} : m, n \in \mathbf{Z}, n \neq 0 \right\}.$$

The colon is often read as "such that". As above, the rational numbers are also denoted by \mathbb{Q} .

Example: The Real Numbers. The **real numbers** are the set of all the real numbers like the rationals but also numbers like $\sqrt{2}$ and π that cannot be expressed as a fraction involving integers. We denote the real numbers by \mathbf{R} . One can formally construct the real numbers from the rational numbers in various ways. Note too that we constructed, in some sense, the integers and the rational numbers from the natural numbers. The existence of the natural numbers is an axiom that we assume in order to set up mathematics. These are matters that are frankly best ignored on a first exposure to these topics as one trusts that humans have some sense of what these sets mean. As with all of the above cases, we also denote the real numbers by \mathbb{R} .

Now we will give some additional examples that highlight the diversity of sets. We again start with some familiar examples.

Example. Let $a, b \in \mathbf{R}$ be real numbers such that $a \leq b$. For concreteness, you can take $a = -1$ and $b = \sqrt{2}$. We then define various open/closed intervals using a, b :

$$\begin{aligned} [a, b] &= \{x \in \mathbf{R} : a \leq x \leq b\} \\ (a, b) &= \{x \in \mathbf{R} : a < x < b\}. \end{aligned}$$

We call $[a, b]$ a closed interval and in words it is the set comprising all real numbers that are between a and b , including a, b . We call (a, b) an open interval and in words is the set of of

numbers strictly between a, b and it does not include a, b . We can define half open, half closed intervals as well like $(a, b]$ which contains b but not a or $[a, b)$ which contains a but not b .

We now move to some slightly more exotic sets. We start with one of the most important basic set called the empty set.

Example: The Empty Set. The empty set is the set with no elements and is denoted typically by \emptyset . The empty set is completely natural and also self-descriptive. However, the empty set crops up in mathematics a lot and can lead to various points of confusion. For now, I hope that it is fairly clear what it is.

Example. Now that we have a basic idea what sets are, we can start randomly creating them. Sticking with sets whose elements are numbers, we could have:

$$S = \{1, 2, 3\}$$

or

$$S = \{\pi\}.$$

When a set consists of just a single element, we sometimes refer to such sets as **singletons**.

Sets can consist of all kinds of things. Later when we look at graphs (not graphs of functions!), people consider colorings of graphs. So we may have a set of possible colors like

$$\{\text{Red, Blue, Yellow, Green}\}.$$

In economics, we may have sets of commodities like

$$\{\text{natural gas, oil, gold}\}.$$

Hopefully you get the point that sets can package together all sorts of things that one may want to consider.

Subsets. One basic way for constructing new sets from old sets is by forming subsets. A **subset** T of a set S , denoted $T \subset S$ is a set whose elements (the elements of T) are also elements of S . For example, $\mathbf{N} \subset \mathbf{Z}$ and $\mathbf{Z} \subset \mathbf{Q}$, and $\mathbf{Q} \subset \mathbf{R}$. Two simple examples of subsets of S are S itself and the empty set \emptyset . That is, $\emptyset, S \subset S$.

Example. Let us take one of our previous example sets $S = \{1, 2, 3\}$. We can list out all of the subsets of S :

$$\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}.$$

1.1. LECTURE 1. SETS AND FUNCTIONS

Power set power set size = $2^{(\text{number of set})}$ ^{Binomial theory} $= (1+1)^n = \sum (C(n,i) * 1^i * 1^{(n-i)}) = C(3,0) + C(3,1) + C(3,2) + C(3,3)$

Given a set S , we have an associated set called **the power set of S** that consists of all of the subsets of S . We denote the power set of S by $\mathcal{P}(S)$. In this example, the elements are not numbers like many of the above examples but are sets of numbers. Again, for $S = \{1, 2, 3\}$, we have

$$\mathcal{P}(S) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$

One curious comment is that the empty set \emptyset is different from the set that contains only the empty set $\{\emptyset\}$. That latter set is a singleton subset of the power set of S , for instance, and as such, is surely not empty. In the analogy that sets are something like bags, the empty set is an empty bag whereas the set containing only the empty set is a bag that contains an empty bag.

Set operations. Once we have subsets of a set S , we can form new subsets via a few basic operations called intersection and union. Given $T_1, T_2 \subset S$, we define the **union** of T_1, T_2 to be the set

$$T_1 \cup T_2 = \{s \in S : s \in T_1 \text{ or } s \in T_2\}.$$

It is the set consisting of the elements of S that are either in the subset T_1 or the subset T_2 . By definition, we have

$$T_1 \subset T_1 \cup T_2, \quad T_2 \subset T_1 \cup T_2.$$

The **intersection** of T_1, T_2 is defined to be the set

$$T_1 \cap T_2 = \{s \in S : s \in T_1 \text{ and } s \in T_2\}.$$

It is the set consisting of the elements of S that are in both T_1 and T_2 . By definition, we have

$$T_1 \cap T_2 \subset T_1, \quad T_1 \cap T_2 \subset T_2.$$

In total, we have

$$T_1 \cap T_2 \subset T_1, T_2 \subset T_1 \cup T_2.$$

Example. Keeping with our basic example set $S = \{1, 2, 3\}$, we can calculate intersections and unions of various pairs of sets. For instance, we have:

$$\begin{aligned} \{1\} \cup \{2\} &= \{1, 2\} \\ \{1\} \cap \{2\} &= \emptyset \\ \{1, 2\} \cup \{2, 3\} &= \{1, 2, 3\} \\ \{1, 2\} \cap \{2, 3\} &= \{2\}. \end{aligned}$$

Another useful operation is what is called the complement of a set. Given a subset $T \subset S$, **the complement of T** is defined to be the set

$$S - T = \{s \in S : s \notin T\}.$$

It is the set consisting of the elements of S that are not elements of T . By definition, we have

$$T \cap (S - T) = \emptyset, \quad T \cup (S - T) = S.$$

The complement is denoted several other different ways in various books. For instance, $S \setminus T$ and T^c sometimes denote the complement of T .

In taking intersections and unions, it is sometimes convenient to work in some unspecified **universal set** that contains the given sets as subsets. We will sometimes make use of this view and will denote the universal set via Ω . For the complement operation, the universal set matters since we are looking at all elements in our "universe" that are not in our given set T . The notation $S - T$ exhibits that dependence that the complement of T can only be defined once you declare what your universe is. In that sense, one should think of picking a universal set as something like making that declaration.

Functions. We next introduce functions on sets. **Given two sets X, Y , a function f from X to Y assigns to element $x \in X$, an element $f(x) \in Y$. We denote a function f from X to Y via the notation**

$$f: X \longrightarrow Y.$$

$f(x) = \text{range (shoot)} = \text{image}$
 $f^{-1}(x) = \text{preimage}$
 $x = \text{domain}$
 $y = \text{codomain}$

People will call functions by various names like maps.

Example: Identity Function and constant functions. Let $S = \{1, 2, 3\}$. We can define functions from S to itself by explicitly giving the assignment. For instance, we could have $f: S \rightarrow S$ given by

$$f(1) = 1, \quad f(2) = 2, \quad f(3) = 3.$$

More generally, if S is any set, we define the **identity function** to be the function $\text{Id}_S: S \rightarrow S$ defined by $f(s) = s$. If $s_0 \in S$ is a fixed element of S , we can define the constant function $f(s) = s_0$ that assigns to each $s \in S$ the same element s_0 . You have seen both of these functions before where $S = \mathbf{R}$.

Example: Characteristic function. Given a set S and a subset T , we define the **characteristic function of T** to be $\chi_T: S \rightarrow \{0, 1\} \subset \mathbf{R}$ by

$$\chi_T(s) = \begin{cases} 1, & s \in T \\ 0, & s \notin T. \end{cases}$$

Sometimes we think of 1,0 as true/false and in that view, the characteristic function tells you if $s \in T$ is a true or false statement. When we consider logic more formally, we will return to this example.

injective

We say that a function $f: X \rightarrow Y$ is **1-1** or **injective** if when $f(x_1) = f(x_2)$, then $x_1 = x_2$. Less formally, an injective function assigns each $x \in X$ a different $y \in Y$. The identity function is injective while a constant function is only injective if X has only one element.

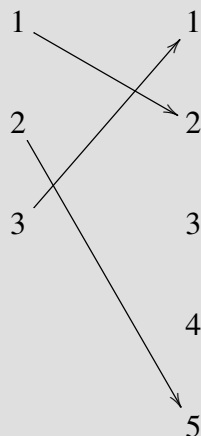
Example. Let $X = \{1, 2, 3\}$ and $Y = \{1, 2, 3, 4, 5\}$. The function $f_1: X \rightarrow Y$ defined by

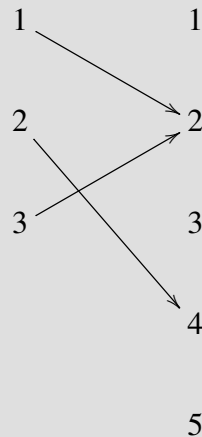
$$f_1(1) = 2, \quad f_1(2) = 5, \quad f_1(3) = 1$$

is injective while the function $f_2: X \rightarrow Y$ defined by

$$f_2(1) = 2, \quad f_2(2) = 4, \quad f_2(3) = 2$$

is not injective. We could represent these functions pictorially:





surjective

We say that a function $f: X \rightarrow Y$ is **onto** or **surjective** if for each $y \in Y$, there exists $x \in X$ such that $f(x) = y$. Less formally, a surjective function uses every Y in the assignment. As before, the identity function is surjective while the constant function is surjective only when X consists of a single element.

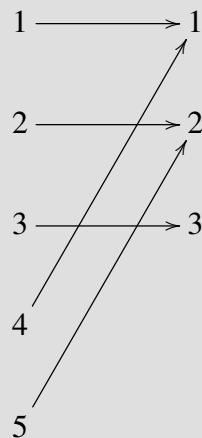
Example. Let X, Y be as before and define the functions $f_1: Y \rightarrow X$ by

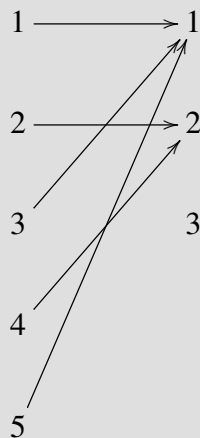
$$f_1(1) = 1, \quad f_1(2) = 2, \quad f_1(3) = 3, \quad f_1(4) = 1, \quad f_1(5) = 2.$$

This function is surjective. However, the function $f_2: Y \rightarrow X$ defined by

$$f_2(1) = 2, \quad f_2(2) = 2, \quad f_2(3) = 1, \quad f_2(4) = 2, \quad f_2(5) = 1$$

is not surjective. Again, pictorially, we have:





bijjective

We say that a function $f: X \rightarrow Y$ is **bijjective** if f is both injective and surjective. We say that two sets X, Y have the **same cardinality** if there exists a bijective function $f: X \rightarrow Y$. For our example sets X, Y above, one consequence of the Pigeon Hole Principle is that any function $f: Y \rightarrow X$ cannot be injective. This fact is because Y is "bigger" than X . Sets that have the same cardinality have the same size in some sense. |X| = |Y|

For each natural number $n \in \mathbf{N}$, we define the set $X_n = \{1, 2, 3, \dots, n\}$. We say that a set S is **finite** if there exists a bijective function $f: S \rightarrow X_n$ for some n . If S is a finite set, then we say that S has **cardinality** n if there is a bijective function $f: S \rightarrow X_n$. Once we have set up some additional mathematical language, we will prove that if S is finite, there is only one possible n for which a bijective function $f: S \rightarrow X_n$ can exist. We denote the cardinality of a finite set S by $|S|$. Finally, we say a set S is **infinite** if S is not finite. Some basic examples of infinite sets are $\mathbf{N}, \mathbf{Z}, \mathbf{Q}, \mathbf{R}$.

More on 2.6

1.2 Lecture 2. Basic Logic.

Basic logic. We now turn to **mathematical logic** and how one proves a mathematical statement. For instance, we may have a statement of the form: if T is a subset of S , then $S - T$ is a subset of S . We can decompose this statement into two parts that formally could be viewed as If P , then Q . Statements can be true or false. Some statements might be undecidable. In order to have a functioning theory, we will set our basic building blocks P to be declarative sentences that are either true or false but not both. Simple examples are $1 = 2$ or $1 < 2$. We will refer to

such things as **statements** or **propositions** (I prefer statement).

We will form **compound statements** by combining statements with some basic operations. Some of these operations are very similar to our basic set operations like union, intersection, and complement. In a compound statement, we will have one or more statements like P, Q that we will think of as **statement variables**. Whether or not a given compound statement is true or false will depend only on the validity of the statement variables. We will represent all of the possible true/false outcomes for the compound statement in tables called **true tables**.

Basic operations. Given a statement P , the **negation** of the statement P will be denoted simply by "not P ". For negation, we have the true table for a statement "not P ":

P	not P
T	F
F	T

One can think of P as some unknown statement that could either be true or false. The above table represents all of the different possibilities for P and not P . The negation operation is sometimes denoted by $\neg P$ or \bar{P} (I prefer the latter).

Our next basic operation is "or" which allows us to combine two statements P, Q into a new statement " P or Q ". We have the following true table for the "or" operation:

P	Q	P or Q
T	T	T
T	F	T
F	T	T
F	F	F

The "or" operation is analogous to the union operation for sets. The "or" operation is sometimes referred to as **disjunction** and is sometimes denoted by $P \vee Q$.

We also have an "and" operation:

P	Q	P and Q
T	T	T
T	F	F
F	T	F
F	F	F

1.2. LECTURE 2. BASIC LOGIC.

The "and" operation is analogous to the intersection operation for sets. The "and" operation is sometimes referred to as **conjunction** and is sometimes denoted by $P \wedge Q$.

It is convenient to have a slightly different "or" operation as the "or" operation we described above is sometimes referred to as **inclusive or** as it permits both P, Q being true. The **exclusive or** operation of statements P, Q will be denoted by $P \oplus Q$ and as the following truth table:

P	Q	$P \oplus Q$
T	T	F
T	F	T
F	T	T
F	F	F

Conditional statements. Our next operation will permit us to form **conditional** statements that in words represents "if P , then Q " or symbolically $P \rightarrow Q$. The truth table for the conditional statement "if P , then Q " is:

P	Q	$P \rightarrow Q$
T	T	T
T	F	F
F	T	T
F	F	T

We refer to P as the **hypothesis** and Q as the **conclusion** in the conditional statement "if P , then Q ". Note that if P is false, regardless of the validity of Q , the conditional statement $P \rightarrow Q$ is true.

The conditional statement "if P , then Q " has several useful equivalent statements that lead us to some general methods of proof. Moreover, we can form inverse or converse statements. The **contrapositive** of the statement "if P , then Q " is the statement "if not Q , then not P ". The truth table for the contrapositive is the following:

P	Q	not $Q \rightarrow$ not P
T	T	T
T	F	F
F	T	T
F	F	T

Note that the validity of "if not Q , then not P " is the same as that of "if P , then Q ". The statement "if Q , then P " is called the **converse** of the statement "if P , then Q ". We have the truth table for

the converse:

P	Q	$Q \rightarrow P$
T	T	T
T	F	T
F	T	F
F	F	T

Note that the converse is different from the original conditional statement since they behave different under the various possibilities for P, Q . We call the statement "if not P , then not Q " the **inverse** of the statement "if P , then Q ". The inverse statement is equivalent to the converse as can be seen by the truth table for the inverse:

P	Q	not $P \rightarrow$ not Q
T	T	T
T	F	T
F	T	F
F	F	T

Finally, we have the **bi-conditional** statement " P if and only if Q " that we symbolically denote by $P \leftrightarrow Q$. The truth table for the bi-conditional statement is:

P	Q	$P \leftrightarrow Q$
T	T	T
T	F	F
F	T	F
F	F	T

The bi-conditional statement " $P \leftrightarrow Q$ " is equivalent to the compound statement " $(P \rightarrow Q) \wedge (Q \rightarrow P)$ " as can be seen from the truth table of the latter:

P	Q	$P \rightarrow Q$	$Q \rightarrow P$	$(P \rightarrow Q) \wedge (Q \rightarrow P)$
T	T	T	T	T
T	F	F	T	F
F	T	T	F	F
F	F	F	F	F

Equivalence of statements. We say that a compound statement P is a **tautology** if regardless of the values of the variable statements that P depends on, the statement P is always true. A

1.2. LECTURE 2. BASIC LOGIC.

simple example of a tautology is the compound statement "P or not P" or in symbols $P \vee \bar{P}$. The true table for this statement is the following:

P	\bar{P}	$P \vee \bar{P}$
T	F	T
F	T	T

We say that a compound statement P is a **contradiction** if regardless of the values of the variable statements that P depends on, the statement P is always false. A simple example of a contradiction is the compound statement "P and not P" or in symbols $P \wedge \bar{P}$. The truth table for this statement is the following:

P	\bar{P}	$P \wedge \bar{P}$
T	F	F
F	T	F

Finally, we say that two compound statements P, Q are **equivalent** if the compound statement $P \leftrightarrow Q$ is a tautology. We denote the equivalence of P, Q by $P \Leftrightarrow Q$. Above, we used this notion when we said that the statements like $P \rightarrow Q$ and $\bar{Q} \rightarrow \bar{P}$ are equivalent. For that example, we have the following truth table:

P	Q	$P \rightarrow Q$	$\bar{Q} \rightarrow \bar{P}$	$(P \rightarrow Q) \leftrightarrow (\bar{Q} \rightarrow \bar{P})$
T	T	T	T	T
T	F	F	F	T
F	T	T	T	T
F	F	T	T	T

Two important equivalences between statements that we will need are the following that are referred to as **De Morgan's Laws**:

$$\begin{aligned}\overline{P \vee Q} &\Leftrightarrow \bar{P} \wedge \bar{Q} \\ \overline{P \wedge Q} &\Leftrightarrow \bar{P} \vee \bar{Q}.\end{aligned}$$

In words, "not (P or Q)" is equivalent to "(not P) and (not Q)". Similarly, "not (P and Q)" is equivalent to "(not P) or (not Q)". We have the following truth tables that verify these assertions:

P	Q	$\overline{P \vee Q}$	$\bar{P} \wedge \bar{Q}$
T	T	F	F
T	F	F	F
F	T	F	F
F	F	T	T

Similarly, we have:

P	Q	$\overline{P \wedge Q}$	$\overline{P} \vee \overline{Q}$
T	T	F	F
T	F	T	T
F	T	T	T
F	F	T	T

Another important equivalence is at the heart of a proof method known as "proof by contradiction". Namely, we have the equivalence

$$P \rightarrow Q \Leftrightarrow \overline{P} \vee Q.$$

Combining this equivalence with De Morgan's Law, we see the following equivalence:

$$\overline{P \rightarrow Q} \Leftrightarrow P \wedge \overline{Q}.$$

Methods of Proof. In mathematics, often we state theorems using "if/then" language. For instance, say x is a real number. If $x > 1$, then $x^2 > 1$. What does it mean for this to be a true statement? We want the statement "If $x > 1$ ", then $x^2 > 1$ " to always be true. We know that if " $x > 1$ " is false, regardless of whether " $x^2 > 1$ " is true or false, the statement "if $x > 1$ ", then $x^2 > 1$ " is true. So we must prove that if " $x > 1$ " is true, then " $x^2 > 1$ " is also true. This method is known as the **direct method**. Alternatively, we know that "if $x > 1$, then $x^2 > 1$ " is equivalent to the contrapositive statement "if $x^2 \leq 1$, then $x \leq 1$ ". If we verify the contrapositive is always true, which again only requires that we check that if " $x^2 \leq 1$ " is true, then " $x \leq 1$ " is also true, we know that our original statement "if $x > 1$, then $x^2 > 1$ " is always true. This method is known as the **contrapositive method**. In proof by contradiction, we instead want to verify that "not(if $x > 1$, then $x^2 > 1$)" is always false. As this statement is equivalent to the statement " $x > 1$ and $x^2 \leq 1$ ", we then can check instead that that statement is always false. In proof by contradiction, one instead assumes that it is true and then derives some contradiction. This method is referred to as **proof by contradiction**.

In summary, in trying to verify that "if P , then Q " is true, we have (at least) three broad methods. The direct method which shows (using existing mathematical knowledge) that if P is true, then Q is true. The contrapositive method which shows (using existing mathematical knowledge) if not Q is true, then not P is true. Last, proof by contradiction, which assumes that P and not Q is true and then (using existing mathematical knowledge) shows a contradiction. Each method has its uses though for those new to proofs, the direct method should be the default. We will have a vast multitude of examples of proofs throughout this course for you to build up some experience and intuition with.

1.3 Lecture 3. Induction.

The Principal of Induction. In the previous lecture, we discussed some methods of proof. Today, we discuss another method of proof that can be applied to certain types of problems. The method of proof is based on the principal of induction and is often referred to an **induction** or an **inductive proof** or **proof by induction**.

The Principal of Induction. Let $S \subset \mathbf{N}$ be a subset of the natural numbers such that the following two properties are satisfied:

(a) The element 1 is in S or symbolically $1 \in S$.

(b) If $k \in S$, then $k + 1 \in S$.

Then $S = \mathbf{N}$.

The principal of induction is clear intuitively and will be assumed to hold. Here is a rough idea of how one may verify that the principal of induction holds. Our argument will be via contradiction and will use some essential arithmetic structure. We want to prove that $S = \mathbf{N}$. If this assertion does not hold, then $\mathbf{N} - S$ is not the empty set. Let $k_0 \in \mathbf{N} - S$ be the smallest such element. Since k_0 is the smallest element, $k_0 - 1$ must be an element of S . By (b), we know that $(k_0 - 1) + 1 \in S$. However, $(k_0 - 1) + 1 = k_0$, contradicting the fact that $k_0 \notin S$. Since we arrived at a contradiction from assuming $\mathbf{N} - S$ is not the empty set, we conclude that $\mathbf{N} - S$ is the empty set. Equivalently, $S = \mathbf{N}$.

Remark. The condition (a) (or (a') below) is usually referred to as the **base case for induction** while the condition (b) is referred to as **the induction step**. The assumption that $k \in S$ in (b) is referred to as the **induction hypothesis**.

In some applications, it may be false that $1 \in S$. If we instead know that:

(a') $n_0 \in S$ for some integer $n_0 \in \mathbf{N}$.

(b) If $k \in S$, then $k + 1 \in S$.

We can conclude that $\{n_0, n_0 + 1, n_0 + 2, \dots\} \subset \mathbf{N}$. In practice, one typically just wants to know

for the set S that there exists some positive integer n_0 such that if $n \geq n_0$, then $n \in S$. The augmented condition (a') establishes this condition for the set S .

Remark. The condition (a) seems silly in that condition (b) clearly contains most of the substance of the principal of induction. Indeed, (a) is simply to rule out the empty set as the empty set always satisfies condition (b).

Okay, we have this principal of induction and now we should see what it is good for. Before giving some basic examples, we will establish some notation that we will use throughout the remainder of this document. It is likely notation that you have seen before. Nevertheless, for completeness, we review it here.

By a **sequence** on a set X , we mean a function

$$f: \mathbf{N} \longrightarrow X.$$

Traditionally, we write $x_j = f(j)$ and refer to $\{x_j\}$ as the sequence. To indicate the role of X , we will say $\{x_j\}$ is a sequence on X or write $\{x_j\} \subset X$.

Remark. There are generalizations of sequences called **nets** that replace the set \mathbf{N} with other sets Λ that are endowed with a certain amount of structure (a partial ordering). Also, the concept of a net is related to the concept of a **filter**.

Given a sequence $\{x_j\}$ on \mathbf{R} , we can build new sequences using arithmetic. For instance, the sequence of n th partial sums is defined to be the sequence

$$y_n = x_1 + x_2 + x_3 + \cdots + x_n = \sum_{j=1}^n x_j.$$

We can define the sequence of n th partial products as well to be

$$z_n = x_1 x_2 x_3 \cdots x_n = \prod_{j=1}^n x_j.$$

The operator \sum is referred to as the **summation operator** and indicates addition while the operator \prod is referred to as the **product operator** and indicates multiplication. For the sequence

$$x_n = n,$$

1.3. LECTURE 3. INDUCTION.

the partial products of this sequence appear often enough in mathematics that we have a short-hand notation for them. Specifically, we define the **factorial** operator to be

$$n! = \prod_{j=1}^n j = 1 \cdot 2 \cdot 3 \cdot \dots \cdot (n-1) \cdot n.$$

Okay, we now give a few examples of some induction proofs. We start with a basic, famous example:

Example. Prove that for every positive integer $n \in \mathbf{N}$, the equation

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

holds.

As this example is our first induction example, I will work it painfully slow and write it in a way that should not be copied. In fact, I will give the more "professional" proof after the painful one. For sake of style only, I will denote the end of a proof with the symbol ♠ though typically one uses \square or \blacksquare .

Painful Proof. Let S be the subset of \mathbf{N} comprised of positive integers n such that

$$\sum_{j=1}^n j = \frac{n(n+1)}{2}. \tag{1.1}$$

Our goal is to prove that $S = \mathbf{N}$ and we will achieve that goal by using the principal of mathematical induction. By the induction principal, we must show two things:

- (a) $1 \in S$.
- (b) If $k \in S$, then $k+1 \in S$.


We start with the verification of (a).

Proof that (a) is true. To see that (a) holds, we simply write out what each side of (1.1) is for $n = 1$. The left hand side of (1.1) is:

$$\sum_{j=1}^1 j = 1.$$

The right hand side of (1.1) is:

$$\frac{1(1+1)}{2} = \frac{2}{2} = 1.$$

Since $1 = 1$ is true, we see that (a) holds. 

Next, we verify that (b) holds.

Proof that (b) is true. To see that (b) is true, we assume that $k \in S$ and must show that $k + 1 \in S$. Explicitly, this information tells us that

$$\sum_{j=1}^k j = \frac{k(k+1)}{2} \tag{1.2}$$

and we want to verify that

$$\sum_{j=1}^{k+1} j = \frac{(k+1)(k+2)}{2}. \tag{1.3}$$

In particular, our primary goal is verifying that (1.3) is true and we are allowed to use (1.2). Now, in an induction proof, one must translate the information that $k \in S$ to some useable information for verifying that $k + 1 \in S$. Each induction problem is different in that regard. Here, we relate them by relating the left hand sides of (1.2) and (1.3). Specifically, we have

$$\sum_{j=1}^{k+1} j = k + 1 + \sum_{j=1}^k j. \tag{1.4}$$

Now, using (1.2) with (1.4), we see that

$$\sum_{j=1}^{k+1} j = k + 1 + \frac{k(k+1)}{2}. \tag{1.5}$$

1.3. LECTURE 3. INDUCTION.

Using basic arithmetic on the right hand side of (1.5), we obtain

$$\begin{aligned}k + 1 + \frac{k(k+1)}{2} &= \frac{2(k+1)}{2} + \frac{k(k+1)}{2} \\&= \frac{2(k+1) + k^2 + k}{2} \\&= \frac{k^2 + 3k + 2}{2} \\&= \frac{(k+1)(k+2)}{2}.\end{aligned}$$

In particular,

$$k + 1 + \frac{k(k+1)}{2} = \frac{(k+1)(k+2)}{2}. \quad (1.6)$$

Using (1.5) and (1.6), we conclude that (1.3) is true. ♠

Finally, since S satisfies (a) and (b) in the principal of induction, we conclude that $S = \mathbb{N}$ and so (1.1) holds for all positive integers n . ♠

A research mathematician in a research paper would likely prove it as follows.

Research Paper Proof. The verification of (1.1) is a straightforward exercise that we leave to the reader. ♠

However, you are not a researcher yet and so I will give a professional proof now.

Professional Proof of (1.1). We prove (1.1) via induction. We first check that (1.1) holds in the base case $n = 1$. To that end, inserting $n = 1$ into both sides of (1.1), we have

$$1 = \sum_{j=1}^1 j = \frac{1(2)}{2} = 1.$$

Next, we verify that the induction step. To that end, we have the induction hypothesis

$$\sum_{j=1}^k j = \frac{k(k+1)}{2}$$

and must deduce that

$$\sum_{j=1}^{k+1} j = \frac{(k+1)(k+2)}{2}.$$

The verification is achieved with following string of equalities:

$$\begin{aligned} \sum_{j=1}^{k+1} j &= k+1 + \sum_{j=1}^k j = k+1 + \frac{k(k+1)}{2} \\ &= \frac{k^2 + 3k + 2}{2} = \frac{(k+1)(k+2)}{2}. \end{aligned}$$



A nifty application of (1.1) is that the sum

$$\sum_{j=1}^n \frac{j}{n},$$

which is the average of the first n integers, is

$$\frac{n+1}{2},$$

which is the average of the first and last terms of the set $\{1, \dots, n\}$.

1.3. LECTURE 3. INDUCTION.

Chapter 2

Counting

2.1 Lecture 4. Binomial Theorem and some basics in counting.

One basic result that you have likely seen is the **Binomial Theorem** that states that

$$(x + y)^n = \sum_{j=0}^n \binom{n}{j} x^{n-j} y^j$$

where the number

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

are sometimes referred to as the **binomial coefficients**. In order to establish the Binomial Theorem, we will introduce some basic methods in counting. The mathematics of counting is often referred to as **combinatorics** and has many applications in mathematics as well as being a subject of interest in its own right.

Given a set X with $|X| = n$, we may want to **select from X , in various ways**, subsets of X of some specified cardinality/size r . For example, we may want to **order** the set X in some way:

$$X = \{x_1, x_2, \dots, x_n\}.$$

2.1. LECTURE 4. BINOMIAL THEOREM AND SOME BASICS IN COUNTING.

There are many ways of selecting an ordering of the elements of X . Let us investigate the number of possible orderings to start. We will use a simple though fundamental approach. However, before doing so, let us consider a few simple examples.

Example. Let $X = \{x_1, x_2\}$ have cardinality 2. We see that there are two possible orderings of X . Namely:

$$\{x_1, x_2\} \text{ or } \{x_2, x_1\}.$$

Example. Let $X = \{x_1, x_2, x_3\}$ have cardinality 3. We see that there are six possibilities for orderings of X . Namely:

$$\{x_1, x_2, x_3\}, \{x_1, x_3, x_2\}, \{x_2, x_1, x_3\}, \{x_2, x_3, x_1\}, \{x_3, x_1, x_2\}, \text{ or } \{x_3, x_2, x_1\}.$$

The first example is perhaps too simple to derive much useful information for. However, we can think about the strategy that we took for listing off all of the possible orderings in the second, slightly more complicated example. We first chose x_1 to be first and then used the fact that we had two possible orderings for the remaining two elements x_2, x_3 . Next, we chose x_2 to be first and then used the fact that we had two possible orderings for the remaining two elements x_1, x_3 . Finally, we chose x_3 to be first and then used the fact that we had two possible orderings for the remaining two elements x_1, x_2 . In fact, once we choose the first element, we are reduced to the lower complexity case of ordering two elements. We have three choices for which one to make first and each gives rise to the number of orderings on two elements. Thus, we have

two examples to get permutation Number of orders of $X = 2 + 2 + 2 = 3 \cdot 2 = 6$.

What about the case when $|X| = 4$? Well, we can proceed as above. We have four choices for which element to make first. Upon making that choice, we have the number of ways of ordering a set of size three possibilities. Let $P(n, n)$ denote the number of ways of ordering a set with n elements. We see that

$$P(4, 4) = 4 \cdot P(3, 3) = 4 \cdot 3 \cdot P(2, 2) = 4 \cdot 3 \cdot 2 = 4!.$$

Of course, we saw already that $P(3, 3) = 6 = 3!$.

Claim: $P(n, n) = n!$. permutation def # of functions $X \rightarrow 1-1$ and onto $\rightarrow X$

Now we have only checked three cases but the strategy we took to compute those examples can be implemented in an induction proof.

Proof of Claim. We will proceed by **induction**. The base case $n = 1$ is trivial since there is exactly one way of ordering a set with 1 element. We **assume that $P(n, n) = n!$** (the induction hypothesis) and then must **show that**

$$P(n+1, n+1) = (n+1)!.$$

To that end, for each $x \in X$, we consider the number of orderings of X where x appears first. If x appears first, then we get an ordering on the remaining n elements and by our induction hypothesis, we know that there are $n!$ possibilities. So each x gives $n!$ orderings where it appears first. We have $n+1$ possible choices for x and each gives $n!$ orderings and so we see that

$$P(n+1, n+1) = \sum_{j=1}^{n+1} n! = (n+1)n! = (n+1)!.$$



===== $P(n, n) \rightarrow P(n, r)$ =====

We next consider a slightly more complicated problem. In a set X of cardinality n , how many ways can we **pick r ordered elements from X** ? These r ordered elements correspond to order subsets of X of cardinality r . We denote this number by $P(n, r)$. We again will deduce the number of such orderings with a similar strategy. We will start by recasting our computation of $P(n, n)$ in a slightly different way. In essence, we will build all possible orderings of n . First, we have n choices for the first element. Upon making such a choice, we have $n-1$ possible choices for the second element. Upon making our second selection, we have $n-2$ choices for the third element. At the j th stage, we will have $n-j+1$ choices and so at the $n-2$ stage, we will have $n-(n-3)=3$ choices. At the $n-1$ stage, we have $n-(n-2)=2$ choices. After that selection, we have only one element left and so must choose it. Now, to get the total number of choices, we multiply these numbers together (you should think about why that is the case!) and so we see that

$$P(n, n) = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 3 \cdot 2 = n!.$$

Now, if we want to make an ordered set of r elements from the elements of X , we can proceed identically. We select a first element and we have n choices for such. We select a second element and we have $n-1$ choices for such. **Continuing, at the r th stage, we have $n-r+1$ possibilities for the final selection and thus conclude that**

$$P(n, r) = n(n-1) \dots (n-r+1).$$

Of course, if $r > n$, it is impossible since X does not have r elements to select. For reasons that we will see shortly, note that

$$P(n, r) = n(n-1) \dots (n-r+1) = \frac{n!}{(n-r)!}.$$

2.1. LECTURE 4. BINOMIAL THEOREM AND SOME BASICS IN COUNTING.

Now, we may want to select r elements from X and disregard the order. We will denote the number of such choices by $C(n, r)$ when $|X| = n$. Note that each ordered collection of r elements of X gives rise to an unordered collection of r elements. From above, we know that there are $r!$ ways of ordering this set of r elements and so we see that each collection of r elements in X has $r!$ ordered sets associated to it. In particular, we must have

ignore the ordering $/r!$

$$C(n, r) = \frac{P(n, r)}{r!} = \frac{n!}{(n-r)!r!} = \binom{n}{r}.$$

Remark. The number $P(n, r)$ is sometimes referred to as the number of r -permutations of a set of order n and one reason why we use the P in our notation. Similarly, $C(n, r)$ is sometimes referred to as the number of r -combinations of a set of order n and one reason why we use the C in our notation. Of course, the numbers $C(n, r)$ are also referred to as the binomial coefficients.

Remark. In our computation of the numbers $C(n, r)$, we made, in some sense, use of a surjective function

$$F: \{r\text{-permutations}\} \longrightarrow \{r\text{-combination}\}$$

where the function F sends an ordered set of r elements of X to the associated subset of r elements by forgetting the order. Since every subset of r elements can be ordered, the function is onto (surjective). We also noted that the size of the set

$$F^{-1}(R) = \{r\text{-permutations that have } R \text{ as their associated set}\}$$

had size $r!$ since a set of size r has $r!$ different orderings. This method of computation is actually much more broad than the specific example here and utilizes group actions. In particular, this calculation is a special case of a theorem in group theory called the Orbit-Stabilizer Theorem. You can read about **group actions** for more on this.

We have a few important identities involving $C(n, r)$ that we mention here. First,

$$C(n, r) = C(n, n-r).$$

You can see this equality by the fact that

$$C(n, n-r) = \frac{n!}{(n-r)!(n-(n-r))!} = \frac{n!}{(n-r)!r!} = C(n, r).$$

Another way to see that this equality holds is a bit cuter. Let $\mathcal{P}_r(X)$ denote the set of subsets of X with cardinality r . If $|X| = n$, it follows by definition of $C(n, r)$ that

$$|\mathcal{P}_r(X)| = C(n, r).$$

We have a function

$$F: \mathcal{P}_r(X) \longrightarrow \mathcal{P}_{n-r}(X)$$

given by

$$F(R) = X - R.$$

We also have the function

$$G: \mathcal{P}_{n-r}(X) \longrightarrow \mathcal{P}_r(X)$$

given by

$$G(R) = X - R.$$

Since $G \circ F$ is the identity function of $\mathcal{P}_r(X)$ and $F \circ G$ is the identity function of $\mathcal{P}_{n-r}(X)$, the functions F, G are bijections and so

$$C(n, r) = |\mathcal{P}_r(X)| = |\mathcal{P}_{n-r}(X)| = C(n, n - r).$$

We now prove the Binomial Theorem.

Proof of the Binomial Theorem. We must prove that

proof 1

$$(x + y)^n = \sum_{j=0}^n \binom{n}{j} x^{n-j} y^j.$$

On the left hand side, we have

$$(x + y)^n = \underbrace{(x + y)(x + y) \dots (x + y)}_{n \text{ times}}.$$

If we expand this product, notice that each term in the expansion corresponds to selecting some number of the x 's and some number of the y 's where the total number of x 's and y 's is n . In particular, the number of $x^{n-j} y^j$ terms corresponds to choosing $n - j$ of the x 's. Note that once you select the places where you will select the x 's, that determines the places where you will select the y 's. Since the number of different ways of choosing $n - j$ elements from a set of n

2.1. LECTURE 4. BINOMIAL THEOREM AND SOME BASICS IN COUNTING.

elements is $C(n, n-j) = C(n, j)$, we see that we have $C(n, j)$ terms in the expansion of the form $x^{n-j}y^j$. Therefore,

$$(x+y)^n = \sum_{j=0}^n \binom{n}{j} x^{n-j} y^j.$$



proof 2 inductive

We could **instead** give an **inductive proof**. The case $n = 1$ is easy to check. If we assume that the Binomial Theorem holds for n , we have

$$(x+y)^n = \sum_{j=0}^n \binom{n}{j} x^{n-j} y^j$$

and we must prove that

$$(x+y)^{n+1} = \sum_{j=0}^{n+1} \binom{n+1}{j} x^{n+1-j} y^j.$$

Since

$$(x+y)^{n+1} = (x+y)(x+y)^n,$$

we see by our induction hypothesis that

$$\begin{aligned} (x+y)^{n+1} &= (x+y)(x+y)^n \\ &= x \left(\sum_{j=0}^n \binom{n}{j} x^{n-j} y^j \right) + y \left(\sum_{j=0}^n \binom{n}{j} x^{n-j} y^j \right) \\ &= \sum_{j=0}^n \binom{n}{j} x^{n+1-j} y^j + \sum_{j=0}^n \binom{n}{j} x^{n-j} y^{j+1}. \end{aligned}$$

We see in the second sum that when we take $j-1$ as our index, we get the term

$$\binom{n}{j-1} x^{n-(j-1)} y^{(j-1)+1} = \binom{n}{j-1} x^{n+1-j} y^j.$$

In particular, we have

$$(x+y)^{n+1} = \sum_{j=0}^{n+1} \left[\binom{n}{j} + \binom{n}{n-j} \right] x^{n+1-j} y^j.$$

Note that $C(n, r) = 0$ if $r > n$ or $r < 0$. In particular, if the Binomial Theorem is true, we must have the identity

$$\binom{n+1}{j} = \binom{n}{j} + \binom{n}{j-1}. \quad (2.1)$$

This identity is often referred to as **Pascal's identity**. We leave the verification of (2.1) to the reader.

power set size = $2^{\text{number of set}} = (1+1)^n = \sum (C(n,i) * 1^i * 1^{(n-i)})$

The Binomial Theorem has some interesting applications. We note two here. First

$$2^n = (1+1)^n = \sum_{j=0}^n \binom{n}{j} 1^{n-j} 1^j = \sum_{j=0}^n \binom{n}{j}.$$

Now, if X is a set of size n , then $C(n, j) = |\mathcal{P}_j(X)|$ is the number of subsets of X of size j . The total number of subsets of X , namely the cardinality $|\mathcal{P}(X)|$ of the power set of X , is the sum

$$|\mathcal{P}(X)| = \sum_{j=0}^n \binom{n}{j} = 2^{|X|}. \quad (2.2)$$

We also have

$$(1+a)^n = \sum_{j=0}^n a^j \binom{n}{j}.$$

Taking $a = -1$, we obtain

$$\sum_{j=0}^n (-1)^j \binom{n}{j} = 0.$$

2.2 Lecture 5. More on sets and functions.

In this lecture, we will return to some basics on sets, cardinalities of sets, and functions. Given a pair of sets X, Y and a function $f: X \rightarrow Y$, we call X the **domain of f** and we call Y the **codomain of f** . The **image of f** is the set $f(X)$ and is sometimes referred to as the **range of f** ; the image of a subset $T \subset X$ also makes sense and is denoted by $f(T)$. For a function $f: X \rightarrow Y$ and a subset $S \subset Y$, we define

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

We often refer to $f^{-1}(S)$ as the **preimage of S under f** .

2.2. LECTURE 5. MORE ON SETS AND FUNCTIONS.

Given sets X, Y, Z and functions $f: X \rightarrow Y, g: Y \rightarrow Z$, the **composition** of f, g is the function

$$g \circ f: X \rightarrow Z$$

defined by

$$(g \circ f)(x) = g(f(x)).$$

Given functions $f: X \rightarrow Y$ and $g: Y \rightarrow X$, we say that g is the **inverse** of f if

$$g \circ f = \text{Id}_X, \quad f \circ g = \text{Id}_Y,$$

where Id_X, Id_Y are the identity functions on X, Y , respectively. That is

$$(g \circ f)(x) = x, \quad (f \circ g)(y) = y$$

for all $x \in X$ and $y \in Y$. We say that f is **invertible** if f has an inverse.

Lemma 2.1. *A function $f: X \rightarrow Y$ is invertible if and only if f is 1-1 and onto.*

Proof. In order to prove this bi-conditional statement, we have two parts to verify. We first verify the direct implication that is sometimes referred to as the "if" part.

Part 1: If f is invertible, then f is 1-1 and onto.

Since f is invertible, there exists an inverse $g: Y \rightarrow X$. We first show that f is onto. Given $y \in Y$, we seek $x \in X$ such that $f(x) = y$. We know that $f(g(y)) = y$ and so we can take $x = g(y)$. Next, we show that f is 1-1. Given $x_1, x_2 \in X$ with $f(x_1) = f(x_2) = y$, we must show that $x_1 = x_2$. We know that $g(f(x)) = x$ for all $x \in X$. In particular,

$$x_1 = g(f(x_1)) = g(y) = g(f(x_2)) = x_2.$$

We now check the reverse or converse implication that is sometimes referred to as the "only if" part.

Part 2: If f is 1-1 and onto, then f is invertible.

To check that f is invertible, we must construct an inverse function $g: Y \rightarrow X$ such that

$$g \circ f = \text{Id}_X, \quad f \circ g = \text{Id}_Y.$$

Since f is onto, we know for each $y \in Y$ that $f^{-1}(\{y\})$. Moreover, since f is 1-1, we know that $f^{-1}(\{y\}) = \{x\}$. We define $g(y) = x$ where $f(x) = y$. By definition, $g \circ f = \text{Id}_X$ and $f \circ g = \text{Id}_Y$. ♠

We often denote the inverse function of an invertible function $f: X \rightarrow Y$ by f^{-1} . Note that the preimage $f^{-1}(S)$ makes sense for any function $f: X \rightarrow Y$ and any subset $S \subset Y$, while the inverse function does not always exist. It is unfortunate that the notation that is traditionally used can lead to confusion.

We now verify a few basic set results: See [here](#) for a lot more formulas/identities involving sets.

Lemma 2.2. *Let $A, B, C \subset \Omega$.*

(a)

$$A \cup B = B \cup A, \quad A \cap B = B \cap A.$$

(b)

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C), \quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

(c)

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

(d)

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

(e)

$$\overline{A \cup B} = \overline{A} \cap \overline{B}.$$

(f)

$$\overline{A \cap B} = \overline{A} \cup \overline{B}.$$

Proof. We will not prove all of these as most are quite basic.

Part (a): We have

$$A \cup B = \{x \in \Omega : x \in A \text{ or } x \in B\} = \{x \in \Omega : x \in B \text{ or } x \in A\} = B \cup A.$$

The intersection version is similar.

Part (b): We have

$$\begin{aligned} A \cup (B \cap C) &= \{x \in \Omega : x \in A \text{ or } x \in B \cap C\} \\ &= \{x \in \Omega : x \in A \text{ or } x \in B \text{ or } x \in C\} \\ &= \{x \in \Omega : x \in A \cup B \text{ or } x \in C\} \\ &= (A \cup B) \cup C. \end{aligned}$$

The intersection version is similar.

Part (c): We want to prove that

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

We will prove this fact by proving

$$A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C), \quad (A \cap B) \cup (A \cap C) \subset A \cap (B \cup C).$$

Let $x \in A \cap (B \cup C)$. Then $x \in A$ and $x \in B \cup C$. Of course, if $x \in B \cup C$, then $x \in B$ or $x \in C$ (perhaps both happen). In particular, $x \in A \cap B$ or $x \in A \cap C$. To check the reverse inclusion, we argue similarly. Take $x \in (A \cap B) \cup (A \cap C)$. Then $x \in A \cap B$ or $x \in A \cap C$. In particular, x is in A and B or x is in A and C . As one of these case must happen, we see that x must be in $A \cap (B \cup C)$. Furthermore, x must also be in B or C , and so $x \in A \cap (B \cup C)$.

Part (d) is similar to Part (c). Parts (e) and (f) are the set analogs of De Morgan's formulas for and/or under negation. ♠

Remark: We can define the intersection and union of finitely many sets $A_1, \dots, A_n \subset \Omega$ recursively and by Lemma 2.1 (b). We denote the union and intersection of A_1, \dots, A_n by

$$\bigcup_{j=1}^n A_j, \quad \bigcap_{j=1}^n A_j.$$

We can prove (using induction) that

$$A \cap \left(\bigcup_{j=1}^n A_j \right) = \bigcup_{j=1}^n (A \cap A_j), \quad A \cup \left(\bigcap_{j=1}^n A_j \right) = \bigcap_{j=1}^n (A \cup A_j).$$

In the above, we used a basic fact in (c). Namely, if $A \subset B$ and $B \subset A$, then $A = B$. One of the basic methods for showing two sets A, B are equal is to prove $A \subset B$ and $B \subset A$.

Given two sets A, B , we define the **product** or **cartesian product** of A, B to be

$$A \times B = \{(a, b) : a \in A, b \in B\}.$$

We have **projection functions**

$$P_A : A \times B \longrightarrow A, \quad P_B : A \times B \longrightarrow B$$

given by

$$P_A((a, b)) = a, \quad P_B((a, b)) = b.$$

If we fix $a_0 \in A$ and $b_0 \in B$, we also have inclusion functions

$$\iota_{b_0}: A \longrightarrow A \times B, \quad \iota_{a_0}: B \longrightarrow A \times B$$

given by

$$\iota_{b_0}(a) = (a, b_0), \quad \iota_{a_0}(b) = (a_0, b).$$

Claim: $P_A \circ \iota_{b_0} = \text{Id}_A$.

We must show that $P_A \circ \iota_{b_0}(a) = a$ for all $a \in A$. To that end, we have

$$P_A \circ \iota_{b_0}(a) = P_A(\iota_{b_0}(a)) = P_A((a, b_0)) = a.$$

Likewise, we have $P_B \circ \iota_{a_0} = \text{Id}_B$. Note that as long as neither A or B is a singleton (has just one element), the maps P_A, P_B are not 1-1 and the maps ι_{a_0}, ι_{b_0} are not onto. So in Lemma 2.1, we really do need both conditions $f \circ g = \text{Id}_Y$ and $g \circ f = \text{Id}_X$.

Lemma 2.3. *If A, B are finite sets, then so is $A \times B$. Moreover, $|A \times B| = |A| \cdot |B|$.*

See [here](#) for this formula and many more (some of which appear below). We will prove Lemma 2.3 using a related counting result. We say that two sets $X, Y \subset \Omega$ are **disjoint** if $X \cap Y = \emptyset$.

Lemma 2.4. *If $A_1, \dots, A_n \subset \Omega$ are disjoint, finite sets, then*

$$A = \bigcup_{j=1}^n A_j$$

is a finite set. Moreover,

$$|A| = \sum_{j=1}^n |A_j|.$$

Proof of Lemma 2.4. We will prove the lemma via (strong) induction on n . For $n = 1$, we have the trivial statement $|A_1| = |A_1|$. Now, we assume the result holds for all $j \leq n$ finite sets and must show that it holds for $n + 1$. To that end, we have A_1, \dots, A_n, A_{n+1} disjoint sets and seek to show

$$\left| \bigcup_{j=1}^{n+1} A_j \right| = \sum_{j=1}^{n+1} |A_j|.$$

2.2. LECTURE 5. MORE ON SETS AND FUNCTIONS.

Set

$$B = \bigcup_{j=1}^n A_j, \quad A = A_{n+1}.$$

By our induction hypothesis, we have that

$$|B| = \sum_{j=1}^n |A_j|.$$

We assert that A, B are disjoint. To see that, we have

$$\begin{aligned} A \cap B &= A \cap \left(\bigcup_{j=1}^n A_j \right) = \bigcup_{j=1}^n (A \cap A_j) \\ &= \bigcup_{j=1}^n (A_{n+1} \cap A_j) = \bigcup_{j=1}^n \emptyset = \emptyset. \end{aligned}$$

Now, A, B are disjoint finite sets and so by our induction hypothesis,

$$|A \cup B| = |A| + |B|.$$

Of course, we have

$$\left| \bigcup_{j=1}^{n+1} A_j \right| = |A \cup B| = |A| + |B| = |A_{n+1}| + \sum_{j=1}^n |A_j| = \sum_{j=1}^{n+1} |A_j|.$$



We next use Lemma 2.4 to prove Lemma 2.3.

Proof of Lemma 2.3. For each $a \in A$, let

$$B_a = \{(a, b) : b \in B\}.$$

The function $f: B_a \rightarrow B$ defined by $f((a, b)) = b$ is 1-1 and onto and so $|B_a| = |B|$. Notice also that the sets $B_a, B_{a'}$ are disjoint for $a \neq a'$. Finally, notice that

$$A \times B = \bigcup_{a \in A} B_a.$$

Therefore,

$$|A \times B| = \sum_{a \in A} |B_a| = \sum_{a \in A} |B| = |A| \cdot |B|.$$



Lemma 2.5. *If A, B are finite sets and $B \subset A$, then $|A - B| = |A| - |B|$.*

We leave this lemma as an exercise.

Proposition 2.6. *If A, B are finite sets, then $A \cup B, A \cap B$ are finite sets. Moreover,*

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Proof. It is clear that $A \cap B$ is a finite set since $A \cap B \subset A$ and subsets of finite sets are finite. Set

$$C = A - (A \cap B), \quad D = B - (A \cap B), \quad E = A \cap B.$$

We assert that C, D, E are disjoint, finite sets. The finiteness follows from Lemma 2.5. That they are disjoint is left as an exercise. We know by Lemma 2.4 that

$$|C \cup D \cup E| = |C| + |D| + |E|.$$

We know by Lemma 2.5 that

$$|C| = |A| - |A \cap B|, \quad |D| = |B| - |A \cap B|.$$

We again leave it as an exercise to verify that

$$C \cup D \cup E = A \cup B.$$

Thus

$$|A \cup B| = |A| - |A \cap B| + |B| - |A \cap B| + |A \cap B| = |A| + |B| - |A \cap B|.$$



2.3 Lecture 6. More on counting

In the previous lecture, we investigated the cardinalities of various finite sets that arose from operations on finite sets like intersection, union, and product. Today, we continue that theme further by reinterpreting some of that material in terms of counting problems and we will also see some new tools like the Pigeon-Hole Principal.

One of the basic results from last lecture was the following. If $A_1, \dots, A_n \subset \Omega$ are finite sets and for each i, j with $i \neq j$, we have $A_i \cap A_j = \emptyset$, then

$$\left| \bigcup_{j=1}^n A_j \right| = \sum_{j=1}^n |A_j|.$$

Recall when this condition on intersections is satisfied, we said that the sets A_1, \dots, A_n are disjoint (or pairwise disjoint). Let

$$m = \left| \bigcup_{j=1}^n A_j \right|, \quad m_j = |A_j|.$$

We know that each element $a \in A_j$ contributes an element to the union since the union is the set of elements that is contained in one of the A_j . So we can try to "build" the set

$$A = \bigcup_{j=1}^n A_j$$

in stages. First, we have

$$A_1 = \{a_{1,1}, \dots, a_{1,m_1}\}.$$

We know that these elements are in A . Next, we add the elements from A_2 . If

$$A_2 = \{a_{2,1}, \dots, a_{2,m_2}\}$$

we can form the set

$$A'_2 = \{a_{1,1}, \dots, a_{1,m_1}, a_{2,1}, \dots, a_{2,m_2}\}.$$

The worry here is that we may have included an element more than once. If $a \in A_1 \cap A_2$, it appears in both list but only contributes once as an element of A . However, since $A_1 \cap A_2 = \emptyset$, we know that we have not repeated any elements. Next, we add the elements from A_3 . If

$$A_3 = \{a_{3,1}, \dots, a_{3,m_3}\},$$

we can form the set

$$A'_3 = \{a_{1,1}, \dots, a_{1,m_1} a_{2,1}, \dots, a_{2,m_2} a_{3,1} \dots, a_{3,m_3}\}$$

These elements are all in A but we again worry about repetition. Repetition only happens if there is some $a \in A_1 \cap A_3$ or $A_2 \cap A_3$; we took care of the possible repetition coming from A_1, A_2 in our second stage. However, we have $A_1 \cap A_3 = A_2 \cap A_3 = \emptyset$ and so we know that no element is repeated. Continuing to the n th stage, we produce

$$A'_n = \{a_{1,1}, \dots, a_{n,m_n}\} = A.$$

As before, the disjointedness assumption ensures that we have no repetition in this list. Notice that we have $m_1 + m_2 + \dots + m_n$ elements and so we see that our formula

$$\left| \bigcup_{j=1}^n A_j \right| = \sum_{j=1}^n |A_j|$$

provided the sets A_1, \dots, A_n are disjoint.

In applications to counting problems, the set A is perhaps of interest as it may represent some collection of possible events/outcomes/configurations that we would like to know. We view the subsets A_1, \dots, A_n are mutually exclusive special cases of the more general collection A . In fact, we used this approach in understanding all of the possible orderings on a set X . If $X = \{x_1, \dots, x_n\}$, we denote by $\text{Ord}(X)$, the set of all orderings of X . Next, we define $\text{Ord}_{x_j}(X)$ as the set of orderings of X where x_j is first. Notice that

$$\text{Ord}(X) = \bigcup_{j=1}^n \text{Ord}_{x_j}(X)$$

and that

$$\text{Ord}_{x_j}(X) \cap \text{Ord}_{x_k}(X) = \emptyset$$

for $j \neq k$. Therefore, we have

$$|\text{Ord}(X)| = \sum_{j=1}^n |\text{Ord}_{x_j}(X)|.$$

We took an inductive approach and by our induction hypothesis

$$|\text{Ord}_{x_j}(X)| = (n-1)!$$

2.3. LECTURE 6. MORE ON COUNTING

and so

$$|\text{Ord}(X)| = \sum_{j=1}^n (n-1)! = n(n-1)! = n!.$$

When the sets A_j are not disjoint, the problem is much more complicated. We proved that for finite sets $A, B \subset \Omega$ that

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Our counting strategy used in the case of disjoint sets can still be implemented. We again list of the elements of A , say

$$A = \{a_1, \dots, a_m\}$$

and list off the elements of B , say

$$B = \{b_1, \dots, b_n\}.$$

Now, the set

$$C = \{a_1, \dots, a_m, b_1, \dots, b_n\}$$

contains all of the elements of $A \cup B$ but we have repetition for each element in $A \cap B$. In particular, we overcount $A \cup B$ by precisely the size of $A \cap B$ with the set C . Thus,

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

What happens with three sets A, B, C ? We again list them and form the larger list of all of them:

$$A = \{a_1, \dots, a_m\}$$

$$B = \{b_1, \dots, b_n\}$$

$$C = \{c_1, \dots, c_p\}$$

$$D = \{a_1, \dots, a_m, b_1, \dots, b_n, c_1, \dots, c_p\}.$$

We throw out all of the repeats coming from $A \cap B$, $A \cap C$, and $B \cap C$. However, notice that things in $A \cap B \cap C$ get thrown out twice. So we under count by the size of $A \cap B \cap C$. In particular, we see that

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

One way to see that this formula holds is as follows. An element $x \in A \cup B \cup C$ is either an element of 1, 2, or 3 of the sets A, B, C . If x is in exactly one of the sets, then it cannot be in any of the sets $A \cap B$, $A \cap C$, $B \cap C$, or $A \cap B \cap C$. Moreover, it contributes once among the sets A, B, C . If x is in two of the three sets, then x contributes to two of the three of A, B, C , one of the

sets $A \cap B$, $A \cap C$, and $B \cap C$. The element x does not contribute to $A \cap B \cap C$. Since we subtract the cardinalities of the intersections $A \cap B$, $A \cap C$, and $B \cap C$, in our formula, we count x exactly one time. Finally, if x is in all three sets, it contributes to all of the sets

$$A, B, C, A \cap B, A \cap C, B \cap C, A \cap B \cap C.$$

The total contribution in our formula is then

$$1 + 1 + 1 - 1 - 1 - 1 + 1 = 1.$$

Remark: One can work out what happens for n sets and as you might guess, the formula is fairly complicated:

$$\left| \bigcup_{j=1}^n A_j \right| = \sum_{r=1}^n \left(\sum_{1 \leq i_1 < \dots < i_r \leq n} (-1)^{r+1} \left| \bigcap_{j=1}^r A_{i_j} \right| \right).$$

This formula is sometimes referred to as the **exclusion-inclusion principal**. It can be proven using the same logic we used in the case of three sets. If x is in exactly s of the sets A_1, \dots, A_n , it contributes $C(s, 1)$ in the sum $r = 1$, $-C(s, 2)$ times in the sum when $r = 2$, and so forth. In particular, the total contribution is

$$C(s, 1) - C(s, 2) + \dots + (-1)^{s+1} C(s, s)$$

However, we know that

$$\sum_{r=0}^s (-1)^r C(s, r) = 0$$

by the Binomial Theorem. Since $C(s, 0) = 1$, we see that

$$C(s, 1) - C(s, 2) + \dots + (-1)^{s+1} C(s, s) = 1.$$

Something we did not consider in the previous lecture explicitly is counting related matters for finite sets via functions. Say we have a finite sets A, B and a function $f: A \rightarrow B$. For each $b \in B$, we have the preimage $f^{-1}(b) \subset A$. Notice that if $b \neq b'$, then

$$f^{-1}(b) \cap f^{-1}(b') = \emptyset.$$

In particular,

$$|A| = \sum_{b \in B} |f^{-1}(b)|. \quad (2.3)$$

We can use this simple observation to prove a simple but useful result. The basic version of the **Pigeon-Hole Principal** is as follows:

2.3. LECTURE 6. MORE ON COUNTING

Theorem 2.7 (Pigeon-Hole Principal). *If $|A| > |B|$ and $f: A \rightarrow B$ is a function, then there exists $b \in B$ such that $|f^{-1}(b)| > 1$.*

Proof. We prove this result via contradiction. If the result is not true, then there exists a function $f: A \rightarrow B$ such that for all $b \in B$, we have $|f^{-1}(b)| \leq 1$. Note that there are exactly two possibilities for $|f^{-1}(b)|$. It has size either 0 or 1. This assumption in tandem with (2.3), we have

$$|A| = \sum_{b \in B} |f^{-1}(b)| \leq \sum_{b \in B} 1 = |B|.$$

Since, $|A| > |B|$, we arrive at a contradiction and thus conclude that no such f can exist. ♠

In contrast to the Pigeon-Hole Principal, there is **Hilbert's hotel** that you can perhaps draw some intellectual amusement from.

In fact, we can do a little better if we take into account the relationship between the size of A, B .

Theorem 2.8 (Generalized Pigeon-Hole Principal). *If $|A| / |B| > m$ and $f: A \rightarrow B$ is a function, then there exists $b \in B$ such that $|f^{-1}(b)| > m$.*

Proof. We proceed as in the proof of Theorem 2.7. If the result is not true, there exists a function $f: A \rightarrow B$ such that for all $b \in B$, we have $|f^{-1}(b)| \leq m$. Coupled with (2.3), we have

$$|A| = \sum_{b \in B} |f^{-1}(b)| \leq \sum_{b \in B} m = m|B|.$$

Since $|A| / |B| > m$, the above inequality is a contradiction and we deduce that no such function can exist. ♠

There are two (previously deduced) applications of (2.3) that we will discuss. First, if A_1, \dots, A_n are finite sets, we can prove that

$$|A_1 \times A_2 \times \dots \times A_n| = \prod_{j=1}^n |A_j|. \quad (2.4)$$

We established earlier that

$$|A \times B| = |A| \cdot |B|.$$

We prove (2.4) by induction. The base case $n = 1$ is trivially verified since it asks us to demonstrate $|A_1| = |A_1|$. Assume the result for $n - 1$ finite sets, we must verify it holds for n sets. We have

$$P_{A_n} : A_1 \times \cdots \times A_{n-1} \times A_n \longrightarrow A_n$$

defined by

$$P_{A_n}(a_1, \dots, a_{n-1}, a_n) = a_n.$$

By (2.3), we have

$$|A_1 \times \cdots \times A_{n-1} \times A_n| = \sum_{a \in A_n} |P_{A_n}^{-1}(a)|.$$

Now,

$$P_{A_n}^{-1}(a) = \{(a_1, \dots, a_{n-1}, a) : a_j \in A_j\}.$$

In particular,

$$|P_{A_n}^{-1}(a)| = |A_1 \times \cdots \times A_{n-1}| = \prod_{j=1}^{n-1} |A_j|,$$

where the second equality is by our induction hypothesis. Therefore, we have

$$|A_1 \times \cdots \times A_{n-1} \times A_n| = \sum_{a \in A_n} \left(\prod_{j=1}^{n-1} |A_j| \right) = \prod_{j=1}^n |A_j|.$$

The second application is the relationship $P(n, r) = r!C(n, r)$ that we established earlier. If $\text{Ord}_r(X)$ is the set of ordered subsets of X of size r and $\mathcal{P}_r(X)$ is the set of (unordered) subsets of X of size r , we have a function

$$\text{Forget} : \text{Ord}_r(X) \longrightarrow \mathcal{P}_r(X)$$

that sends an order set of size r to the associated unordered subset of size r by "forgetting" the ordering. Thus,

$$|\text{Ord}_r(X)| = \sum_{R \in \mathcal{P}_r(X)} |\text{Forget}^{-1}(R)|.$$

Since the size of $\text{Forget}^{-1}(R)$ is just the number of orderings of R , which is $r!$, we see that

$$|\text{Ord}_r(X)| = \sum_{R \in \mathcal{P}_r(X)} r! = r! |\mathcal{P}_r(X)|.$$

Of course, $P(n, r) = |\text{Ord}_r(X)|$ and $C(n, r) = |\mathcal{P}_r(X)|$.

Yet another application of (2.3) is the following simple result about functions between sets of the same cardinality.

2.3. LECTURE 6. MORE ON COUNTING

Lemma 2.9. *If A, B are finite sets with $|A| = |B|$ and $f: A \rightarrow B$ is a function, then the following statements are equivalent:*

(a) f is bijective.

(b) f is injective.

(c) f is surjective.

Proof. It suffices to prove $(a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (a)$.

Part 1: (a) implies (b). This part is trivial as by definition a bijective function is both injective and surjective.

Part 2: (b) implies (c). We assume that f is injective and must show that f is surjective. Since f is injective, we have that $|f^{-1}(b)| \leq 1$ for all $b \in B$. By (2.3), we have

$$|A| = \sum_{b \in B} |f^{-1}(b)|.$$

If $|f^{-1}(b_0)| = 0$ for some b_0 , the above equality would yield

$$|A| = \sum_{b \in B - \{b_0\}} |f^{-1}(b)| \leq |B - \{b_0\}| = |B| - 1.$$

Since $|A| = |B|$, we see that such an inequality is impossible and so $|f^{-1}(b)| \neq 0$ for any $b \in B$. In particular, f is surjective.

Part 3: (c) implies (a). We assume that f is surjective and must prove that f is bijective. Since a bijective function is both injective and surjective, we see that we only must show that f is injective. By surjectivity, we see that $|f^{-1}(b)| \geq 1$ for all $b \in B$. If f is not injective, then for some $b_0 \in B$, we have $|f^{-1}(b_0)| > 1$. By (2.3), we obtain

$$|A| = \sum_{b \in B} |f^{-1}(b)| > \sum_{b \in B} 1 = |B|.$$

Since $|A| = |B|$, the above inequality is impossible and so f must be injective. ♠

The above proof was organized to show how one typically proves that three or more statements are equivalent. That said, there is a slightly better way to prove Lemma 2.9.

If (b) is equivalent to (c), we see that either (b) or (c) must be equivalent to (a) since (a) is just "(b) and (c)". To see that (b) and (c) are equivalent, we only need to think about (2.3). Indeed, we have

$$|A| = \sum_{b \in B} |f^{-1}(b)|.$$

The equality $|A| = |B|$ with the assumption $|f^{-1}(b)| \geq 1$ for all $b \in B$ implies that $|A| \geq |B|$ with equality if and only if $|f^{-1}(b)| = 1$ for all $b \in B$. Likewise, the assumption $|f^{-1}(b)| \leq 1$ for all $b \in B$ implies that $|A| \leq |B|$ with equality if and only if $|f^{-1}(b)| = 1$ for all $b \in B$.

2.4 Lecture 7. Recursion and Generating Functions

Given a sequence a_n , a basic question is how to describe each entry. One way to do this is through recursion. Specifically, the term a_n is determined by the values of the previous a_i 's. A first example is the linear recursion

Definition 2.1. A linear recurrence of degree k is a sequence which satisfies a relation of the following form.

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \cdots + c_k a_{n-k} + F(n)$$

where the c_i are constants.

The term **linear** refers to the constant c_i and linear in the variables a_i . If $F(n) = 0$, the relation is called **homogeneous**; otherwise, it is called **non homogeneous** and removing the $F(n)$ provides the **associated homogeneous** relation. The distinction is addressed by the following theorem

Theorem 2.10. Suppose $a_n^{(p)}$ satisfies a non homogeneous linear recurrence. Then every solution to this recurrence is of the form

$$a_n = a_n^{(p)} + a_n^{(h)}$$

where $a_n^{(h)}$ is a solution to the associated homogeneous recurrence.

2.4. LECTURE 7. RECURSION AND GENERATING FUNCTIONS

To prove the statement, take the difference between any solution a_n and the given particular solution $a_n^{(p)}$. The $F(n)$ terms cancel and it is straightforward that it satisfies the homogeneous relation.

Given the above theorem, solving a non homogeneous relation requires the acquisition of one particular solution and all the solutions to the associated homogeneous relation. In order to solve such a recurrence, one first supposes that the solution has the form $a_n = r^n$ for some r . If this were the case, then one obtains

$$\begin{aligned}r^n &= c_1 r^{n-1} + c_2 r^{n-2} + \cdots + c_k a_{n-k} \\r^k &= c_1 r^{k-1} + c_2 r^{k-2} + \cdots + c_k\end{aligned}$$

Thus, any r which would solve the recurrence must solve this *associated polynomial*. Since such a polynomial has k roots - not necessarily all distinct - there is a limit to the possible solutions to the recurrence. We will work out that all solutions are essentially a linear combination of roots of this associated polynomial.

Given a sequence a_n , the associated generating function is a function

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

The purpose of a generating function is that relations on the sequence then translate into relations on the associated generating function. Then, using calculus, one hopes to solve the generating function, and make productive statements about the original sequence.

Example. Suppose that a_n is the constant sequence. Then the generating function is the usual geometric sequence

$$A(x) = \sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$$

Suppose we are given the following linear homogeneous recurrence

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \cdots + c_k a_{n-k}$$

Then we may approach this problem using generating functions. Have $A(x)$ denote the associated generating function for a_n . First, multiply across the relation by x^n

$$a_n x^n = c_1 a_{n-1} x^n + c_2 a_{n-2} x^n + \cdots + c_k a_{n-k} x^n$$

This relation holds for all $n \geq k$, so we may sum over all these possible n

$$\sum_{n=k}^{\infty} a_n x^n = \sum_{n=k}^{\infty} c_1 a_{n-1} x^n + \sum_{n=k}^{\infty} c_2 a_{n-2} x^n + \cdots + \sum_{n=k}^{\infty} c_k a_{n-k} x^n$$

We may factor out the extra x factors in all the terms on the right, reindex, and we get

$$\sum_{n=k}^{\infty} a_n x^n = x \sum_{n=k-1}^{\infty} c_1 a_{n-1} x^{n-1} + x^2 \sum_{n=k-2}^{\infty} c_2 a_{n-2} x^{n-2} + \cdots + x^k \sum_{n=0}^{\infty} c_k a_{n-k} x^n$$

Notice, if we add $a_0 + a_1 x + \cdots + a_{k-1} x^{k-1}$ to either side, the left hand side is completed to form $A(x)$. We may repeat the trick for all of the above sums, adding and subtracting the necessary terms to complete each sum to form an $A(x)$ factor. This procedure then leaves us with something of the form

$$A(x) + C(x) = c_1 x A(x) + c_2 x^2 A(x) + \cdots + c_k x^k A(x)$$

where the C is some polynomial in x . It is important to note that depends on a_i , only for $i \leq k$. That is, if we are given an initial condition for a_0 through a_k , $C(x)$ is a polynomial in x and is independent of the general a_n . Subtracting $A(x)$ from either side and dividing, we obtain

$$\frac{C(x)}{-1 + c_1 x + c_2 x^2 + \cdots + c_k x^k}$$

By applying partial fractions, we may reduce the above to terms of the form

$$\frac{B(x)}{(x-r)^j}$$

where r is a root of the associated polynomial. These terms may be written as generating functions via the above example; by equating the n th term in the series on either side of the equation, one obtains a closed form for the recurrence.

We discuss other applications of generating functions. First, if we note that the binomial coefficients can be defined via the generating function

$$(x+1)^n = \sum_{j=0}^n \binom{n}{j} x^j$$

2.4. LECTURE 7. RECURSION AND GENERATING FUNCTIONS

Since $(x+1)^n$ is an infinitely differentiable function for any real n , and the Taylor series gives us a way to create a power series for any such function, we may extend the definition to any real n

$$(x+1)^n = \sum_j \binom{n}{j} x^j$$

A rapid succession of power rule applications yields the following formula

$$\binom{n}{j} = n(n-1)(n-2)\cdots(n-j+1)/j! \quad j \neq 0$$

we define it to be 1 for $j = 0$.

Finally, we consider the following counting problem. Suppose we would like to know the number of ways that three numbers can add to 5, subject to the condition that each number must be between 1 and 3. In order to count this, we consider the product

$$(x^1 + x^2 + x^3)(x^1 + x^2 + x^3)(x^1 + x^2 + x^3)$$

The coefficient of x^5 will be a positive natural number. Specifically, it is all the different ways that a product consisting of a term from each can result in x^5 . Since the exponents each term were selected to match the given restrictions, the coefficient of x^5 precisely the count we are looking for. Thus, all we must do is expand the above product

$$(x^1 + x^2 + x^3)^3 = x^3 + 3x^4 + 6x^5 + 7x^6 + 6x^7 + 3x^8 + x^9$$

Thus, the answer to the question is that there are 6 ways.

A few observations are worth noting. First, this approach is highly flexible: simply adjust the terms with the restrictions given. Second, it gives more information than was requested. In addition to knowing that there are 6 ways to form 5, there are 3 ways to form 4 and 8, 6 ways to form 7 and 7 ways to form 6. Finally, though it may seem like there is no gain, the problem has moved from not knowing where to begin, to only needing a large amount of tedious work to find a solution. Fortunately, there are many free computer resources that will expand a polynomial on command.

Chapter 3

An Excursion into Probability Theory

3.1 Lecture 8. Basic probability theory

In this lecture, we start a new topic (once again). We will investigate some aspects of basic (discrete) probability theory. We start with a finite set S that we view as some space of possibilities/states. S could be the number of possible sequences $\{x_1, x_2, x_3\}$ with 3-terms of elements in X or S could be the set of subsets of X of size r . More concretely, S could be the set of possible lottery ticket choices or the possible hands in a card game like five card draw or Texas Hold'em. An event will be a subset E of S . An event could be the number of 3-term sequences that contains a given element x or the lottery ticket numbers that have a particular number in them like 17. In these examples, each of the possibilities/states in the set S are equally likely to happen though we may wish to consider more complicated situations. For now, we will assume that each possible outcome is equally likely and under that assumption, we define the probability of an event E to be

$$P(E) = \frac{|E|}{|S|}.$$

Example. We can take a coin and flip it seven times, recording whether or not it comes up heads or tails. Our space of outcomes is the possible strings like "HTHHTTH" where H represents heads and T represents tails. It is clear that there are 2^7 different outcomes. What is the probability that the first coin comes up heads and the third comes up tails? These events have "H" first and "T" third. There are 2^5 outcomes of this form and the probability of this

3.1. LECTURE 8. BASIC PROBABILITY THEORY

event is

$$P(E) = \frac{2^5}{2^7} = \frac{1}{4}.$$

What about events that have H three times? It is simple to see that $C(7,3)$ is the number of outcomes with three heads and so the probability of the event is

$$P(E) = \frac{35}{128}.$$

Given an event E , what is the probability that an outcome will not be in E ? That is, what is the probability of $S - E$? Well, we know that

$$|E| + |S - E| = |S|$$

since $E, S - E$ are disjoint sets whose union is all of S . Therefore,

$$\frac{|E|}{|S|} + \frac{|S - E|}{|S|} = 1.$$

In particular,

$$P(E) + P(S - E) = 1.$$

We see that we can write the probability of $S - E$ in terms of the probability of E :

$$P(S - E) = 1 - P(E).$$

Example. What is the probability that we get at least one tails when flipping a coin 7 times? Well, we know that if we don't get at least one tail, then all of the flips yields heads. There is one outcome of this type and so the probability that we get at least one tail is

$$P(E) = 1 - \frac{1}{2^7} = \frac{127}{128}.$$

Given events E_1, E_2 , what is the probability of the event $E_1 \cup E_2$? We know that

$$|E_1 \cup E_2| = |E_1| + |E_2| - |E_1 \cap E_2|.$$

Therefore,

$$P(E_1 \cup E_2) = \frac{|E_1 \cup E_2|}{|S|} = \frac{|E_1|}{|S|} + \frac{|E_2|}{|S|} - \frac{|E_1 \cap E_2|}{|S|} = P(E_1) + P(E_2) - P(E_1 \cap E_2).$$

Example. Monty Hall Paradox. You are a contestant on a game show and there are three doors. Behind one of the doors is a prize while there is nothing behind the remaining two doors. You select one of the doors and then the game show host, who knows which door is the winning door, knowing opens a door with nothing behind it. You are then asked if you want to switch. What should you do? You know that the probability that the door you selected, say door A has the prize is $1/3$ while the probability that the other two doors, say B, C , have the prize is $2/3$. The game show host has revealed to you that one of those other two doors, say C , has nothing behind it. So you should pick the door B since you have a $2/3$ probability of winning with door B while you only have a $1/3$ probability of winning with door A .

Remark. The word paradox is misleading and refers to the somewhat counter-intuitive answer for what the contestant should do. It also is referred to as the Monty Hall Problem or Puzzle.

We now generalize our notion of probabilities on the space S . We will assume that S is finite. For each outcome $x \in S$, we assign to x a probability $P(x) \in [0, 1]$. We also assume that

$$\sum_{x \in S} P(x) = 1.$$

If S is assumed to instead be countable (i.e., the cardinality of the natural numbers \mathbf{N}), the same assumptions can be used. If we write

$$S = \{x_1, x_2, \dots\}$$

then we can view P as a function $P: S \rightarrow [0, 1]$ such that

$$\sum_{x \in S} P(x) = 1.$$

We call P a **probability distribution**. We will call S with the probability distribution P a **probability space**.

Example. Uniform distribution. If $S = \{x_1, \dots, x_n\}$ and

$$P(x_j) = \frac{1}{n}$$

for each $j \in \{1, \dots, n\}$, then it is a simple matter to verify that P is a probability distribution. This probability distribution is called a **uniform distribution**. Note that the first condition that $P(x) \in [0, 1]$ is immediate from our definition of P . The second condition is easily checked:

$$\sum_{j=1}^n P(x_j) = \sum_{j=1}^n \frac{1}{n} = \frac{n}{n} = 1.$$

3.1. LECTURE 8. BASIC PROBABILITY THEORY

Given an event $E \subset S$, we define the probability of the event E by

$$P(E) = \sum_{x \in E} P(x).$$

It is a simple matter to see that we still have the formulas

$$P(S - E) = 1 - P(E)$$

and

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2).$$

Likewise, if $\{E_j\}$ are a countable collection of disjoint events, we have

$$P\left(\bigcup_j E_j\right) = \sum_j P(E_j).$$

Conditional probabilities. Assume that we have a pair of events E, F and that $P(F) > 0$. We may want to know among the outcomes in F , what is the probability that the outcome is also in E . We define the **conditional probability** $P(E | F)$ to be

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

Example. Take the numbers $\{1, \dots, 12\}$. Let F be the event that a number in our set is divisible by 2 and let E be the event that a number is divisible by 3. We see that

$$P(F) = \frac{1}{2}, \quad P(E) = \frac{1}{3}.$$

Now, the event that a number is divisible by both 2, 3 is the same as the event that the number be divisible by 6 and so $P(E \cap F) = 1/6$. Therefore,

$$P(E | F) = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

We see also that

$$F = \{2, 4, 6, 8, 10, 12\}, \quad E = \{3, 6, 9, 12\}, \quad E \cap F = \{6, 12\}.$$

The conditional probability $P(E | F)$ is meant to give the probability an outcome in F is an outcome in E . We see that among the six outcomes in F , there are two that are also in E . Thus, we should have

$$P(E | F) = \frac{2}{6} = \frac{1}{3}.$$

If E is instead the event that a number is divisible by 4, we see that

$$P(E) = \frac{1}{4}$$

while

$$P(E | F) = \frac{1}{2}.$$

In this case, $E \subset F$ and so $P(E \cap F) = P(E)$.

We say that two events E, F are **independent** if $P(E \cap F) = P(E)P(F)$. In our example above, if E is the event that a number be divisible by 3 and F is the event that a number be divisible by 2, we saw that

$$P(E) = \frac{1}{3}, \quad P(F) = \frac{1}{2}, \quad P(E \cap F) = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3}.$$

So these events are independent. On the other hand, if E is the event that a number be divisible by 4, we saw that

$$P(E) = \frac{1}{4}, \quad P(F) = \frac{1}{2}, \quad P(E \cap F) = \frac{1}{4} \neq \frac{1}{8}.$$

In this case, the two events are not independent.

More generally, we say a collection of events $\{E_j\}$ are **mutually independent** if they are pairwise independent and so

$$P\left(\bigcap_j E_j\right) = \prod_j P(E_j).$$

Example. When flipping a coin, each flip is independent from the other flips. Specifically, if the event of getting a head or tail on the j th flip is independent of the the event of getting a head or tail on the i th flip.

An important general example is performing repeatedly an experiment (like flipping a coin) where there are two possible outcomes each time the experiment is performed. If p represents the probability of one of the outcomes and q represents the probability of the other outcome,

we have $p + q = 1$. We see then that if we run n experiments, the probability that the outcome p occurs exactly k times can be described in terms of binomial coefficients:

binomial

$$P(p \text{ happens } k \text{ times}) = C(n, k) p^k q^{n-k}.$$

These types of probability spaces are called **Bernoulli trials** and the associated probability distribution is called a **binomial distribution**. As we saw above, if we flip a coin 7 times and assume that heads/tails are equally likely, then we have

$$P(\text{heads happens exactly 3 times}) = C(7, 3) \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^4 = \frac{C(7, 3)}{2^7}.$$

Function P: S → [0, 1]

A **random variable** on a probability space (S, P) is a **function $X: S \rightarrow \mathbf{R}$** .

Example. If $S = \{1, \dots, 12\}$ with a uniform distribution, we have the random variable X that assigns to each $s \in S$ the remainder when divided by 3. So

$$X(1) = X(4) = X(7) = X(11) = 1, \quad X(2) = X(5) = X(8) = X(11) = 2, \quad X(3) = X(6) = X(9) = X(12) = 0.$$

$X(s) = \alpha$

The **distribution** associated to a random variable is the set of pairs $(\alpha, P(X = \alpha))$, where $P(X = \alpha)$ is the probability that the random variable X equals α . Alternatively,

$$P(X = \alpha) = P(X^{-1}(\alpha)).$$

**$X^{-1}: \mathbf{R} \rightarrow S$
 $P(X^{-1}(\mathbf{R}))$**

We see that

$$\sum_{\alpha} P(X = \alpha) = 1.$$

In our simple example above, we see that

$$P(X = 1) = P(X = 2) = P(X = 3) = \frac{1}{3}.$$

Example. Birthday Paradox. In this example, we will determine the probability that two people share a birthday amongst n people. In our calculation, we will assume that the birthdays among the n people are independent of one another and also that there are 366 days. We will also not consider the year the person is born. Now, to calculate the probability that two people share a birthday among n people, we will compute the probability that n people have n distinct birthdays. Let us call that event $E_{0,n}$ and let S_n denote the probability space of possible outcomes

for the birthdays among n people. Let $E_{1,n}$ denote the event that two people share a birthday among n people. We note that $S_n - E_{1,n} = E_{0,n}$ and so

$$P(E_{1,n}) = 1 - P(E_{0,n}).$$

We know that if n people have distinct birthdays, then there are

$$366 \cdot 365 \cdot \dots \cdot (366 - n + 1)$$

different outcomes with each of the n people having a distinct birthday. Now, the total number of outcomes is $(366)^n$ and so

$$P(E_{0,n}) = \frac{\prod_{j=1}^n (366 - j)}{(366)^n}.$$

Thus,

$$P(E_{1,n}) = 1 - \frac{\prod_{j=1}^n (366 - j)}{(366)^n}.$$

Here is a small table of the probabilities of $E_{1,n}$ for certain values of n :

n	$P(E_{1,n})$
5	0.02714
10	0.1169
15	0.2529
20	0.4114
25	0.5687
30	0.7063
35	0.8144
40	0.8912
45	0.941
50	0.9704
55	0.9863

You can calculate these probabilities [here](#) and can read more about the Birthday Problem [here](#).

It turns out that at $n = 23$, the probability is better than $1/2$.

Remark. Again, the paradoxical nature of the above example is not that it is invalid but that the answer is (possibly) counter-intuitive.

Finally, we end with a simple yet important observation.

Lemma 3.1. *Let (S, P) be a probability space and E an event. If $P(S - E) < 1$, then there exists $x \in E$. That is, there exists an outcome in E .*

Lemma 3.1 is the basis for a method called the **probabilistic method** for proving certain objects/elements exist with certain properties without explicitly constructing one. We will use the probabilistic method later when we do graph theory.

3.2 Lecture 9. More probability theory

We continue our study of discrete probability theory. We start with a basic, useful result called **Bayes' Theorem**.

Theorem 3.2 (Bayes' Theorem). *If (S, P) is a probability space and E, F are events with $P(E), P(F) > 0$, then*

$$P(E | F) = \frac{P(F | E)P(E)}{P(F)}.$$

Proof. Recall that

$$P(E | F) = \frac{P(E \cap F)}{P(F)}, \quad P(F | E) = \frac{P(E \cap F)}{P(E)}.$$

In particular, we see that

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



A useful alternative form of Bayes' Theorem is

$$P(E | F) = \frac{P(F | E)P(E)}{P(F | E)P(E) + P(F | \bar{E})P(\bar{E})} \tag{3.1}$$

where $\bar{E} = S - E$. To obtain (3.1) from Theorem 3.2, we must prove that

$$P(F) = P(F | E)P(E) + P(F | \bar{E})P(\bar{E}). \tag{3.2}$$

To that end, note that

$$(F \cap E) \cup (F \cap \bar{E}) = F$$

and that

$$(F \cap E) \cap (F \cap \bar{E}) = \emptyset.$$

Thus,

$$P(F) = P(F \cap E) + P(F \cap \bar{E}). \quad (3.3)$$

By definition, we have (note that we now need to assume that $P(\bar{E}) > 0$ as well)

$$P(F | E) = \frac{P(F \cap E)}{P(E)}, \quad P(F | \bar{E}) = \frac{P(F \cap \bar{E})}{P(\bar{E})}.$$

Thus,

$$P(F \cap E) = P(F | E)P(E), \quad P(F \cap \bar{E}) = P(F | \bar{E})P(\bar{E}).$$

The above two equalities inserted into (3.3) yield (3.2). With (3.2), we obtain the asserted (3.1).

Remark. There are many fine explicit applications of Bayes' Theorem on-line. The link given before the statement of Theorem 3.2, for instance, has a few nice examples of some useful, real world applications of Bayes' Theorem.

Next, we consider the expected value and variance of a random variable X on a probability space (S, P) . We define the **expected value** of X to be

$$E(X) = \sum_{s \in S} P(s)X(s). \quad \text{def by image } X$$

When S is infinite, we expected value of X is only defined when the series is absolutely convergent. The **deviation of X as s** is defined to be $X(s) - E(X)$.

Alternatively, we can express the expected value via the summation

$$E(X) = \sum_{\alpha} P(X = \alpha)\alpha. \quad \text{def by preimage } X^{-1}$$

That the right hand side equals the expected value is clear since we can decompose S as a disjoint union

$$S = \bigcup_{\alpha} X^{-1}(\alpha).$$

We know that

$$P(X = \alpha) = P(X^{-1}(\alpha)) = \sum_{s \in X^{-1}(\alpha)} P(s)$$

3.2. LECTURE 9. MORE PROBABILITY THEORY

and so

$$P(X = \alpha)\alpha = \sum_{s \in X^{-1}(\alpha)} P(s)\alpha = \sum_{s \in X^{-1}(\alpha)} P(s)X(s).$$

Finally,

$$\sum_{s \in S} P(s)X(s) = \sum_{\alpha} \sum_{s \in X^{-1}(\alpha)} P(s)X(s) = \sum_{\alpha} P(X = \alpha)\alpha.$$

Given two random variables X_1, X_2 and fixed real numbers α, β , we can form a new random variable $\alpha X_1 + \beta X_2$. The expected value of $\alpha X_1 + \beta X_2$ is easily computed from the expected values of X_1 and X_2 . Specifically,

$$E(\alpha X_1 + \beta X_2) = \alpha E(X_1) + \beta E(X_2).$$

We can also form the random variable $X_1 X_2$. However, it need not be the case that $E(X_1 X_2) = E(X_1)E(X_2)$. We say two random variables X_1, X_2 are **independent** if for each $\alpha, \beta \in \mathbf{R}$, we have

$$P(X_1 = \alpha \text{ and } X_2 = \beta) = P(X_1 = \alpha)P(X_2 = \beta).$$

X1, X2: S -> R

Proposition 3.3. *If X_1, X_2 are independent random variables on a probability space (S, P) , then $E(X_1 X_2) = E(X_1)E(X_2)$.*

sense: orthogonal

Proof. This result is fairly easy to prove. We have

$$E(X_1 X_2) = \sum_{\alpha} P(X_1 X_2 = \alpha)\alpha.$$

Now,

$$\sum_{\alpha} P(X_1 X_2 = \alpha)\alpha = \sum_{\alpha_1, \alpha_2} \alpha_1 \alpha_2 P(X_1 = \alpha_1 \text{ and } X_2 = \alpha_2).$$

$$(X_1 X_2)(s) = X_1(s)X_2(s)$$

Using the independence of X_1, X_2 , we have

$$\sum_{\alpha_1, \alpha_2} \alpha_1 \alpha_2 P(X_1 = \alpha_1 \text{ and } X_2 = \alpha_2) = \sum_{\alpha_1, \alpha_2} \alpha_1 P(X_1 = \alpha_1) \alpha_2 P(X_2 = \alpha_2).$$

We can express the right hand side of the above equality as a double sum:

$$\sum_{\alpha_1, \alpha_2} \alpha_1 P(X_1 = \alpha_1) \alpha_2 P(X_2 = \alpha_2) = \sum_{\alpha_1} \left(\alpha_1 P(X_1 = \alpha_1) \sum_{\alpha_2} \alpha_2 P(X_2 = \alpha_2) \right).$$

Now,

$$\alpha_1 P(X_1 = \alpha_1) \sum_{\alpha_2} \alpha_2 P(X_2 = \alpha_2) = \alpha_1 P(X_1 = \alpha_1) E(X_2).$$

Thus,

$$\begin{aligned} \sum_{\alpha_1} \left(\alpha_1 P(X_1 = \alpha_1) \sum_{\alpha_2} \alpha_2 P(X_2 = \alpha_2) \right) &= \sum_{\alpha_1} \alpha_1 P(X_1 = \alpha_1) E(X_2) \\ &= E(X_2) \sum_{\alpha_1} \alpha_1 P(X_1 = \alpha_1) = E(X_1) E(X_2). \end{aligned}$$



The **variance** of a random variable X is defined to be

$$V(X) = \sum_{s \in S} [X(s) - E(X)]^2 P(s).$$

The **standard deviation** of X is defined to be the square-root of the variance. That is,

$$\sigma(X) = \sqrt{V(X)}.$$

Lemma 3.4. *If X is a random variable on a probability space, then* $V(X) = E(X^2) - E(X)^2$.

Proof. We have

$$\begin{aligned} V(X) &= \sum_{s \in S} (X(s) - E(X))^2 P(s) \\ &= \sum_{s \in S} [X^2(s)P(s) - 2X(s)E(X)P(s) + E(X)^2P(s)] \\ &= \sum_{s \in S} X^2(s)P(s) - 2E(X) \sum_{s \in S} X(s)P(s) + E(X)^2 \sum_{s \in S} P(s) \\ &= E(X^2) - 2E(X)^2 + E(X)^2 = E(X^2) - E(X)^2. \end{aligned}$$



We also have, for $E(X) = \mu$, that $V(X) = E((X - \mu)^2)$. To see this equality, we have (using the linearity of expected values):

$$\begin{aligned} E((X - \mu)^2) &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2. \end{aligned}$$

3.2. LECTURE 9. MORE PROBABILITY THEORY

Note that here we view μ^2 as the constant random variable that assigns to each $s \in S$, the value μ^2 . In particular,

$$E(\mu^2) = \sum_{s \in S} \mu^2 P(s) = \mu^2 \sum_{s \in S} P(s) = \mu^2$$

since

$$\sum_{s \in S} P(s) = 1.$$

Remark. One should really think of integration theory in the above. If you integrate a constant function c over an interval $[a, b]$, we get $\int_a^b c dx = c(b - a)$. Our assumption that S be a probability space is something like assuming our interval has length one or $b - a = 1$. The expected value of a random variable, which is JUST a function on S , then is analogous to $\int_a^b f(x) dx$ where $b - a = 1$.

Theorem 3.5 (Bienaymé's Formula). *If X_1, X_2 are independent, random variables, then*

$$V(X_1 + X_2) = V(X_1) + V(X_2).$$

orthogonal sense in linear algebra

You can find more on this formula [here](#).

Proof. We have:

$$\begin{aligned} V(X_1 + X_2) &= E((X_1 + X_2)^2) - E(X_1 + X_2)^2 \\ &= E(X_1^2) + 2E(X_1 X_2) + E(X_2^2) - (E(X_1) + E(X_2))^2 \\ &= E(X_1^2) + 2E(X_1)E(X_2) + E(X_2^2) - E(X_1)^2 - 2E(X_1)E(X_2) - E(X_2)^2 \\ &= E(X_1^2) - E(X_1)^2 + E(X_2^2) - E(X_2)^2 = V(X_1) + V(X_2). \end{aligned}$$



Next, we have a **basic inequality** often called **Chebyshev's Inequality**.

Theorem 3.6 (Chebyshev's Inequality). *If X is a random variable, then*

$$P(|X(s) - E(X)| \geq r) \leq \frac{V(X)}{r^2}.$$

隨機變量的「幾乎所有」值都會「接近」平均

Proof. Let A_r denote the set of $s \in S$ such that $|X(s) - E(X)| \geq r$ and note that the inequality we seek is simply

$$V(X) \geq r^2 P(A_r).$$

The variance of X is given by

$$V(X) = \sum_{s \in S} (X(s) - E(X))^2 P(s) = \sum_{s \in A_r} (X(s) - E(x))^2 P(s) + \sum_{s \notin A_r} (X(s) - E(X))^2 P(s).$$

Now,

$$\sum_{s \notin A_r} (X(s) - E(X))^2 P(s) \geq 0$$

and so

$$\sum_{s \in A_r} (X(s) - E(x))^2 P(s) + \sum_{s \notin A_r} (X(s) - E(X))^2 P(s) \geq \sum_{s \in A_r} (X(s) - E(x))^2 P(s).$$

Also,

$$\sum_{s \in A_r} (X(s) - E(X))^2 P(s) \geq \sum_{s \in A_r} r^2 P(s) = r^2 P(A_r).$$

In particular

$$V(X) \geq r^2 P(A_r)$$

as desired. ♠

3.2. LECTURE 9. MORE PROBABILITY THEORY

Chapter 4

An Excursion into Algebra

We now start a new module that will focus on more algebraic matters. We start with a review of arithmetic.

4.1 Lecture 10. Arithmetic

Recall that \mathbf{Z} is the set of integers. We have two basic binary operations on the integers, namely addition and multiplication. We denote addition by $a + b$ and multiplication by ab for a pair of integers a, b . The natural numbers \mathbf{N} sit inside of \mathbf{Z} and are closed under both addition and multiplication.

Addition, in some sense, is the easier of the two operations to understand. In addition, the number 0 plays a special role. It is the unique integer a such that $a + b = b$ for all $b \in \mathbf{Z}$. The element 0 is sometimes referred to as the additive identity. For each integer $a \in \mathbf{Z}$, there is a unique $b \in \mathbf{Z}$ such that $a + b = 0$. We usually write b as $-a$ and call this number the **additive inverse** of a . In multiplication, both 0 and 1 have special roles. For 1, it is the unique number $a \in \mathbf{Z}$ such that $ab = b$ for all $b \in \mathbf{Z}$. We say that **1 is the multiplicative identity**. However, for most numbers $a \in \mathbf{Z}$, there is no $b \in \mathbf{Z}$ such that $ab = 1$. In fact, only 1 and -1 have this special property.

A number $a \in \mathbf{Z}$ is called **prime** if whenever $a = a_1 a_2$, then either $a_1 = \pm 1$ or $a_2 = \pm 1$. Some

basic facts:

每个大于1的自然数均可写为質數的积

Fundamental Theorem of Arithmetic. Every integer $m \in \mathbf{Z}$ can be expressed uniquely (up to permutations of the primes) as a products of finitely many primes p_1, \dots, p_{s_m} to positive integer powers:

$$m = \pm 1 \prod_{j=1}^{s_m} p_j^{\alpha_j}.$$

Infinitude of the Primes. There are infinitely many prime integers.

The Division Algorithm. Given to integers $m, n \in \mathbf{Z}$, there exist unique integers q, r such that $m = qn + r$ where $0 \leq r < |n|$. We say r the **residue** of m under division by n . We could also say that n goes into m q -times with remainder r .

Using the division algorithm, for each fixed n , we can define a collection of subsets of \mathbf{Z} :

$$\bar{r} = \{m \in \mathbf{Z} : m = qn + r\}.$$

We call the sets $\bar{0}, \bar{1}, \dots, \overline{n-1}$ the **residue classes** of n or the modulo n residue classes. In practice, in writing \bar{a} , we mean $\overline{r_a}$, where r_a is the residue of a modulo n . One can think of the representatives r_a for the residue class \bar{a} as something like a reduced form. For instance, if $n = 5$, we have the residue classes $\bar{0}, \bar{1}, \bar{2}, \bar{3}, \bar{4}$ and these representatives are in what one might call reduced form. The class $\bar{13}$ is the same as the class $\bar{3}$ or $\overline{-2}$. For the number -2 , we see that if $-2 = 5k + r$ where $r \in \{0, 1, 2, 3, 4\}$, we see that $k = -1$ and $r = 3$.

It turns out, we can define **binary operations that play the role of addition and multiplication**. Let

$$\mathbf{Z}/n\mathbf{Z} = \{\bar{0}, \bar{1}, \dots, \overline{n-1}\} \quad \text{quotient group dfn}$$

denote the set of residue classes modulo n . We define addition and multiplication by

$$\bar{a} + \bar{b} = \overline{a+b}, \quad \bar{a}\bar{b} = \overline{ab}.$$

Let us make sense of these operations with a few general calculations. First, we know that the elements of \bar{a} are of the form

$$a' = qn + r_a,$$

where $r_a \in \{0, 1, \dots, n-1\}$ is the residue of a modulo n . Similarly, the elements in \bar{b} are of the form

$$b' = q'n + r_b,$$

where $r_b \in \{0, 1, \dots, n-1\}$ is the residue of b modulo n . Note that in both of the above forms, r_a, r_b, n are fixed and q, q' vary over all of \mathbf{Z} . We see that

$$a' + b' = n(q + q') + r_a + r_b.$$

Since the r_a, r_b are fixed, the numbers $a' + b'$ will always have the same residue modulo n , namely $r_a + r_b$. In particular, the reduced form is either $\overline{r_a + r_b}$, if $r_a + r_b < n$ or $\overline{n - r_a - r_b}$, if $r_a + r_b \geq n$. For multiplication, we have

$$a'b' = (qn + r_a)(q'n + r_b) = qq'n^2 + qr_bn + q'r_an + r_ar_b.$$

Everything but r_ar_b has a factor of n and so $a'b'$ has the same residue modulo n for any pair of choices a', b' . Below are the addition and multiplication tables for $n = 5, n = 6$:

First, we have the addition table for the modulo 5 residue classes:

$+, 5$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$
$\overline{0}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$
$\overline{1}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{0}$
$\overline{2}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{0}$	$\overline{1}$
$\overline{3}$	$\overline{3}$	$\overline{4}$	$\overline{0}$	$\overline{1}$	$\overline{2}$
$\overline{4}$	$\overline{4}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$

Notice how symmetric the table is. Next, we have the modulo 5 residue class multiplication table:

$\cdot, 5$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$
$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$
$\overline{2}$	$\overline{0}$	$\overline{2}$	$\overline{4}$	$\overline{1}$	$\overline{3}$
$\overline{3}$	$\overline{0}$	$\overline{3}$	$\overline{1}$	$\overline{4}$	$\overline{2}$
$\overline{4}$	$\overline{0}$	$\overline{4}$	$\overline{3}$	$\overline{2}$	$\overline{1}$

Ignore the row and column of zeroes, there is a fairly symmetric four by four square. Next, we have the addition table for the modulo 6 residue classes:

4.1. LECTURE 10. ARITHMETIC

$+, 6$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$
$\bar{0}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$
$\bar{1}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$	$\bar{1}$
$\bar{2}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$	$\bar{0}$	$\bar{1}$
$\bar{3}$	$\bar{3}$	$\bar{4}$	$\bar{5}$	$\bar{0}$	$\bar{1}$	$\bar{2}$
$\bar{4}$	$\bar{4}$	$\bar{5}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$
$\bar{5}$	$\bar{5}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$

The table is quite symmetric as in the addition table for modulo 5 addition. Finally, we have the multiplication table for the modulo 6 residue classes:

$\cdot, 6$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$
$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$
$\bar{1}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$
$\bar{2}$	$\bar{0}$	$\bar{2}$	$\bar{4}$	$\bar{0}$	$\bar{2}$	$\bar{4}$
$\bar{3}$	$\bar{0}$	$\bar{3}$	$\bar{0}$	$\bar{3}$	$\bar{0}$	$\bar{3}$
$\bar{4}$	$\bar{0}$	$\bar{4}$	$\bar{2}$	$\bar{0}$	$\bar{4}$	$\bar{2}$
$\bar{5}$	$\bar{0}$	$\bar{5}$	$\bar{4}$	$\bar{3}$	$\bar{2}$	$\bar{1}$

As before, ignoring the column and row of zeroes, there is the part of the table that is encoding multiplication. Unlike the very symmetric table for 5, the table for 6 has oddities not found in any of the above tables. There are rows and columns with numbers/classes that appear more than once. We see that $2 \cdot 3$ is 0 in modulo 6 multiplication. Same with $4 \cdot 3$.

The modulo n residue classes $\mathbf{Z}/n\mathbf{Z}$ with the addition operation form what is called an **abelian group** or **commutative group**. We have the following properties that in total give the definition of an abelian group:

- (a) There is an element $\bar{0} \in \mathbf{Z}/n\mathbf{Z}$ such that for each $a \in \mathbf{Z}/n\mathbf{Z}$, we have $a + \bar{0} = a$.

- (b) For each $a \in \mathbf{Z}/n\mathbf{Z}$, there is a $b \in \mathbf{Z}/n\mathbf{Z}$ such that $a + b = \bar{0}$.

- (c) For each $a, b \in \mathbf{Z}/n\mathbf{Z}$, we have

$$a + b = b + a.$$

(d) For each $a, b, c \in \mathbf{Z}/n\mathbf{Z}$, we have

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

The modulo n residue classes also have a multiplication operation that satisfies the following properties:

(e) There exists $\bar{1}$ in $\mathbf{Z}/n\mathbf{Z}$ such that for all a in $\mathbf{Z}/n\mathbf{Z}$, we have

$$a \cdot \bar{1} = a.$$

(f) For each $a, b \in \mathbf{Z}/n\mathbf{Z}$, we have

$$ab = ba.$$

(g) For each $a, b, c \in \mathbf{Z}/n\mathbf{Z}$, we have

$$(ab)c = a(bc).$$

(h) For each $a, b, c \in \mathbf{Z}/n\mathbf{Z}$, we have

$$a(b + c) = ab + ac.$$

The whole package $(\mathbf{Z}/n\mathbf{Z}, +, \cdot)$ gives $\mathbf{Z}/n\mathbf{Z}$ the structure of a **commutative ring with identity**. Identity refers to the existence of the element $\bar{1}$. Commutative refers to the multiplication operation being commutative as the additive operation in a general ring is always commutative. The integers $(\mathbf{Z}, +, \cdot)$ are also a commutative ring with identity. So are the rationals, reals, and complex numbers. In fact, those rings satisfy another property:

(i) For each non-zero $a \in \mathbf{R}$, there exists $b \in \mathbf{R}$ such that $ab = 1$.

When a commutative ring $(R, +, \cdot)$ satisfies the condition (i), then we say R is a **field**. The integers are not a field since a number like 2 or 17 does not have a multiplicative inverse. In fact, only ± 1 do as we noted above.

In our above examples, we can see that $\mathbf{Z}/5\mathbf{Z}$ is a field. However, $\mathbf{Z}/6\mathbf{Z}$ is not since 2, 3, 4 do not have multiplicative inverses.

4.2 Lecture 11. More on congruence relations

Given a pair of integers $a, b \in \mathbf{Z}$, we define the **greatest common divisor** to be

$$\text{GCD}(a, b) = \max \{m \in \mathbf{N} : m \mid a, m \mid b\}$$

where $m \mid a$ (respectively, $m \mid b$) means that m divides a (respectively, m divides b). Note that $\text{GCD}(a, b)$ exists since the set of such m is bounded. We say that $a, b \in \mathbf{Z}$ are **relatively prime** if $\text{GCD}(a, b) = 1$. For a pair of integers $a, b \in \mathbf{Z}$, define the **least common multiple** of a, b to be

$$\text{LCM}(a, b) = \min \{m \in \mathbf{N} : a \mid m, b \mid m\}.$$

Since $a, b \mid ab$, we know that $\text{LCM}(a, b) \leq |ab|$. We leave it as an exercise to prove that

$$|ab| = \text{GCD}(a, b) \text{LCM}(a, b). \quad ?$$

We saw above that $\mathbf{Z}/n\mathbf{Z}$ is in general not a field. This issue presents some difficulties in performing basic algebraic operations. For instance, given $a, b, c \in \mathbf{Z}/n\mathbf{Z}$, if $ac = bc \pmod n$, when can we conclude that $a = b \pmod c$? Let us first note that it is not always true. If we take $\bar{1}, \bar{2}, \bar{3} \in \mathbf{Z}/4\mathbf{Z}$, note that

$$\bar{1} \neq \bar{3} \pmod 4$$

but

$$\bar{2} = \bar{2} \cdot \bar{1} = \bar{2} \cdot \bar{3} = \bar{2} \pmod 4.$$

Notice that if we have $ac = bc \pmod n$ and there exists $d \in \mathbf{Z}/n\mathbf{Z}$ such that $cd = 1 \pmod n$, then we would have

$$(ac)d = (bc)d \pmod n$$

and by associativity, we have

$$a(cd) = b(cd) \pmod n.$$

Therefore, we have $a = b \pmod n$. Similarly, if there exists $d \in \mathbf{Z}/n\mathbf{Z}$ such that $d \neq 0$ but $cd = 0 \pmod n$, then we can take $a = d$ and $b = 0$. For that pair, we have $a \neq b \pmod n$ but

$$ac = bc \pmod n.$$

Indeed, we have $cd = 0 \pmod n$ but $c, d \neq 0 \pmod n$.

The primary arithmetic issue that we have is the following. If we are given $ac = bc \pmod n$, then we know that $n \mid ac - bc$. In particular, $n \mid c(a - b)$. However, unless n is prime, there is no reason that $n \mid (a - b)$ or $n \mid c$.

Lemma 4.1. Let $a, b, c \in \mathbf{Z}$, $n \in \mathbf{N}$ and set $d = \text{GCD}(c, n)$, $\ell = n/d$. If $ac = bc \pmod n$, then $a = b \pmod \ell$.

Proof. By definition, we have that n divides $c(a - b)$. In particular, there exists $q \in \mathbf{Z}$ such that $c(a - b) = qn$. Now, dividing both sides by $d = \text{GCD}(c, n)$, we have $c'(a - b) = q\ell$, where

$$c' = \frac{c}{d}.$$

By Exercise 2 on Pset 5, we know that c', ℓ are relatively prime. Since c', ℓ are relatively prime, it follows then that ℓ divides $a - b$ and so $a = b \pmod \ell$. ♠

Corollary 4.2. If $a, b, c, n \in \mathbf{Z}$, $\text{GCD}(c, n) = 1$, and $ac = bc \pmod n$, then $a = b \pmod n$.

We define an important function called the **Euler ϕ -function**. Given a positive integer $n \in \mathbf{N}$, we set

$$\phi(n) = |\{m \in \mathbf{N} : \text{GCD}(m, n) = 1, m \leq n\}|.$$

That is, $\phi(n)$ is the number of positive integers that are less than n and also are relatively prime to n .

Examples: The following table gives some values of $\phi(n)$ for some small integers:

n	1	2	3	4	5	6	7	8	9	10	11
$\phi(n)$	1	1	2	2	4	2	6	4	6	4	10

For instance, if $n = 5$, we have

$$\{1, 2, 3, 4\}$$

and for $n = 6$, we have

$$\{1, 5\}.$$

In fact, for all of the primes p in our table, we have $\phi(p) = p - 1$. That holds in general.

Lemma 4.3. If p is a prime, then

$$\phi(p) = p - 1.$$

Proof. Since p is prime, for each $1 \leq m < p$ that $\text{GCD}(m, p) = 1$. Thus, $\phi(p) = p - 1$. ♠

4.2. LECTURE 11. MORE ON CONGRUENCE RELATIONS

The function ϕ appears to be fairly complicated. To push Lemma 4.3 a bit further, we investigate $\phi(p^\alpha)$ for $\alpha \geq 1$. For simplicity, we start with $\alpha = 2$. Now, p^2 will be relatively prime with any number that is not divisible by p . We know that the numbers that are divisible by p that are between 1 and p^2 are the numbers

$$p < 2p < 3p < \dots (p-2)p < (p-1)p < p^2.$$

We see that there are p of these numbers and so

$$\phi(p^2) = p^2 - p = p(p-1).$$

For a general power $\alpha > 2$, we know that p^α is relatively prime with all of the numbers that are relatively prime to p . We can again list off all of the numbers that p divides that are between 1 and p^α :

$$\begin{aligned} p < 2p < \dots < (p-1)p < p^2 < (p+1)p < \dots \\ < (p^2-1)p < p^2p < \dots < (p^{\alpha-1}-1)p < p^{\alpha-1}p = p^\alpha. \end{aligned}$$

Specifically, this list is all of the numbers kp where $k \in \{1, \dots, p^{\alpha-1}\}$. Therefore, we see that

$$\phi(p^\alpha) = p^\alpha - p^{\alpha-1} = p^{\alpha-1}(p-1).$$

Remark: Let $S = \{1, \dots, p^\alpha\}$ and let P be the uniform probability distribution on S that takes value $P(x) = 1/p^\alpha$ for each $x \in S$. If E is the event that $p \mid x$ for $x \in S$, we see that

$$P(E) = \frac{1}{p}.$$

Now, we know that

$$|E| = p^\alpha - \phi(p^\alpha)$$

and

$$P(E) = \frac{|E|}{p^\alpha}.$$

Combining all of that yields:

$$P(E) = \frac{1}{p} = \frac{|E|}{p^\alpha} = \frac{p^\alpha - \phi(p^\alpha)}{p^\alpha} = 1 - \frac{\phi(p^\alpha)}{p^\alpha}.$$

Therefore

$$\frac{1}{p} = 1 - \frac{\phi(p^\alpha)}{p^\alpha}.$$

Solving for $\phi(p^\alpha)$, we obtain

$$\phi(p^\alpha) = p^\alpha - p^{\alpha-1}.$$

With the above probabilistic view, we immediately see too that

$$\phi(p^\alpha) = p^\alpha \left(1 - \frac{1}{p}\right) = |S| P(S - E).$$

We now return to the computation of $\phi(n)$. There are a number of ways of deriving a formula for $\phi(n)$. We will use a probabilistic approach similar to the one used in the remark above. By the Fundamental Theorem of Arithmetic, we can factor n into a product of primes to powers:

$$n = \prod_{j=1}^r p_j^{\alpha_j}. \quad (4.1)$$

Now, we want to count the number of numbers $m \leq n$ that are relatively prime to n . By Exercise 6 on Pset 5, we see that $\text{GCD}(n, m) = 1$ if and only if $\text{GCD}(p_j, m) = 1$ for all $1 \leq j \leq r$. Let $S = \{1, \dots, n\}$, P be the uniform probability distribution, and E_j be the event that $p_j \mid x$. We see that

$$P(E_j) = \frac{1}{p_j}$$

and that the events $\{E_j\}$ are pairwise independent. From above, setting

$$E = \bigcup_{j=1}^r E_j,$$

we know that

$$\phi(n) = |S| P(S - E).$$

Now, by de Morgan's Law, we have

$$P(S - E) = P\left(\bigcap_{j=1}^r (S - E_j)\right).$$

Since the events are pairwise independent, we have

$$P\left(\bigcap_{j=1}^r (S - E_j)\right) = \prod_{j=1}^r \left(1 - \frac{1}{p_j}\right).$$

Thus,

$$\phi(n) = n \cdot \left(\prod_{j=1}^r \left(1 - \frac{1}{p_j}\right)\right). \quad (4.2)$$

As lovely as the above formula is, there is an **equally nice alternative**. For that, employing (4.1) in (4.2), we obtain

$$\begin{aligned} \phi(n) &= \left(\prod_{j=1}^r p_j^{\alpha_j}\right) \left(\prod_{j=1}^r \left(1 - \frac{1}{p_j}\right)\right) \\ &= \prod_{j=1}^r p_j^{\alpha_j} \left(1 - \frac{1}{p_j}\right) = \prod_{j=1}^r \phi(p_j^{\alpha_j}). \end{aligned}$$

$24 = 2^3 * 3$
 $\text{fi}(24) = \text{fi}(2^3) * \text{fi}(3)$
 $= 2^2(2-1) * 2$

Hence, in total:

in exam will be 100 ~ 1000

Exam: quick way!

$$\phi(n) = \prod_{j=1}^r \phi(p_j^{\alpha_j}). \quad (4.3)$$

Now, we would like to use the function $\phi(n)$ for a cool application. Let G_n be the subset of $\mathbf{Z}/n\mathbf{Z}$ of classes that are relatively prime with n . We know that $|G_n| = \phi(n)$. The set G_n is closed under multiplication and contains the identity element 1. For any element $a \in G_n$, we can consider the sequence a^j in $\mathbf{Z}/n\mathbf{Z}$. Since G_n is finite, there is a smallest positive integer $j_0 \leq \phi(n) + 1$ such that $a^{j_0} \in \{a, a^2, \dots, a^{j_0-1}\}$. That says that $a^{j_0} = a^j \pmod n$ for some $1 \leq j < j_0$. In particular,

$$a^j(a^{j_0-j} - 1) = 0 \pmod n.$$

Since a^j is relatively prime to n , this implies that

$$a^{j_0-j} = 1 \pmod n. \quad (4.4)$$

The conditions on the selection of j_0 in tandem with (4.4) force $j = 1$. In particular, $a^{j_0} = a \pmod n$ and $a^{j_0-1} = 1 \pmod n$. We conclude that G_n is an abelian group of order $\phi(n)$. The integer $j_0 - 1$ is the order of the element a in the abelian group G_n . The group G_n is called the **group of units** of the commutative ring $\mathbf{Z}/n\mathbf{Z}$. Since G_n is a group of order $\phi(n)$, it follows that $a^{\phi(n)} = 1$. This result is called the Euler–Fermat Theorem.

$$5^{1011} = 5 \pmod{1009}$$
$$5^{1011} = 5^3 * 5^{1008} = 5^3 \pmod{1009}$$

Theorem 4.4 (Euler–Fermat). *If $\text{GCD}(a, n) = 1$, then $a^{\phi(n)} = 1 \pmod{n}$.*

We also obtain the following special case called Fermat’s Theorem.

Corollary 4.5 (Fermat). *If p is a prime and $\text{GCD}(a, p) = 1$, then $a^{p-1} = 1 \pmod{p}$. In general, for all $a \in \mathbb{Z}$, we have $a^p = a \pmod{p}$.*

Corollary 4.5 follows immediately from Theorem 4.4 and Lemma 4.3.

4.3 Lecture 12. The Chinese Remainder Theorem

4.3. LECTURE 12. THE CHINESE REMAINDER THEOREM

红字是logic flow

Chapter 5

Graph Theory

In our final module, we will studying a class of versatile objects called graphs.

graph -> vertex set, vertex, edge set, edge

5.1 Lecture 13. Definitions and Examples

end points of edge -> loop

Intuitively, a basic graph is a collection of points called vertices and a collection of edges/paths between the vertices. It need not be the case that a given pair of vertices is connected via an edge. Formally, a **graph** $\Gamma(V, E)$ is a pair of sets $V = \{v_j\}$ and $E = \{e_j\}$. The set V is called the **vertex set** and an element $v \in V$ is referred to as **vertex**. The set E is called the **edge set** and an element of E is referred to an **edge**. For each edge e , we have a pair of vertices $v_i, v_j \in V$ that are called the **end points** of e . We sometimes write $e = [v_i, v_j]$. When $v_i = v_j$, we call the associated edge e a **loop**. Note that $[v_j, v_i]$ also represents an edge between the vertices v_i, v_j and so the order does not matter (it will if we orient the edge as we will below).

Remark. A pair of vertices could have several distinct edges between them. The notation $[v, w]$ in that case does not specify which of those edges we are referring to. In general, it is better to write an edge as simply $e = [v, w]$.

Example. Complete Graphs. For any set V , we can form what is called the **complete graph on V** as follows. For any pair of distinct elements $v, v' \in V$, there is a unique edge $e \in E$ with $e = [v, v']$. That is, for every pair of distinct vertices, there is an edge that connects them. When

每两点间有且只有一条线

$|V| = n$, we call the complete graph on V simply the **complete graph on n vertices**. We will denote the complete graph on n vertices by K_n .

Example. Empty graph. For any set V , we can form what is called the **empty graph on V** by declaring $E = \emptyset$. This graph is a collection of vertices but there are no edges. This is maybe not the most interesting graph.

Example. Cycles. Given a finite vertex set $V = \{v_1, \dots, v_n\}$, we can form a graph called an **n -cycle** as follows. For each $i \leq n$, for the vertices v_i, v_{i+1} , we have a **unique** edge $e = [v_i, v_{i+1}]$ where we take the indices modulo n . By "taking the indices modulo n ", we simply mean by this that if $i < n$, then we have an edge $e = [v_i, v_{i+1}]$ and in the case $i = n$, we have an edge $e = [v_n, v_1]$. You can image this graph as a regular n -gon where the vertices are the points of a polygon with n vertices and all the edges are equal. For $n = 3$, this would be an equilateral triangle and for $n = 4$, this would be a square. You can also imagine this as a circle with n points on the circle. We will denote the cycle graph on n vertices by C_n .

Example. Line graph. Given a finite vertex set $V = \{v_1, \dots, v_n\}$, we can form a graph that we will call the **n -point line graph** as follows. For each $i < n$, we have an edge $e = [v_i, v_{i+1}]$. This graph is very close to the cycle graph in the previous example except that we do not have the edge $e = [v_n, v_1]$. This graph looks like a line segment with n points on it and with endpoints v_1, v_n . We will denote the line graph on n vertices by L_n .

Example. Complete Bipartite Graphs. Given **two sets V_1, V_2** , we can form a graph called the associated **complete bipartite graph** as follows. For each $v \in V_1$ and each $w \in V_2$, we have **an edge $e = [v, w]$** . That is, each point on V_1 is connected to each point on V_2 . When $|V_1| = m$ and $|V_2| = n$, we denote the associated complete bipartite graph by $K_{m,n}$.

Example. Bipartite Graphs. More generally, given two sets V_1, V_2 , we can set $V = V_1 \cup V_2$. If $\Gamma(V, E)$ is a graph with vertex set V such that for each edge $e \in E$, we have $e = [v, w]$ where $v \in V_1$ and $w \in V_2$, we will call $\Gamma(V, E)$ a bipartite graph.

Now there are lots of different structures and properties we can study on graphs. Given a graph $\Gamma(V, E)$, a **path** is an ordered set of vertices $\{v_0, \dots, v_r\}$ such that for each $j < r$, there is an edge $e = [v_j, v_{j+1}]$. We refer to v_0 as the **initial point** of the path and v_r as the **terminal point**. Sometimes, **we will simply say that v_0, v_r are the endpoints of the path**. Now, we should really be careful as this may not uniquely specify the path since between v_j, v_{j+1} , there could be multiple edges e of the form $e = [v_j, v_{j+1}]$. We should really think of **a path as an ordered set of edges $\{e_1, \dots, e_r\}$ such that we have $e_i = [v_i, v_{i+1}], e_{i+1} = [v_{i+1}, v_{i+2}]$** .

二分图

connected -> not connected ??? -> path component -> subgraph

We say that a graph $\Gamma(V, E)$ is **connected** if for each pair of vertices $v, w \in V$, there is a path in $\Gamma(V, E)$ with end points v, w . When $\Gamma(V, E)$ is **not connected**, we can define sets

$$\Gamma_v = \{w \in V : \text{there is a path with end points } v, w\}.$$

These sets partition the vertex set in disjoint subsets V_1, \dots, V_r such that for any edge $e \in E$, we have $e = [v, w]$ with $v, w \in V_j$ for some j . If we set E_j to be the set of edges with endpoints in V_j , then $\Gamma(V_j, E_j)$ is a connected graph and is called a **path component** of the graph $\Gamma(V, E)$. These graphs $\Gamma(V_j, E_j)$ are what we call subgraphs of $\Gamma(V, E)$. By a **subgraph** $\Gamma(V', E')$ of a graph $\Gamma(V, E)$, we mean that $V' \subset V$, $E' \subset E$, and for each edge $e \in E'$, the endpoints of e are in V' .

Remark. Often we work with connected graphs $\Gamma(V, E)$ since many problems can be reduced to the path components.

Given a graph $\Gamma(V, E)$, we say that a subgraph $\Gamma(V, E)$ is a **cycle** if $V = \{v_1, \dots, v_n\}$ and the edge set E has a unique edge of the form $e = [v_i, v_{i+1}]$ with indices taken modulo n . In particular, we require that all of the edges be distinct. We say that the cycle has **length** n in this case. We can also think of these cycle subgraphs as closed paths in the graph. They are actually better than that as we insist that our path not backtrack ever. A path that backtracks will traverse a edge more than once. If $\Gamma(V, E)$ is a connected graph without any cycles of length $n \geq 1$, then we call $\Gamma(V, E)$ a **tree**.

If $\Gamma(V, E)$ is a graph such that each path component is a tree, then we call $\Gamma(V, E)$ a **forest**.

Remark. Most of the above examples are not trees. The line graph L_n is a tree but $K_n, K_{m,n}, C_n$ are never trees (except possibly for really small values of m, n).

For each vertex $v \in V$, we set

$$V_v = \{w \in V : \text{there is an edge of the form } [v, w]\}.$$

This is the set of vertices in V that are connected to v by an edge. We say that a graph is **regular** if for each $v, w \in V$, we have $|V_v| = |V_w|$. If $\Gamma(V, E)$ is a regular graph and $k = |V_v|$, we often say that $\Gamma(V, E)$ is **k -regular**. Regular graphs arise, for instance, naturally from some construction in group theory via the Cayley graph $\text{Cayley}(G, S)$, where G is a group and S is a generating set. When S is finite (we say G is finitely generated in this case), the associated graph is a $|S|$ -regular graph. We will return to k -regular graphs a bit later.

Given a subgraph $\Gamma(V', E')$ of a graph $\Gamma(V, E)$, the **complement graph** is defined as follows.

We set the vertex set of the complement graph to be $V_0 = V - V'$ and the edge set E_0 to be all edges with endpoints in V_0 . It is straightforward to see that $\Gamma(V_0, E_0)$ is a subgraph of $\Gamma(V, E)$.

There are various ways we can augment a graph that come up in both theoretical and practical pursuits. We can **orient** a graph $\Gamma(V, E)$ by assigning to each edge $e \in E$ a direction. For each edge $e = [v, w]$, we can think of the edge as either starting at v or starting at w . Once we pick a starting vertex, the other vertex is the terminal vertex and we say that the edge is oriented. If $\Gamma(V, E)$ is a graph where each edge has an orientation, we say that $\Gamma(V, E)$ is an **oriented graph**. In an oriented graph, it is common to write for an edge $e \in E$ the endpoints of the edge as e_-, e_+ where e_- is the starting vertex and e_+ is the terminal vertex. The reason we write e_- is because the edge is leaving that vertex while at e_+ it is coming in (like a sink/source for zeroes of vector fields). In practice, an orientation on a graph is denoted by putting direction arrows on the edges.

Given a graph $\Gamma(V, E)$ and a pair of vertices $v, w \in V$, the **distance** between v, w is defined to minimal length of a path with endpoints v, w . If $\Gamma(V, E)$ is not connected, then there will be pairs of vertices that have infinite distance as the convention is that if no such path exists, we declare the points to be infinitely far apart. We also define the distance between v and itself as 0 since we can take a 0-path (a path that just stays at the vertex v). This notion of distance provides us with a function

$$d_V: V \longrightarrow [0, \infty]$$

that we call the **path distance function** or simply the distance function. You will verify on the problem set that d_V is a metric on the vertex set.

We can assign weights to the edges as in applications it is sometimes useful to think of some edges as longer than others (maybe the edges represent routes between places that do actually have a distance between them that the weight would represent). We say that $\Gamma(V, E)$ is a **weighted graph** if $\Gamma(V, E)$ is a graph and we have a function

$$w: E \longrightarrow (0, \infty).$$

We call $w(e)$ the **weight of e** . For a weighted graph, we can define the distance between two vertices as follows. Given a path c with associated edges e_1, \dots, e_r , we define the **length of c** to be

$$\ell(c) = \sum_{j=1}^r w(e_j).$$

Notice that the length of an edge e_j is simply the weight $w(e_j)$. We then define the distance between a pair of vertices v, w to be the length of the shortest path with end points v, w . When

$\Gamma(V, E)$ is a weighted graph, we will simply call this distance function the associated distance function. In particular, when we have a weighted graph, distances will take into account the weights unless we specify otherwise.

5.2 Lecture 14. (Two lectures) Matchings

Today, we will discuss matching problems that can be modeled by bipartite graphs. There are many specific examples and we will pick one on jobs. We will be given two sets V_1, V_2 where V_1 will represent a collection of employees and V_2 will represent as collection of tasks that need to be done. Each employee $x \in V_1$ will be capable of performing some of the tasks in V_2 . For x , we connect x to a task t by an edge if x is capable of performing t . We get a bipartite graph $\Gamma(V, E)$ where $V = V_1 \cup V_2$ and the edge set E is given as above. A **matching** is a subgraph say $M = \Gamma(V_0, E_0)$ of $\Gamma(V, E)$ such that no two edges share any endpoints. In the context of our example, each employee $x \in V_0$ is assigned exactly one task and no task has more than one person doing it. A **complete matching** if a matching $M = \Gamma(V_0, E_0)$ such that $|V_0| = |V|$. That is, every person is matched with a task.

Hall's Marriage Theorem resolves completely when a bipartite graph $\Gamma(V, E)$ has a complete matching. Before stating Hall's result, we need a general another general concept in graph theory. Given $x \in V$, we set $N(x)$ to be the set of vertices $y \in V$ such that $[x, y] \in E$. These are the vertices that are distance one from x in the metric d we defined last lecture. We sometimes refer to $N(x)$ as the **neighbors of x** . More generally, given $A \subset V$, we define the set

$$N(A) = \{y \in V - A : d(x, y) = 1 \text{ for some } x \in A\}.$$

These are the vertices in V that are connected to A by an edge. Some of the vertices in A could be connected by an edge but we do not count those as neighbors of A since there are in A .

We can now state Hall's theorem.

Theorem 5.1 (Hall's Marriage Theorem). *Let $\Gamma(V, E)$ be a bipartite graph with partitioning vertex sets V_1, V_2 . There is a complete matching $M = \Gamma(V_0, E_0)$ of $\Gamma(V, E)$ if and only if for each $A \subset V_1$, we have $|A| \leq |N(A)|$.*

Before we prove this result, a bit of clarification is needed. By partitioning vertex sets V_1, V_2 , we simply mean that $\Gamma(V, E)$ is a subset of the complete bipartite graph $K_{|V_1|, |V_2|}$. That is, all of

5.2. LECTURE 14. (TWO LECTURES) MATCHINGS

the edges $e \in E$ are of the form $e = [v_1, v_2]$ where $v_1 \in V_1$ and $v_2 \in V_2$. A complete matching in the contexts of general bipartite graphs is defined as follows. A **matching** $M = \Gamma(V_0, E_0)$ is a subgraph such that if $e_1, e_2 \in E_0$ are edges with endpoints v_1, w_1 and v_2, w_2 , then all four points v_1, v_2, w_1, w_2 are distinct. In other words, two distinct edges do not share any endpoints. A **complete matching** is a matching $M = \Gamma(V_0, E_0)$ with $|V_0| = |V|$.

In what follows, we will refer to the hypothesis $|N(A)| \geq |A|$ as the **matching condition**.

Proof of Theorem 5.1. We start with the easy direction of prove that if there is a complete matching, then for any subset $A \subset V_1$, we must have $|A| \leq |N(A)|$. As there is a matching subgraph $M = \Gamma(V_0, E_0)$ with $V_0 = V$, we know that for each $x \in A$, there is an edge e with endpoints $[x, w]$ for $w \in V_2$. In particular, $w \in N(A)$. By the matching condition, if we take another $x' \in A$, the associated edge with have endpoints $[x', w']$ with $w \neq w'$. Thus, in M , we see that A has at least $|A|$ neighbors. As the number of neighbors is a non-decreasing function in passing to the larger graph $\Gamma(V, E)$, we see that $|A| \leq |N(A)|$.

The other direction is the content of the theorem and will be proven by induction on the size of V_1 . If $|V_1| = 1$, then $V_1 = \{x\}$. By assumption we know that there is at least one edge connecting x to V_2 and we can simply pick such an edge for a complete matching. We now assume that for any bipartite graph $\Gamma(V, E)$ with $|V_1| < k$, there exists a complete matching and must prove that the same holds if $|V_1| = k$. We will split into two cases:

Case 1. For every subset $A \subset V_1$, we have $|N(A)| \geq |A| + 1$.

Case 2. There exists a subset $A \subset V_1$ with $|N(A)| = |A|$.

We first treat Case 1. We take $A = V_1$ and select a vertex $x \in V_1$. By assumption, we know that there is an edge connecting x to some vertex say y in V_2 . We remove all the edges from x to V_2 and all the edges from y to V_1 . The result is a new graph with $V'_1 = V_1 - \{x\}$ and $V'_2 = V_2 - \{y\}$. In order to apply our induction hypothesis, we need to know for each set $A' \subset V'_1$ that $|N(A')| \geq |A'|$. In passing from the pair (V_1, V_2) to the pair (V'_1, V'_2) , we eliminated the vertices x, y and all of the edges involving x, y . Viewing A' inside the graph bipartite graph for the pair (V_1, V_2) , by assumption, we have $|N_{V_1, V_2}(A')| \geq |A'| + 1$. If y is not a neighbor of A' in (V_1, V_2) , then $|N_{V'_1, V'_2}(A')| \geq |A'| + 1$. If y is a neighbor of A' in (V_1, V_2) , then we have $|N_{V'_1, V'_2}(A')| \geq |A'|$. In either case we see the matching hypothesis holds and so by induction, we have a complete matching for the bipartite graph associated to (V'_1, V'_2) . To obtain a matching

for the original bipartite graph for (V_1, V_2) , we simply add the matching of x, y .

We next treat Case 2. By assumption, we have a set $A \subset V_1$ such that $|N(A)| = |A|$. For each vertex $x \in A$, there is an edge $e \in E$ with $e = [x, y]$ for $y \in N(A)$ since the matching condition implies $|N(\{x\})| \geq 1$. As $|N(A)| = |A|$, the endpoints y_1, y_2 for distinct vertices $x_1, x_2 \in A$ must be distinct. In particular, we have a matching of the vertices in A with the vertices in $N(A)$. Set $V'_1 = V_1 - A$ and $V'_2 = V_2 - N(A)$. If we can ensure that the bipartite graph associated to (V'_1, V'_2) has a complete matching, then we obtain a complete matching for (V_1, V_2) using the matching we have between $A, N(A)$. By induction, it suffices to verify that the bipartite graph for (V'_1, V'_2) satisfies the matching condition. Given a subset $B \subset V'_1$, we denote again let $N_{V'_1, V'_2}(B)$ be the neighbors of B in the bipartite graph associated to (V'_1, V'_2) . Our present goal is to establish $|N_{V'_1, V'_2}(B)| \geq |B|$. Consider the set $C = A \cup B$ in the bipartite graph (V_1, V_2) . Since $A, N(A)$ are matched already inside the graph (V_1, V_2) , we see that

$$|N(C)| = |N(A)| + N_{V'_1, V'_2}(B).$$

By the matching condition, we have

$$|N(C)| \geq |C| = |A| + |B|.$$

Combining these two observations, we see that

$$|N(C)| = |N(A)| + |N_{V'_1, V'_2}(B)| \geq |A| + |B|.$$

By assumption, $|A| = |N(A)|$ and so

$$|N_{V'_1, V'_2}(B)| \geq |B|.$$

Therefore, the bipartite graph associated to (V'_1, V'_2) satisfies the matching condition. Hence, by induction, we have a complete matching. As noted above, using the matching for A yields a complete matching on the original graph. This completes the proof of Hall's theorem. ♠

Hall's theorem has many equivalent reformulations and also many applications.

Example. We start with an ordinary deck of 52 cards. We have four suits and 13 possible face values for the cards ranging from an ace to a king. Now, I will make 13 piles of 4 cards. For each pile of 4 cards, we will view that as a vertex and the total set of 13 piles will represent

V_1 . For each possible face value, we will have a vertex in a set V_2 . Note that both V_1, V_2 have 13 vertices. For each card in a pile represented by the vertex $x \in V_1$, we will draw an edge to the associated face value vertex in V_2 . In particular, each vertex $x \in V_1$ has 4 edges with endpoints in V_2 . The graph we have is a bipartite graph and one can check that it satisfies the matching condition. By Hall's theorem, there is a complete matching for this bipartite graph. Consequently, we can select a card from each of the 13 piles so that every possible face value occurs and no face value occurs twice.

5.3 Lecture 15. Ramsey Theory

Before launching into the graph theory formalities of Ramsey Theory, we will first discuss some of the philosophies in and around Ramsey Theory. If you take the real plane and tile it with honeycomb tiles, square tiles or triangle tiles for instance, how large of a portion of space can you take without having a full tile? What if you instead split up the natural numbers into two disjoint sets, how much structure do either of the two sets have?

Structure can mean many things and so we will begin to introduce some formalism. We start with some finite set V which we can think of as a collection of people. Any two people $v, w \in V$ can have two possible relationships that we will denote by a color, say, red or blue. If v and w can either have a red relationship or a blue relationship, and we will denote that by an edge between v, w with the color of their relationship. We can allow for a wider variety of relationships, say, some set of colors $\{C_1, \dots, C_r\}$. Returning to our 2-color example, the above graph is a complete graph of $|V|$ vertices and the edges are either red or blue. The graph models real relationships that we may be studying and have some expectation that there should be some structure to the colorings. It could just be random. In this setting, structure is represented by subgraphs with all the same edges either red or blue.

Let K_n be a complete graph on n vertices and E_n the associated edge set. An **r -coloring** of K_n is a function

$$C: E_n \longrightarrow \{C_1, \dots, C_r\},$$

where the set $\{C_1, \dots, C_r\}$ is called the **set of colors**. We say that a subgraph $\Gamma(V, E)$ of K_n is **monochromatic** if $C|_E$ is constant and the color is given by $C(e) = C_j$ for any $e \in E$. In that case, we say that $\Gamma(V, E)$ is monochromatic C_j .

The **Ramsey number** $R(p_1, \dots, p_r)$ is the smallest number $R = R(p_1, \dots, p_r)$ such that if C

is any r -coloring of the complete graph K_R , then there exists a monochromatic C_j complete subgraph on p_j vertices for some $1 \leq j \leq r$.

It does not follow immediately that Ramsey numbers are even finite. Before establish finiteness of Ramsey numbers and some estimates, we interpret the philosophy of what the finiteness of the Ramsey numbers means. What it says that if the number of people is sufficiently large, then no matter what collection of r relationships I want to study, I know that there will be a group of p_j people that all have the same relationships. We view those groups as structure on some scale p_j .

If we return to the tiling of the plane, we could take triangles as our tiles. For each edge of a tile, we could think of taking that part of the tiling as a blue edge and when we don't take that part of the tiling, we can think of it as a red edge. In this way our tiles get some coloring of their edges. Ramsey numbers tell us that if we take too much of the tiling, we are forced to have full tile while also if we take too little, we will leave a full tile behind. Specifically, it says that either you are take a full tile or you leave a full tile. The same holds for squares, and other shapes.

Now back to the Ramsey numbers. It is easy to see that

$$R(p'_1, \dots, p'_r) \leq R(p_1, \dots, p_r)$$

when

$$p'_j \leq p_j$$

for all j . One can prove inductively on the sum

$$\sum_{j=1}^r p_j = N$$

to prove that $R(p_1, \dots, p_r)$ is finite. In the proof, one gets upper bounds like

$$R(p, q) \leq R(p-1, q) + R(p, q-1).$$

These can be used to get bounds recursively.

Very few Ramsey numbers have been computed explicitly. For instance, $R(3, 3) = 6$ which can be shown directly using upper bounds and explicit constructions. The number $R(4, 4) = 18$ is computed via a similar strategy. However, it is only known that

$$43 \leq R(5, 5) \leq 55$$

5.3. LECTURE 15. RAMSEY THEORY

and

$$102 \leq R(6, 6) \leq 178.$$

Also, $R(3, 3, 3) = 17$ is the only non-trivial Ramsey number to be explicitly computed for more than two colors.

One can give a lower bound for Ramsey numbers using what is now called the **probabilistic method**. To prove get a lower bound on $R(k, k)$ for instance, you pick a number n satisfying the question

$$\binom{n}{k} 2^{1 - \binom{k}{2}} < 1.$$

You then argue that if you randomly 2-color K_n , with positive probability, there are is neither a monochromatic red or blue complete subgraph K_k . That works out to the explicit estimate

$$k2^{k/2} \leq R(k, k).$$

In the probabilistic method, you take the sample S of colorings of K_n equipped with a uniform probability distribution. The problem is then just one where we have to count the number of colorings with a monochromatic red or blue complete subgraph K_k and compare it to the size of the total space. By the selection of n , the number of ones with monochromatic subgraphs is strictly smaller. Hence, there must exist a coloring without any monochromatic complete subgraphs K_k and thus $n < R(k, k)$.