Discrete Mathematics

Lecture Notes

on

Sets and Functions

M. Andrew Moshier

October 2014

Overview

The mathematical universe consists of various things: numbers, functions, graphs, lists and so on. A *set* is collection of things. For example, the collection of all natural numbers is a set. A *function* is a correlation of the members of one set with members of another set. These two abstract concepts (sets and functions) form a conceptual framework in which virtually all of mathematics can be built. So an understanding of sets and functions is key to a rigorous approach to most other parts of mathematics. This conceptual framework can itself be put on a formal, precise footing called the Category of Sets and Functions. In these lectures, we investigate the Category of Sets and Functions, so that we can use these things as the basic building blocks of everything else we do.

Version: 0:0.0 (December 16, 2014) Format: print

Contents

Part I

Natural Numbers and Induction

1

The Natural Numbers

The *natural numbers* have to do with counting: $0, 1, 2, 3, \ldots$. They do not include negatives or fractions or irrationals. In this lecture, the structure of natural numbers is the topic. To hone in on that structure, we look at structures similar to the natural numbers, but that fail to capture some basic aspects of counting. Bogus structures are ruled out by axioms that distinguish the structure of natural numbers from all others.

Goals

Lecture

- Present the natural numbers as comprising a structure suited to counting.
- Identify similar structures that can not properly represent counting.
- Rule out "bad" structures via axioms.

Study

- Gain facility in the course's *successor* notation, including translating between successor notation and base 10 notation.
- Commit to memory the axioms of natural numbers.
- Demonstrate ability to recognize failures of the axioms.

1.1 The Basic Picture

Natural numbers are pictured like stepping stones in Figure 1.1.

$$0 \longrightarrow 1 \longrightarrow 2 \longrightarrow 3 \longrightarrow \cdots$$

Figure 1.1: A picture of the natural numbers

Figures 1.2, 1.3 and 1.4 illustrate three ways *not* to picture the natural numbers.
These incorrect pictures can be ruled out by explaining the basic structure of counting.

Basic Vocabulary 1.1

Nat The natural numbers have the following basic structure.

- There is a special natural number. We denote this by 0.
- For any natural number n, there is a unique *next* natural number. We call this the *successor of* n. In these lectures, we denote the successor of n by $n^{\frown}$.

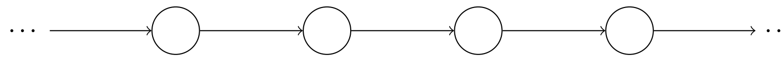


Figure 1.2: Nowhere to start

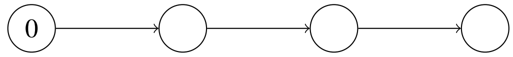


Figure 1.3: Nowhere to go

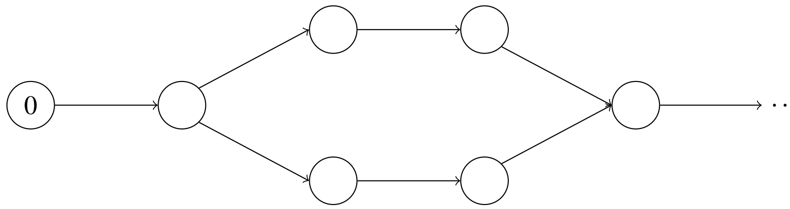


Figure 1.4: Forks in the path

Exercises for Lecture 1

Convert the following to successor notation.

1. 9
2. 10
3. $4+3$
4. $n+4$

Convert the following to base 10 notation.

1. $0^{\frown\frown\frown\frown}$
2. $n^{\frown\frown\frown\frown\frown}$
3. $5^{\frown\frown}$
4. $0^{\frown}+0^{\frown\frown}$

1.2 Narrowing the possibilities

Figures 1.5 and 1.6 illustrate problems that **??** does not avoid.

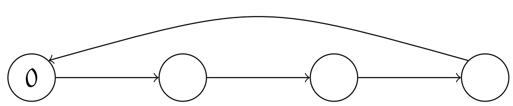


Figure 1.5: A strange way to count

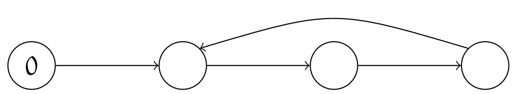


Figure 1.6: Another strange way to count

Exercises for Lecture 1

Explain, in one or two sentences each, why Figures 1.5 and 1.6 satisfy Axiom **??**.

Figure 1.5 is flawed because 0 has a *predecessor*: a value n satisfying $0^{\curvearrowright\curvearrowright\curvearrowright\curvearrowright} = 0$. Figure 1.6 is flawed because an element has two distinct predecessors: $0^{\curvearrowright} = 0^{\curvearrowright\curvearrowright\curvearrowright\curvearrowright}$. We can insist that these flaws do not happen in the natural numbers. That is, we rule them out with axioms.

Axiom 1.1

For every natural number n, $n^{\curvearrowright} \neq 0$.

Axiom 1.2

For any natural numbers m and n, if $m^{\curvearrowright} = n^{\curvearrowright}$ then $m = n$.

These axioms eliminate Figures 1.5, 1.6 and similar pictures. But there is still a subtle problem. Consider Figure 1.7.

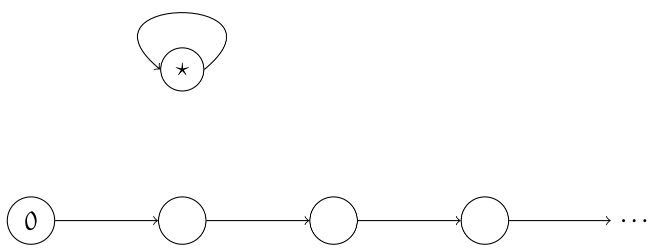


Figure 1.7: A model of the natural numbers?

This picture satisfies the two axioms. Yet, it is certainly not a picture of natural numbers because it has “extra stuff” in it ($\star$).

To rule out “extra stuff”, we formulate our final axiom for natural numbers. The idea is to diagnose the problem as follows. Were we to erase the circle labelled $\star$ and any the arrows leading to and from it, the remaining part of Figure 1.7 would still satisfy live up to the**??**. This is exactly what we mean by “extra stuff”: elements that can be removed without violating violating the signature (the essential structure). This leads to our last axiom.

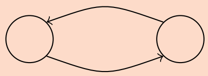
Axiom 1.3

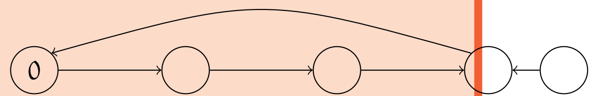
[The Axiom of Induction] No natural numbers can be removed without violating **??**.

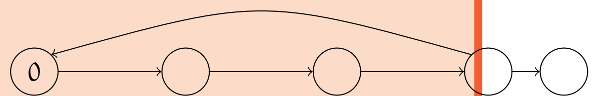
Exercises for Lecture 1

Each of the following pictures fails to satisfy one or more of our axioms. For each, explain which axioms are violated.

1. 

2. 

3. 

4. 

I have in mind a picture for Signature **??** and that satisfies Axioms 1.1 and 1.2. Furthermore, in that picture, I have in mind and element n for which (a) $n \neq 0$ and (b) n has no predecessor (that is, $n \neq m^{\curvearrowright}$ for every m). Convince me that the picture fails to satisfy Axiom 4.3.

The latest exercise shows that in the natural numbers, if $n \neq 0$, then $n = m^{\curvearrowright}$ for some m. In other words, every non-zero natural number has a predecessor.

2

Arithmetic

Adding and multiplying arise from counting. In this section, we explore how to define them purely in terms of counting.

Goals

Lecture
- Present addition and multiplication via defining equations.
- Practice using the defining equations to calculate sums and products.

Study
- Understand addition and multiplication as characterized by defining equations.
- Be able to explain how addition and multiplication relate to counting.
- Exhibit competence in calculating sums and products from the defining equations.

2.1 Basic Arithmetic Operations

Definition 2.1

The *sum* of two natural numbers, m and n, is a natural number (denoted by $m + n$). For every natural number m, the following are true:

$$\begin{aligned} m + 0 &= m \\ m + k^{\frown} &= (m + k)^{\frown} && \text{for any natural number } k \end{aligned}$$

The *product* of two natural numbers, m and n, is a natural number (denoted by $m \cdot n$). For every natural number m, the following are true:

$$\begin{aligned} m \cdot 0 &= 0 \\ m \cdot k^{\frown} &= m + (m \cdot k) && \text{for any natural number } k \end{aligned}$$

A moment's thought about arithmetic should convince you that these equations are reasonable. Certainly $m + 0 = m$ and $m \cdot 0 = 0$ should be true for any m. The second equation for $+$ can be read as saying "to add m to the successor of k, simply add m to k, then take the successor." The second equation for $\cdot$ can be read as saying "to multiply m by the successor of k, simply multiply m by k, and add m to the result."

But it is not the case that any arbitrary pair of equations involving 0 and $k^{\frown}$ will define anything sensible. In the cases of $+$ and $\cdot$, though, the Axiom of Induction ensures that there are indeed unique operations that satisfy the equations. A proof of this fact is not particularly illuminating right now, so let us agree to take for granted that the equations spelled out here actually define addition and multiplication.

Example 2.1

Do the defining equations for addition really explain how to add? Let's use them to calculate $4+3$:

$$\begin{aligned}
4+3 &= 4+0^{\frown\frown\frown} && [\text{3 abbreviates } 0^{\frown\frown\frown}] \\
&= 4^{\frown}+0^{\frown\frown} && [m+k^{\frown}=m^{\frown}+k] \\
&= 4^{\frown\frown}+0^{\frown} && [\text{Same reason}] \\
&= 4^{\frown\frown\frown}+0 && [\text{Same reason}] \\
&= 4^{\frown\frown\frown} && [m+0=m] \\
&= (0^{\frown\frown\frown\frown})^{\frown\frown\frown} && [\text{4 abbreviates } 0^{\frown\frown\frown\frown}] \\
&= 0^{\frown\frown\frown\frown\frown\frown\frown} && [\text{Remove unneeded parentheses}] \\
&= 7 && [\text{7 abbreviates } 0^{\frown\frown\frown\frown\frown\frown\frown}]
\end{aligned}$$

Example 2.2

A product can be calculated similarly. Consider $2 \cdot 2$.

$$\begin{aligned}
2\cdot 2 &= 2\cdot 0^{\frown\frown} && [\text{2 abbreviates } 0^{\frown\frown}] \\
&= 2+(2\cdot 0^{\frown}) && [m\cdot k^{\frown}=m+(m\cdot k)] \\
&= 2+(2+(2\cdot 0)) && [\text{Same reason}] \\
&= 2+(2+0) && [m\cdot 0=0] \\
&= 2+2 && [m+0=m] \\
&= 2+0^{\frown\frown} && [\text{2 abbreviates } 0^{\frown\frown}] \\
&= 2^{\frown}+0^{\frown} && [m+k^{\frown}=m^{\frown}+k] \\
&= 2^{\frown\frown}+0 && [\text{Same reason}] \\
&= 2^{\frown\frown} && [m+0=m] \\
&= (0^{\frown\frown})^{\frown\frown} && [\text{2 abbreviates } 0^{\frown\frown}] \\
&= 0^{\frown\frown\frown\frown} && [\text{Remove unnecessary parentheses}] \\
&= 4 && [\text{4 abbreviates } 0^{\frown\frown\frown\frown}]
\end{aligned}$$

We certainly will not want to calculate this way in real life. After all, it took twelve steps just to figure $2 \cdot 2 = 4$. But these examples and the following exercises show how addition and multiplication are closely tied to simple counting.

Exercises for Lecture 2

Calculate these sums, following the previous example to write each step of your calculation explicitly. Include the reason for each step (as in the previous example).Take care to lay out the chain of equalities correctly, and do not skip any steps.

1. $2+4$
2. $4+2$
3. $3+(3+1)$
4. $(3+3)+1$
5. $0+3$

Notice that it takes more steps to calculate $2+4$ than $4+2$, even though we already knew they would produce the same answer. Explain why.

Calculate the following values, writing each step explicity.

1. $2 \cdot 3$
2. $0 \cdot 2$
3. $2 \cdot (2 \cdot 2)$
4. $3 \cdot (2+1)$
5. $(3 \cdot 2) + (3 \cdot 1)$

Write a definition of exponentiation via defining equations. Follow the pattern of definition I have written for addition and multiplication.

3

Laws of Arithmetic

Before working the last exercises, you knew that $3 \cdot (2+1)$ and $3 \cdot 2 + 3 \cdot 1$ would come out the same because of a law of arithmetic known as *distributivity*. Addition and multiplication satisfy several other laws.

Goals

Lecture

- Present the most common Laws of Arithmetic for natural numbers.
- Explain the method of *proof by simple induction*
- Prove a representative sample of the laws by simple induction.

Study

- Become familiar with the common names for the Laws of Arithmetic.
- Pay particular attention to the Laws of Positivity and Cancellativity (they may be the least familiar to you).
- Demonstrate the ability to identify the main parts of a proof by simple induction.
- Demonstrate the ability to construct the parts of a proof by simple induction.
- Prove the remaining laws for yourself.

3.1 Basic Laws

The following list summarizes several useful laws of arithmetic on the natural numbers. They are organized to emphasize similarities between addition and multiplication.

Laws

For any natural numbers, m, n and p:

Associativity	$m+(n+p)=(m+n)+p$ $m\cdot(n\cdot p)=(m\cdot n)\cdot p$	**Commutativity**	$m+n=n+m$ $m\cdot n=n\cdot m$
Identity	$m+0=m$ $m\cdot 1=m$	**Positivity**	if $m+n=0$ then $m=0$ if $m\cdot n=1$ then $m=1$
Cancellativity	if $m+p=n+p$ then $m=n$ if $m\cdot p^{\frown}=n\cdot p^{\frown}$ then $m=n$		
Distributivity	$m\cdot(n+p)=(m\cdot n)+(m\cdot p)$		
Case Distinction	if $m\neq 0$ then $m=k^{\frown}$ for some k		

Most of these laws are familiar and are listed with their common names. The Law of Case Distinction was the subject of Exercise 1.2. *Go back and look at that exercise again.* The Law of Positivity for multiplication is not a common name, but I have used it to emphasize the analogies between addition and multiplication. Also Case Distinction does not really have a common name. I made that up.

3.2 Inductive Proofs

Suppose we wish to prove that every natural number has some property. For example, let us suppose we wish to prove that every natural number is *mimsy*. I have no idea what a mimsy number is, but let us try to prove this anyway. We could try proving that 0 is mimsy, 1 is mimsy, 2 is mimsy, and so on. But this won't work because our proof will never end. In fact, it is not so obvious that we, humans with finite minds, can ever prove that some property is true for *all* natural numbers, since it seems to involve checking infinitely many individual cases.

The Axiom of Induction provides a way forward in spite of our limitations. Suppose we were to show that the mimsy natural numbers all by themselves constitute a picture of Signature **??**. Then there could not be any natural numbers left out, for otherwise, we could erase all the non-mimsy natural numbers and still have a picture of **??**. This is exactly what the Axiom of Induction forbids: we can not erase *anything* without breaking the signature.

So to prove that all natural numbers are mimsy, we simply need to prove that

- 0 is mimsy, and
- for all natural numbers k, if k is mimsy so is $k^{\frown}$.

From these, we conclude that the mimsy natural numbers by themselves form a picture of **??**. So the Axiom of Induction ensures that all natural numbers are mimsy.

To make inductive proofs easier to understand, we often write them using a three step outline, as illustrated here.

- [Basis] Prove that 0 is mimsy.
- [Inductive Hypothesis] Assume that k is mimsy.
- [Inductive Step] Prove that $k^{\frown}$ is mimsy. [You may use the assumption that k is mimsy in this part of the proof.]

More practical examples are next.

Proposition 3.1

Addition is associative.

Proof: We need to show that $m+(n+p)=(m+n)+p$ for all m, n and p. Let us suppose that m and n are fixed values (not known to us). We now prove that the values p for which $m + (n + p) = (m + n) + p$ holds form a picture of **??**.

- [Basis] $m + (n + 0) = m + n = (m + n) + 0$. Both steps are due to the defining equations of $+$.
- [Inductive Hypothesis] Assume $m + (n + k) = (m + n) + k$.
- [Inductive Step] We must show that $m + (n + k^\frown) = (m + n) + k^\frown$.

$$\begin{aligned} m + (n + k^\frown) &= m + (n + k)^\frown && \text{[Def. of +]} \\ &= (m + (n + k))^\frown && \text{[Same]} \\ &= ((m + n) + k)^\frown && \text{[Inductive Hypothesis]} \\ &= (m + n) + k^\frown && \text{[Def. of +]} \end{aligned}$$

Therefore (by the Axiom of Induction), $m + (n + p) = (m + n) + p$ holds for all p. Since the argument does not depend on any extra assumptions about m and n, it holds for all m and n. □

We say this proof is *by induction on* p to emphasize that the variable p is the focus of attention. The other variables are not directly involved in the structure of the proof.

The remainder of this section further illustrates the technique of simple arithmetic induction via proofs of other laws of arithmetic.

Proposition 3.2

0 is the identity for addition.

Proof: We must prove that $m + 0 = m = 0 + m$ for all m. The first equality is true by the definition of $+$. But the second equality, $m = 0 + m$, is not explicitly one of the defining facts about $+$. So we proceed by induction on m.

- [Basis] $0 + 0 = 0$ is true by definition of $+$.
- [Inductive Hypothesis] Assume $0 + k = k$.
- [Inductive Step] We must show that $0 + k^\frown = k^\frown$.

$$\begin{aligned} 0 + k^\frown &= (0 + k)^\frown && \text{[Def. of +]} \\ &= k^\frown && \text{[Inductive hypothesis]} \end{aligned}$$

Therefore, $0 + m = m$ holds for all m. □

To prove that addition is commutative, we need an additional fact about how successor and addition interact.

Lemma 3.1

For any m and n, $(m + n)^\frown = m^\frown + n$.

Proof: By induction on n:

- [Basis]

$$\begin{aligned}(m+0)^{\frown} &= m^{\frown} && \text{[Def. of +]}\\ &= m^{\frown}+0 && \text{[Def. of +]}\end{aligned}$$

- [Inductive Hypothesis] Assume $(m+k)^{\frown} = m^{\frown} + k$ for some k.
- [Inductive Step] We must show that $(m+k^{\frown})^{\frown} = m^{\frown} + k^{\frown}$.

$$\begin{aligned}(m+k^{\frown})^{\frown} &= ((m+k)^{\frown})^{\frown} && \text{[Def. of +]}\\ &= (m^{\frown}+k)^{\frown} && \text{[Inductive Hypothesis]}\\ &= m^{\frown}+k^{\frown} && \text{[Def. of +]}\end{aligned}$$

So $(m+n)^{\frown} = m^{\frown} + n$. Because the proof does not depend on any assumption about m, it is valid for all m. □

Proposition 3.3

Addition is commutative.

Proof: We need to show that $m+n = n+m$ for all m and n. This time, the proof is by induction on m. Fix a value for n.

- [Basis] $0+n = n = n+0$ holds because of Fact 3.2 and the definition of $+$.
- [Inductive Hypothesis] Assume that $k+n = n+k$ for some k.
- [Inductive Step] We must show that $k^{\frown}+n = n+k^{\frown}$.

$$\begin{aligned}k^{\frown}+n &= (k+n)^{\frown} && \text{[Lemma 3.1]}\\ &= (n+k)^{\frown} && \text{[Inductive Hypothesis]}\\ &= n+k^{\frown} && \text{[Def. of +]}\end{aligned}$$

Therefore, $m+n = n+m$ for all m. Since this argument does not depend on any assumptions about n, it is valid for all n. □

Proposition 3.4

Addition is cancellative.

Proof: We need to prove that if $m+p = n+p$, then $m = n$. This proof is a little subtler than the previous ones. But notice that is still follows the same form.

The proof is by induction on p. Assume that m and n are some fixed natural numbers.

- [Basis] Suppose $m+0 = n+0$. Then immediately by definition of $+$, $m = n$.
- [Inductive Hypothesis] Assume that the following statement is true for some k: if $m+k = n+k$ then $m = n$.
- [Inductive Step] We must show that if $m+k^{\frown} = n+k^{\frown}$ then $m = n$. Suppose $m+k^{\frown} =$

$n + k^{\frown}$ [call this (*) for reference]. Then

$$\begin{aligned}
(m+k)^{\frown} &= m + k^{\frown} && \text{[Def. of } +\text{]}\\
&= n + k^{\frown} && \text{[By the supposition (*)]}\\
&= (n+k)^{\frown} && \text{[Definition of } +\text{]}
\end{aligned}$$

Hence, by Axiom 1.2 $m + k = n + k$. So by the Inductive Hypothesis, $m = n$.

Therefore, $m + p = n + p$ implies $m = n$ for all p. Since this argument does not depend on any assumptions regarding m and n, it is valid for all m and n. □

To prove that multiplication is commutative and cancellative, we will need the following technical facts (analogous to Lemmas 3.2 and 3.1).

Lemma 3.2

$0 \cdot n = 0$

Proof: The proof is by induction on n.

- [Basis] $0 \cdot 0 = 0$ by definition of $\cdot$.
- [Inductive Hypothesis] Assume that $0 \cdot k = 0$ for some k.
- [Inductive Step] We must show that $0 \cdot k^{\frown} = 0$.

$$\begin{aligned}
0 \cdot k^{\frown} &= 0 + 0 \cdot k && \text{[Definition of } \cdot\text{]}\\
&= 0 + 0 && \text{[Inductive Hypothesis]}\\
&= 0 && \text{[Definition of } +\text{]}
\end{aligned}$$

□

Lemma 3.3

$m^{\frown} \cdot n = m \cdot n + n$

Proof: The proof is by induction on n.

- [Basis] $m^{\frown} \cdot 0 = 0 = 0 + 0 = m \cdot 0 + 0$ all follow from the definitions of $+$ and $\cdot$.
- [Inductive Hypothesis] Assume that $m^{\frown} \cdot k = m \cdot k + k$ for some k.
- [Inductive Step] We must show that $m^{\frown} \cdot k^{\frown} = m \cdot k^{\frown} + k^{\frown}$.

$$\begin{aligned}
m^{\frown} \cdot k^{\frown} &= m^{\frown} + m^{\frown} \cdot k && \text{[Def. of } \cdot\text{]}\\
&= (m + m^{\frown} \cdot k)^{\frown} && \text{[Lemma 3.1]}\\
&= (m + (m \cdot k + k))^{\frown} && \text{[Inductive Hypothesis]}\\
&= ((m + m \cdot k) + k)^{\frown} && \text{[Associativity of } +\text{]}\\
&= (m \cdot k^{\frown} + k)^{\frown} && \text{[Def. of } \cdot\text{]}\\
&= m \cdot k^{\frown} + k^{\frown} && \text{[Def. of } +\text{]}
\end{aligned}$$

□

Other laws are left as exercises.

Exercises for Lecture 3

Prove that 1 is the identity for multiplication. That is $1 \cdot m = m = m \cdot 1$.

Prove that multiplication distributes over addition $[m \cdot (n + p) = m \cdot n + m \cdot p]$ by induction on p. You can use the fact that addition is associative and commutative because we have already proved those.

1. Prove the basis: $m \cdot (n + 0) = m \cdot n + m \cdot 0$.
2. Write the inductive hypothesis.
3. Prove the inductive step: $m \cdot (n + k^\frown) = m \cdot n + m \cdot k^\frown$

Prove that multiplication is associative $[m \cdot (n \cdot p) = (m \cdot n) \cdot p]$ by induction on p.

1. Prove the basis: $m \cdot (n \cdot 0) = (m \cdot n) \cdot 0$.
2. Write the inductive hypothesis.
3. Prove the Inductive Step: $m \cdot (n \cdot k^\frown) = (m \cdot n) \cdot k^\frown$. Hint: Use the Law of Distribution, which you just proved.

Prove that multiplication is commutative. Hint: Use the two facts we proved riht before these exercises.

For the record, we also prove the cancellation law for multiplication. This is a bit harder than the exercises, but you should try to find your own proof before looking at mine.

Proposition 3.5

If $m \cdot p^\frown = n \cdot p^\frown$, *then* $m = n$.

Proof: The proof is by induction on n.

- [Basis] Suppose $m \cdot p^\frown = 0 \cdot p^\frown$. From Fact 3.2, $m + m \cdot p = m \cdot p^\frown = 0$. So the Law of Positivity for addition ensures that $m = 0$.
- [Inductive Hypothesis] Assume that for some k, the following is true: if $m \cdot p^\frown = k \cdot p^\frown$, then $m = k$.
- [Inductive Step] Suppose that $m \cdot p^\frown = k^\frown \cdot p^\frown$. Then

$$\begin{aligned} m \cdot p^\frown &= k^\frown \cdot p^\frown && \text{[By assumption]} \\ &= k \cdot p^\frown + p^\frown && \text{[Lemma 3.3]} \\ &= (k \cdot p^\frown + p)^\frown && \text{[Definition of +]} \\ &\neq 0 && \text{[Axiom Nat 1.1]} \end{aligned}$$

  Consequently, $m \neq 0$, for otherwise we would have $m \cdot p^\frown = 0$. Since m is not equal to 0, it is equal to some successor (by the Case Distinction Law). Let j be the predecessor of m, so that $j^{\curvearrowright} = m$.

$$\begin{aligned} j \cdot p^{\curvearrowright} + p^{\curvearrowright} m \cdot p^{\curvearrowright} & & \text{[Lemma 3.3]} \\ = k^{\curvearrowright} \cdot p^{\curvearrowright} & & \text{[By supposition]} \\ = k \cdot p^{\curvearrowright} + p^{\curvearrowright} & & \text{[Lemma 3.3]} \end{aligned}$$

Because addition is cancellative, $j \cdot p^{\curvearrowright} = k \cdot p^{\curvearrowright}$. Now, the Inductive Hypothesis ensures that $j = k$. Hence $m = j^{\curvearrowright} = k^{\curvearrowright}$.

□

4

Lists

Natural numbers constitute an important example of something more general, where objects are built up from simpler ones. The Axiom of Induction captures the idea of building "up" and provides an important method for proving facts about natural numbers.

In this lecture, we develop an analogous way to think about *lists*.

Goals

Lecture Goals

- Introduce a formal counterpart to the informal concept of a list
- Emphasize the close analogy between lists and natural numbers
- Introduce basic operations on lists.

Study Goals

- Demonstrate facility with basic list manipulation including calculating length and concatenation of lists.

4.1 Lists

In this section, we concentrate on the fundamental concept of *lists*. The idea is really meant to be the familiar one, so a list of "to do" items is a list. The alphabetized names on a class roster is a list. We will write lists using square brackets. So for example, $[2, 3, 5, 7]$ is the list of the prime numbers less than 10 in ascending order. For lists, we expect the order to matter. So $[7, 5, 3, 2]$ is a different list.

Something that occurs on a list is called an *item* of the list. We can even specify where it is. So we can talk about the "first", "second" item, and so on, assuming the list has enough items.

Because we have already agreed that natural numbers begin with 0, it turns out to make many things easier if we change the way we talk about items on a list to gibe with the natural numbers. So instead of refering to the "first" item, we might call it the "initial" item. Furthermore, we will number them to start with 0. What I mean is that if $L = [2, 3, 5, 7]$, we will write L_0, L_1, L_2, L_3 for the elements 2, 3, 5, 7, respectively. In short, the "initial" item is indexed by the "initial" natural number 0. The next item after that is indexed by next natural number, $0^\frown$, and so on.

Like natural numbers, lists can be built up by starting with an empty list and incrementally adding items. We have choices for how we might formalize the idea. We will follow a standard that has developed in computer science. Clearly, since we use square brackets to punctuate lists, the empty list should be written as $[\,]$. To add an item to a list, we will conventionally put it on the front.

Given the list $[x, y, z]$, we may build a new list with initial item w and the given list as the rest, resulting in $[w, x, y, z]$. The operation of *prepending* an item to a list is denoted by a colon (:). So $w : [x, y, z]$ *is* the list $[w, x, y, z]$.

The empty list, together with prepending items, gives us a way to construct any list we want.

Example 4.1

Here are some examples.

- $5 : 6 : [4, 5]$ is the same as $5 : [6, 4, 5]$, which is the same as $[5, 6, 4, 5]$.
- $[]$ is the empty list
- $1 : []$ is the same as $[1]$
- $1 : 2 : 3 : 4 : []$ is the same as $[1, 2, 3, 4]$.

Notice that every list is either empty ($[]$) or not. If not, it has the form $x : L$ where x is the initial item and L is the rest of the list. This suggests a signature for lists, not so different from the signature for natural numbers.

Basic Vocabulary 4.1

sig:list Lists have the following basic structure.

- There is a special list, which we call *the empty list* and denote by $[]$.
- For any thing x and any list L, there is another list, obtained by *prepending* x to L. We denote the result by $x : L$.

As with the natural numbers, we need to think about axioms that prevent strange behavior. These are exactly analogous to the axioms of natural numbers. First, $[]$ can not be obtained by adding a new initial item to another list. So

Axiom 4.1

For any list L and any thing x, $[] \neq x : L$.

Likewise, a list that is not empty can only be built one way.

Axiom 4.2

For any things x and y and lists L and M, if $x : L = y : M$, then $x = y$ and $L = M$.

For example, if I tell you that $[2, 3, 4, 5] = x : L$, then you know immediately that $x = 2$ and $L = [3, 4, 5]$.

Finally, lists need an induction axiom that ensures that all lists are built up from $[]$.

Axiom 4.3

[The Axiom of List Induction] No lists can be removed without violating **??**.

This axiom justifies conducting proofs about all lists by a scheme almost identical to simple arithmetic induction. That is, to prove some property is true about all lists, it is enough to show

- [Basis] The property is true about $[]$.
- [Inductive Hypothesis] Assume that the property is true for from list K.

- [Inductive Step] Prove that for any thing x, the property is true about $x : K$. [You may use the assumption about K in this part of the proof.]

Operations on lists can now also be defined by schemes similar to how we defined addition and multiplication on natural numbers. For example, every list has a length. Writing $\mathsf{len}(L)$ for the length of a list, $\mathsf{len}([2, 3, 4]) = 3$. A precise definition is now easy to formulate.

Definition 4.1

For a list L. the *length* of L, denoted by $\mathsf{len}(L)$, is the natural number. This satisfies the following equalities.

$$\begin{aligned} \mathsf{len}([\,]) &= 0 \\ \mathsf{len}(x : L) &= \mathit{len}(L)^{\frown} \end{aligned}$$

Example 4.2

$$\begin{aligned} \mathsf{len}([2,3,4]) &= \mathsf{len}(2 : [3,4]) \\ &= \mathsf{len}([3,4])^{\frown} \\ &= \mathsf{len}(3 : [4])^{\frown} \\ &= \mathsf{len}([4])^{\frown\frown} \\ &= \mathsf{len}(4 : [\,])^{\frown\frown} \\ &= \mathsf{len}(,])^{\frown\frown\frown} \\ &= 0^{\frown\frown\frown} \\ &= 3 \end{aligned}$$

Another common operation on lists is *concatenation*: $[2, 3, 4] \otimes [4, 1, 3] = [2, 3, 4, 4, 1, 3]$, whereby the two lists are simply glued together in their original orders. This is defined precisely by the following.

Definition 4.2

For lists L and M, their *concatenation*, denoted by $L \otimes M$, is a list. For all lists M, the following are true.

$$\begin{aligned} [\,] \otimes M &= M \\ (x : K) \otimes M &= x : (K \otimes M) \qquad \text{for any thing } x \text{ and any list } K \end{aligned}$$

Example 4.3

To calculate $[4,5,2,1] \otimes [3,4,1]$, we can follow a method similar to arithmetic:

$$\begin{aligned}
[4,5,2,1] \otimes [3,4,1] &= (4:5:2:1:[\,]) \otimes [3,4,1] && [[4,5,2,1] \text{ abbreviates } 4:5:2:1:[\,]] \\
&= 4:((5:2:1:[\,]) \otimes [3,4,1]) && [\text{Def. of } \otimes] \\
&= 4:5:((2:1:[\,]) \otimes [3,4,1]) && [\text{Same}] \\
&= 4:5:2:((1:[\,]) \otimes [3,4,1]) && [\text{Same}] \\
&= 4:5:2:1:([\,] \otimes [3,4,1]) && [\text{Same}] \\
&= 4:5:2:1:[3,4,1] && [\text{Same}] \\
&= [4,5,2,1,3,4,1] && [\text{Abbreviation}]
\end{aligned}$$

Now we can prove some useful facts about lists.

Lemma 4.1

On lists, $[\,]$ is the identity for $\otimes$,

Proof: By definition $[\,] \otimes L = L$ always true. But $[\,]$ must also satisfy $L \otimes [\,] = L$ always. We can proceed by induction on L. The proof should look familiar (see the proof of Lemma 3.2).

- [Basis] $[\,] \otimes [\,] = [\,]$ is true by definition of $\otimes$.
- [Inductive Hypothesis] Assume $K \otimes [\,] = K$ for some list K.
- [Inductive Step] Suppose x is some thing. We need to show that $(x:K) \otimes [\,] = x:K$.

$$\begin{aligned}
(x:K) \otimes [\,] &= x:(K \otimes [\,]) && [\text{by definition of } \otimes] \\
&= x:K && [\text{by the Inductive Hypothesis}]
\end{aligned}$$

Thus (by the Axiom of List Induction), the lists for which $L \otimes [\,] = L$ constitute all lists. □

Lemma 4.2

On lists, $\otimes$ is associative.

Proof: We prove $L \otimes (M \otimes N) = (L \otimes M) \otimes N$ using induction on L. This should look familiar. It is almost identicial to the proofs that addition and multiplication are associative.

- [Basis] $[\,] \otimes (M \otimes N) = M \otimes N = ([\,] \otimes M) \otimes N$. Both steps are by the definition of $\otimes$.
- [Inductive hypothesis] Suppose $K \otimes (M \otimes N) = (K \otimes M) \otimes N$ for some particular list K.
- [Inductive step]

$$\begin{aligned}
(x:K) \otimes (M \otimes N) &= x:(K \otimes (M \otimes N)) && \text{Def. of } \otimes \\
&= x:((K \otimes M) \otimes N) && \text{Inductive Hypothesis} \\
&= (x:(K \otimes M)) \otimes N && \text{Def. of } \otimes \\
&= ((x:K) \otimes M) \otimes N && \text{Def. of } \otimes
\end{aligned}$$

So $L \otimes (M \otimes N) = (L \otimes M) \otimes N$ is true for all L. Since the proof does not depend on any special propertis of M and N (except that they are both lists), the result is true for all lists M and N. $\square$

Here is another nice fact that we can prove by induction relating length to concatenation.

Lemma 4.3

For any lists L and M, $\mathsf{len}(L \otimes M) = \mathsf{len}(L) + \mathsf{len}(M)$.

Proof: [This claim is probably fairly obvious to you. Nevertheless, to illustrate the technique of list induction again, we prove it explicitly.]

- [Basis] $\mathsf{len}([]) + \mathsf{len}(M) = 0 + \mathsf{len}(M) = \mathsf{len}(M) = \mathsf{len}([] \otimes M)$. These are by definition of $\otimes$ and $+$.
- [Inductive Hypothesis] Suppose $\mathsf{len}(K \otimes M) = \mathsf{len}(K) + \mathsf{len}(M)$ holds for some particular list K.
- [Inductive Step]

$$\begin{aligned}
\mathsf{len}((x : K) \otimes M) &= \mathsf{len}(x : (K \otimes M)) && \text{Def. of } \otimes \\
&= \mathsf{len}(K \otimes M)^{\frown} && \text{Def. of } \mathsf{len} \\
&= (\mathsf{len}(K) + \mathsf{len}(M))^{\frown} && \text{Inductive Hypothesis} \\
&= \mathsf{len}(K)^{\frown} + \mathsf{len}(M) && \text{Lemma 3.1} \\
&= \mathsf{len}(x : K) + \mathsf{len}(M) && \text{Def. of } \mathsf{len}
\end{aligned}$$

$\square$

Often we will use a list somewhat informally without all the punctuation. For example, we might say "Consider a list $a_0, a_1, \ldots, a_{n-1}$ of real numbers." If we do not intend to use the list itself for anything special, but only want to think about the numbers a_0 through a_n, then there is no need to be formal about it. Also, there is no harm in writing something like this: a_5, a_6, a_7, a_8, where the indices start at 5. The default is to start at 0, but that is merely a convention.

4.2 List Itemization

In a list L, the items are in order. So we can refer to items by their position in the list. There are two standards in mathematics for doing this. Either we start counting from 1 or from 0. Although it may seem unintuitive to start from 0 (meaning that the "initial" item of a list is item number 0), this actually makes many calculations simpler. For that reason, most programming languages use this convention for a lists and arrays. So I will consistently start with 0.

The idea can be made precise as follows.

Definition 4.3

Suppose L is a list and $i < \mathsf{len}(L)$. Then L_i is an item on the list defined as follows.

$$\begin{aligned}
&[]_i \text{is never defined because } 0 \not< \mathsf{len}([]) \\
(x : L)_0 &= x \\
(x : L)_{k^{\frown}} &= L_k && \text{provided that } L_k \text{ is defined}
\end{aligned}$$

This is a precise way of explaining that in a list, for example $L = [a, b, c, d, e]$, we can refer to an item by its *index*, so that $L_0 = a$, $L_1 = b$ and so on, up to $L_4 = e$. Notice that L_k is undefined if $k \geq \mathsf{len}(L)$.

Example 4.4

Suppose $L = [a, b, c, d, e]$. We can calculate L_3 explicitly step by step.

$$\begin{aligned} L_3 &= [a, b, c, d, e]_3 \\ &= (a : b : c : d : e : [\,])_{0^{\frown\frown\frown}} \\ &= (b : c : d : e : [\,])_{0^{\frown\frown}} \\ &= (c : d : e : [\,])_{0^{\frown}} \\ &= (d : e : [\,])_0 \\ &= d \end{aligned}$$

Of course, this is just a very careful (you might even say fussy) way to find item number 3 in the list. In every day use, we humans would not do this. We would simply count forward from the beginning of the list.

Exercises for Lecture 4

Suppose $L = [3, 2, 3, 3, 5]$ and $M = [0, 1, 2, 3, 4, 5]$. Calculate the following explicitly step by step.

1. $\mathsf{len}(L)$
2. L_4
3. $(L \otimes M)_9$

4.3 Lists of a Particular Type

We will commonly need to consider lists in which all elements are similar, such as a list consisting of natural numbers. For example, because we know how arithmetic operations work on natural numbers, we can also define operations on lists of natural numbers using arithmetic. Similar extensions are possible for other operations defined on other types of elements.

To illustrate, suppose L is a list of natural numbers. We can define the *sum* of items on the list in the obvious way, so that the sum of the list $[2, 3, 4]$ is $2 + 3 + 4 = 9$. We make this precise with the following.

Definition 4.4

For a list L of natural numbers, the *sum of* L, denoted by $\sum L$, is a natural number, satisfying

$$\begin{aligned} \sum [\,] &= 0 \\ \sum m : L &= m + \sum L \qquad \text{for any natural number } m \text{ and any list of natural numbers } L \end{aligned}$$

We will introduce variations and extensions of this notation this later. For now, we look only at lists.

Exercises for Lecture 4

Prove using list induction that for any lists of natural numbers,

$$\sum \mathsf{L} + \sum \mathsf{M} = \sum(\mathsf{L} \otimes \mathsf{M})$$

Define the product of lists of natural numbers, following the pattern of our definition for $\sum \mathsf{L}$. The standard notation for a product is $\prod \mathsf{L}$. The result should be that $\prod[2,3,4]$ equals 24. Pay close attention the base case $\prod[\,]$.

Using your definition of products, prove by list induction that for any lists of natural numbers,

$$\prod \mathsf{L} \cdot \prod \mathsf{M} = \prod(\mathsf{L} \otimes \mathsf{M})$$

.

We can also consider lists of integers, lists of real numbers, and so on. We can even think about lists of lists. For example, $[[2,3,4],[4,3,2],[5]]$ is a list consisting of two items: $[2,3,4]$, $[4,3,2]$ and $[5]$. Written this using :, this is list is $[2,3,4] : [4,3,2] : [5] : [\,]$. Suppose we have a list of lists like this we can define the concatenation of all the items. For this example, the result should be $[2,3,4,4,3,2,5]$. The definition of this is exactly analogous to sums and products.

Definition 4.5

For a list $\mathcal{L}$ of lists, the *fold of* $\mathcal{L}$, denoted by $\bigotimes \mathcal{L}$, is a list, satisfying

$$\bigotimes[\,] = [\,]$$

$$\sum \mathsf{M} : \mathcal{L} = \mathsf{M} \otimes \otimes\mathcal{L} \qquad \text{for any list } \mathsf{M} \text{ and any list of lists } \mathcal{L}$$

Compare the definitions of $\sum$, $\prod$ and $\otimes$. They differ only in terms of (i) what is the result for an empty list and (ii) what binary operation is used in the second equation.

Suppose we are given a list $\mathcal{L}$ of lists of natural numbers (like the example just above the latest definition). Then its fold is a list of natural numbers. So this can be summed. That is, $\bigotimes \mathcal{L}$ is a list of natural numbers, and $\sum(\otimes\mathcal{L})$ is a natural number. But we might also apply the summation operation to each list on $\mathcal{L}$ separately, resulting in another list of natural numbers. The idea of applying an operation to each element of a list is called "mapping". In this case, we intend to "map" the operation $\sum$ across lists of lists of natural numbers. Here is a suitable definition.

Definition 4.6

For a list $\mathcal{L}$ of lists of natural numbers, the *mapping of* $\sum$ *on* $\mathcal{L}$, denoted by $\mathsf{map}_{\sum}(\mathcal{L})$ is a list of natural numbers, satisfying

$$\bigotimes[\,] = [\,]$$

$$\sum \mathsf{M} : \mathcal{L} = ()\sum \mathsf{M}) : \mathcal{L} \qquad \text{for any list of natural numbers } \mathsf{M} \text{ and any list of lists of natural numbers } \mathcal{L}$$

Exercises for Lecture 4

Calculate $\sum(\otimes[[3,4,5],[6,3]])$.

Calculate $\mathsf{map}_{\sum}([[3,4,5],[6,3]])$.

Calculate $\sum(\mathsf{map}_{\sum}([[3,4,5],[6,3]]))$.
Prove that $\sum(\otimes\mathcal{L}) = \sum(\mathsf{map}_{\sum}(\mathcal{L}))$ for any list of lists of natural numbers $\mathcal{L}$.

4.4 An Efficient Presentation of Natural Numbers and Lists

The structure of natural numbers and the structure of lists are very similar. This similarity can be exploited to develop a simple way of summarizing their properties.

For natural numbers, 0 and $n^{\curvearrowright}$ are the only ways to construct natural numbers. Operations like addition and multiplication are defined in terms of 0 and $^{\curvearrowright}$, so they do not contribute directly to the *construction* of natural numbers. So we refer to 0 and $\curvearrowright$ as *constructors*.

Axioms 1.1 and 1.2 spell out how these constructors behave. Namely, Axiom 1.1 captures the idea that the two constructors are entirely different from the other: $0 \neq n^{\curvearrowright}$. Axiom 1.2 captures the idea that $^{\curvearrowright}$ constructs distinct natural numbers from distinct natural numbers: $m^{\curvearrowright} = n^{\curvearrowright}$ implies $m = n$ (or equivalently, $m \neq n$ implies $m^{\curvearrowright} \neq n^{\curvearrowright}$).

So the basic ingredients of natural numbers are the constructors 0 and $^{\curvearrowright}$ with the understanding that (a) each produces different results and (b) from different ingredients, $^{\curvearrowright}$ produces different results. In fact, point (b) also applies to 0 trivially, because 0 does not use any ingredients.

We can summarize everything we want to say about natural numbers concisely in the following way.

Definition 4.7

The *natural numbers* are defined *inductively* by

$$n = 0 \mid n^{\curvearrowright}$$

In this notation, the vertical bar separates the different constructors for natural numbers. The first constructor (0) does not depend on anything else. The second constructor depends on a natural number n and produces a new one $n^{\curvearrowright}$. So this gives a very concise description of the signature of natural numbers. Implicitly, this notation is meant to indicate that the two alternatives are completely distinct. This is Axiom 1.1. Also implicitly, the notation is meant indicate that $n^{\curvearrowright}$ produces distinct results from distinct n's. This is Axiom 1.2. By declaring saying that this defines natural numbers *inductively*, we also mean that no natural numbers can be removed without violating the signature.

Now let's consider lists. Again, there are two ways to construct lists. $[]$ and $x : L$ for any thing x and any list L. Likewise, the constructrs are distinct, and $x : L = y : M$ is true if and only if both $x = y$ and $L = M$. So we can encapsulate the definition of lists similarly.

Definition 4.8

The *lists* are defined *inductively* by

$$L = [] \mid x : L \qquad\qquad \text{for any thing } x$$

Later in the course, we will make definitions like this rigorous, and will give other similar ones. For now, I just draw your attention to the similarity between natural numbers and lists, and point out that proofs by induction work thanks to the structure of these definitions.

Part II

Basics of Sets and Functions

5

Sets

Goals

Lecture

- Describe informally the category of sets.
- Define list set notation.
- Introduce the idea of a subset.
- Introduce the axiom of extensionality for sets and some of its consequences.

Study

- Demonstrate ability to determine equality of sets.
- Develop facility in basic set theoretic notation.

Sets are the mathematician's way of thinking about *collections* of objects. Examples will be the set of natural numbers, the set of pairs of natural numbers, the set of real numbers, and so on.

An example is a set representing poker cards. We may denote it by

$$\begin{aligned}\mathsf{Deck} = \{&A\clubsuit, 2\clubsuit, 3\clubsuit, 4\clubsuit, 5\clubsuit, 6\clubsuit, 7\clubsuit, 8\clubsuit, 9\clubsuit, 10\clubsuit, J\clubsuit, Q\clubsuit, K\clubsuit,\\ &A\diamondsuit, 2\diamondsuit, 3\diamondsuit, 4\diamondsuit, 5\diamondsuit, 6\diamondsuit, 7\diamondsuit, 8\diamondsuit, 9\diamondsuit, 10\diamondsuit, J\diamondsuit, Q\diamondsuit, K\diamondsuit,\\ &A\spadesuit, 2\spadesuit, 3\spadesuit, 4\spadesuit, 5\spadesuit, 6\spadesuit, 7\spadesuit, 8\spadesuit, 9\spadesuit, 10\spadesuit, J\spadesuit, Q\spadesuit, K\spadesuit,\\ &A\heartsuit, 2\heartsuit, 3\heartsuit, 4\heartsuit, 5\heartsuit, 6\heartsuit, 7\heartsuit, 8\heartsuit, 9\heartsuit, 10\heartsuit, J\heartsuit, Q\heartsuit, K\heartsuit\}\end{aligned}$$

The elements are arranged here conveniently, but we could just as well have listed the cards in any "shuffled" order. The set of them would be the same.

Functions are the mathematicians way of thinking about operations, such as successor, addition, summation, and so on.

Mathematicians also use functions to model attributes of the things in a collection, like "the color of", "the mass of", "the location of", "the father of", "the favorite book of the person to the left of" and so on. For our example of cards in a poker deck, "rank of" or "suit of" are two attributes. So we might write $\mathsf{rank}(A\diamondsuit) = A$ and $\mathsf{suit}(A\diamondsuit) = \diamondsuit$. In general, $\mathsf{rank}(x)$ and $\mathsf{suit}(x)$ pick out these attributes of any card x. Thus $\mathsf{Rank} = \{A, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K\}$ and $\mathsf{Suit} = \{\clubsuit, \diamondsuit, \spadesuit, \heartsuit\}$ are also sets. There is little sense in saying that rank and suit are "operations", but they are functions *from* Deck to Rank and Deck to Suit, respectively.

The functions rank and suit capture some structure of the elements of Deck. In particular, for any rank r and any suit s, there is exactly one card c so that $\mathsf{rank}(c) = r$ and $\mathsf{suit}(c) = s$. For example, if $\mathsf{rank}(c) = 4$ and $\mathsf{suit}(c) = \spadesuit$, then we know exactly what c must be. So the two functions, in a sense, explain what a card is. We will use functions and sets to discuss more complicated structures, but the idea will be very similar to this simple example.

Taken together, sets and functions constitute a fundamental structure in contemporary mathematics called the *Category of Sets and Functions*. This is a slight lie. Actually, there are many different categories of sets that differ in subtle ways. But for most mathematics, the differences are irrelevant. So in practice, it is safe to talk as if there is just one category of sets. The Category of Sets and Functions sometimes abbreviated as **Set**.

To understand sets and functions as they are used in every day mathematics, we need to answer some questions:

- What do we mean by saying that a set is a collection?
- What do we mean by saying that two sets are equal?
- What do we mean by saying that a function behaves like an attribute?
- What do we mean by saying that two functions are equal?

The answers to these leads to some basic principles for reasoning about sets and functions. Other principles allow us to construct sets and functions to reflect specific structure. For example it will turn out that C (the set of cards) can be constructed as $R \times S$ (the set of ranks times the set of suits). We could be more formal and present these principles as *axioms*, but the word "axiom" has a special connotation in mathematics that we do not need here. Nevertheless, everything we say in these lectures could be presented formally.

5.1 Set Basics

A set consists of things that are "in" the set. All other things are "not in" the set. In our running example, $A\spadesuit$ is in the set C, but 25 is not in C. Let us make the idea precise.

Basic Vocabulary 5.1

A *set* is a mathematical entity X with the following feature. For any thing x, either x *is in* X or x *is not in* X. We write $x \in A$ if x is in X and $x \notin X$ if x is not in X.

The symbol $\in$ is used in mathematics exclusively to indicate membership in a set. You will not see it used in any other way.

For variety, all of the following phrases mean the same thing:

- x is in X
- x is an *element of* X
- x is a *member of* X
- X *contains* x
- x *belongs to* X

Basic Vocabulary 5.1 describes how we can talk about sets and elements, and how to use the notation of membership, but does not tell us that any sets actually exist. Our first remedy for this is to make room for finite sets.

Principle 5.1

[Finite Sets] For any list $L = [a_0, \ldots, a_{n-1}]$, there is a set, denoted by $\{a_0, \ldots, a_{n-1}\}$, so that $x \in \{a_0, \ldots, a_{n-1}\}$ if and only if $x = a_i$ for some $i < n$. More precisely,

- $x \notin \{\}$ for any x ($\{\}$ is said to be *empty*);
- $x \in \{a_0, \ldots, a_n\}$ if and only if $x = a_0$ or $x \in \{a_1, \ldots, a_n\}$.

Example 5.1

Here are some examples of sets built from finite lists:

- $\{\}$ – an empty set;
- $\{1, 2, 5\}$ – a set consisting of three elements;
- $\{\{\}\}$ – a set consisting of one element, which is $\{\}$;
- $\{1, 2, 4, \{1, 2\}\}$ – a set consisting of four elements, 1, 2, 4 and the set $\{1, 2\}$.
- $\{4, 5, \{\}, []\}$ – a set consisting of four elements. Note that the set $\{\}$ and the list $[]$ are not the same things.
- $\{1, 2, 3, 4, 3, 2, 1\}$ – a set consisting of four elements, listing an element twice is redundant.
- The sets Deck, Rank and Suit from the introduction.

The study of finite sets is surprisingly complex, and comprises a large part of the branch of mathematics called *combinatorics*. We will touch on some basics of combinatorics later in the course.

Various infinite sets of numbers also exist. Their existence follows from general principles of set theory. We do not try to justify that claim explicitly for now. Instead, we introduce them informally along with the standard symbols we use to denote them.

Definition 5.1

The following sets are denoted by the special symbols:

$$\begin{aligned}
\mathbb{N} &= \text{the set of natural numbers}\\
\mathbb{Z} &= \text{the set of integers}\\
\mathbb{Q} &= \text{the set of rational numbers}\\
\mathbb{R} &= \text{the set of real numbers}\\
\mathbb{C} &= \text{the set of complex numbers}
\end{aligned}$$

Exercises for Lecture 5

1. Let $A = \{1, \{2, 3\}, 4\}$. Determine which of the following assertions are true.
   a) $1 \in A$
   b) $2 \in A$
   c) $\{\} \in A$
   d) $\{2, 3\} \in A$
   e) $A \in A$
2. In the following examples of sets with elements following a pattern, write an expression for the same set that makes the pattern clearer.
   a) $\{0, 2, 4, \ldots, 100\}$
   b) $\{1, 2, 4, 8, \ldots, 256\}$

c) $\{0, 1, 3, 6, 10, \ldots, 55\}$

3. $\bullet \in$ Suit (the finite set defined above).

5.2 Subsets and Extensionality

A set is meant to be a collection: some things are in, some are not. That's all we can say. Unlike a list, a set has no "initial" element. For example, the set $\{1, 2, 3\}$ is the same as the set $\{2, 3, 1\}$, because both have the same elements. This is one important difference between lists and sets: $[1, 2, 3]$ and $[2, 3, 1]$ are *not* the same list because order matters in lists. To make this precise, we need to be clear about when sets are equal. To help, we introduce an important definition.

Definition 5.2

For sets X and Y, we say that X *is a subset of* Y provided that every element of X is an element of Y. We write this as $X \subseteq Y$, and say that X *is included in* Y. We may also write $Y \supseteq X$ to mean the same thing, and say that Y *is a superset of* B.

If X is *not* a subset of Y, we write $X \not\subseteq Y$. If $X \subseteq Y$ and $Y \not\subseteq X$, then X is called a *proper subset of* Y. To indicate that X is a *proper* subset of Y, we may write $X \subsetneq Y$. Some people write $X \subset Y$ for proper subsets, but we will never use that symbol.

To say $X \subseteq Y$ is exactly to say that for any x, if $x \in X$ then $x \in Y$. In plain English, we may translate it informally as "all Xs are Ys." For example, suppose P is the set of all professors, and H is the set of all human beings. Then $P \subseteq H$ is the (dubious) assertion that "all professors are human beings".

Example 5.2

Here are some examples and counter-examples of the subset relation.

- $\{1, 2, 3\} \subseteq \{0, 1, 2, 3\}$
- $\{\} \subseteq \{0\}$
- $X \subseteq X$ for any set X because, trivially, every element of X is an element of X
- $\{\} \subseteq X$ for any set X because every element of $\{\}$ (there are none) is an element of X
- $\{1, 2, 3\} \not\subseteq \{0, 2, 3\}$ because $1 \in \{1, 2, 3\}$ but $1 \notin \{0, 2, 3\}$
- $\{1, 2, 3\} \subseteq \{2, 3, 1\}$
- $\{\spadesuit\} \subseteq S$

Exercises for Lecture 5

For each of the following pairs of sets, determine whether or not the first is a subset of the second. Explain your answer.

4. $\{0, 1\}$ and $\{1, 0\}$
5. $\{a, b, c, d\}$ and $\{a, b, d, e, c\}$
6. $\{\}$ and $\{\{\}\}$
7. $\{0, 3, 6, 10\}$ and $\{10, 9, 8, 7, 6, 5, 4, 2, 1, 0\}$

We can summarize two useful properties of $\subseteq$ as follows.

- [Reflexivity] For any set X, $X \subseteq X$. We say $\subseteq$ is *reflexive*.
- [Transitivity] For any sets X, Y and Z, if $X \subseteq Y$ and $Y \subseteq Z$, then $X \subseteq Z$. We say $\subseteq$ is *transitive*.

Another familiar example of a reflexive, transitive relation is $\leq$ on the natural numbers. In fact there are many examples of reflexive, transitive relations throughout mathematics. The relation $\leq$ is also *anti-symmetric*, meaning that if $m \leq n$ and $n \leq m$ then $m = n$. Suppose $X \subseteq Y$ and $Y \subseteq X$. Then, by definition X and Y have exactly the same elements. By our understanding of sets as collections, X and Y must be equal. So we state this as another axiom.

Principle 5.2

[**The Axiom of Set Extensionality**] For sets X and Y, if $X \subseteq Y$ and $Y \subseteq X$, then $X = Y$. In other words, $\subseteq$ is anti-symmetric.

Based on this, we can already establish a useful fact: there is exactly one empty set. To set the tone for what follows, we make this a formal claim.

Lemma 5.1

There is exactly one empty set.

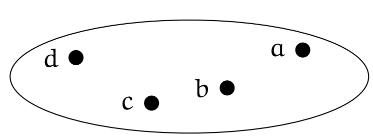
Proof: We have already noted that the set built from an empty list $\{\}$ has no elements. So there is at least one empty set.

Suppose E is a set with no elements. Then $E \subseteq \{\}$ because every element of E (there are none) is an element of $\{\}$. Similarly, $\{\} \subseteq E$ because every element of $\{\}$ (again, there are none) is an element of E. So by Principle 5.2 $E = \{\}$. □

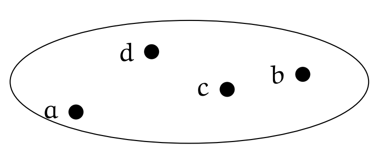
Definition 5.3

The set $\{\}$ is also denoted by $\emptyset$.

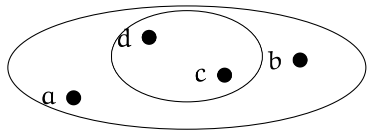
Set extensionality makes precise the idea that a set by itself does not have any structure other than what members it possesses. To emphasize this, sometimes it is useful to depict a set with elements scattered about something like



with the elements scattered about. Evidently, a re-arrangement of the elements does not change the depicted set. So



is the same set. Depicting a subset of a set is a simple matter of drawing a smaller boundary around some of the elements as in the following.



Exercises for Lecture 5

Draw depictions of the following sets

8. $\{1, 4, 5, 2, 3\}$
9. $\{1, 2, 3, \ldots, 23\}$
10. $\{a, b, c, d, e\}$ and $\{c, e, f, g\}$ on the same diagram
11. $\{a, e, b, c, e\}$ [sic]
12. $\{1, 3, 6, 7\}$ and $\{1, 3, 5, 6, 7, 9\}$ on the same diagram
13. $\{\perp, \top$
14. $\{\bullet\}$
15. $\{\top, \perp, 3, 5, 1, \bullet\}$

6

Functions

Goals

Lecture

- Introduce basic structure of functions
- Define the identity functions and function composition
- Introduce internal diagrams of functions.

Study

- Be able to determine equality of functions
- Use internal diagrams to depict function composition

Functions (perhaps in your calculus courses) are often talked about as *operations*. For example,

$$f(x) = x^2 - 1$$

can be seen as an operation that transforms a number x into its square. But it can also be seen as an attribute (the "square of x"). The "operational" view is informal, and often useful. As we will see, though, it gets an important aspect of functions wrong because two entirely different operations may define the same function.

Informally, a function "takes" an element of a given set as input and "produces" an element of a given set as output. So the function f defined by $f(x) = x^2 + 2x + 1$ might "take" the natural number 2 and "produce" the natural number 10. That is, $f(3) = 3^2 + 1 = 10$. We begin by making this idea formal, introducing the vocabulary of functions.

Basic Vocabulary 6.1

- For a set X and a set Y, there are things called *functions from* X *to* Y. We write $f\colon X \to Y$ or $A \xrightarrow{f} B$ to indicate that f is a function from X to Y.
- For $f\colon X \to Y$, the set X is called the *domain* of f and Y is called the *codomain* of f.
- For any function $f\colon X \to Y$ and any element $a \in X$, f and a determine an element of Y, written $f(a)$, and read "f of a".

A function may sometimes also be called a *map*, a *transformation*, or an *operation*. As we will see, however, *operation* is somewhat misleading, so we usually avoid it.

Often, a function $f\colon A \to B$ is *defined* by a rule, just as they are in other parts of mathematics. We typically, write such rules by giving the function a name (very often f because we are lazy) and spelling out the rule at the same time. So we write things like

$$f(x) = x^2 + 4x + 2$$

to define a function $f\colon \mathbb{R} \to \mathbb{R}$ (recall that $\mathbb{R}$ is the set of real numbers). But sometimes it is useful to have a rule without giving it a name. To do that, we will use the "maps to" arrow $\mapsto$. So we may define the same function f by saying that f *is given by the rule*

$$x \mapsto x^2 + 4x + 2.$$

We will not go so far as to write $f = (x \mapsto x^2 + 4x + 2)$ because this more easily understood by writing $f(x) = x^2 + 4x + 2$. The rule $x \mapsto x^2 4x + 2$ is the same as the rule $y \mapsto y^2 + 4y + 2$. The variable only serves as a place holder, so its particular name does not matter.

There are two fundamental (trivial) types of rules that can be used to build functions.

Principle 6.1

- For any set X, there is a function $\mathrm{id}_X\colon X \to X$ defined by the rule $x \mapsto x$. This is called the *identity* function on A.
- For any two functions $f\colon X \to Y$ and $g\colon Y \to Z$, there is a function $g \circ f\colon X \to Z$ defined by the rule $x \mapsto g(f(x))$. This is called the *composition of* g *and* f (or sometimes g *following* f).

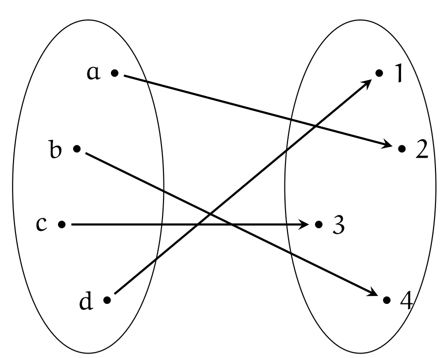
Notice that $g \circ f$ is only defined when the *domain* of g matches the *codomain* of f.

Exercises for Lecture 6

1. Suppose $f\colon W \to X$, $g\colon X \to Y$ and $h\colon Y \to Z$ are functions. Then $h \circ (g \circ f)$ and $(h \circ g) \circ f$ are functions from W to Z. Do you think they are equal? Explain your answer in a few clearly written sentences.

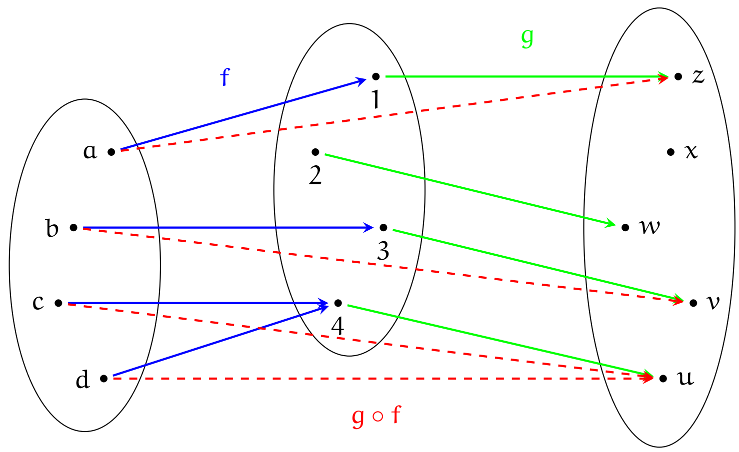
6.1 Internal Diagrams

To depict a function on small sets, we can use the simple internal diagrams of the last section with arrows indicating the input/output relationship. For example,



depicts a function from the set $\{a, b, c, d\}$ to the set $\{1, 2, 3, 4\}$.

Composition can be illustrated using internal diagrams. For example,



Exercises for Lecture 6

Use internal diagrams for the following exercises.

2. Depict four different functions from the set $\{1, 2, 3\}$ to the set $\{\perp, \top\}$. [Draw four different diagrams.]

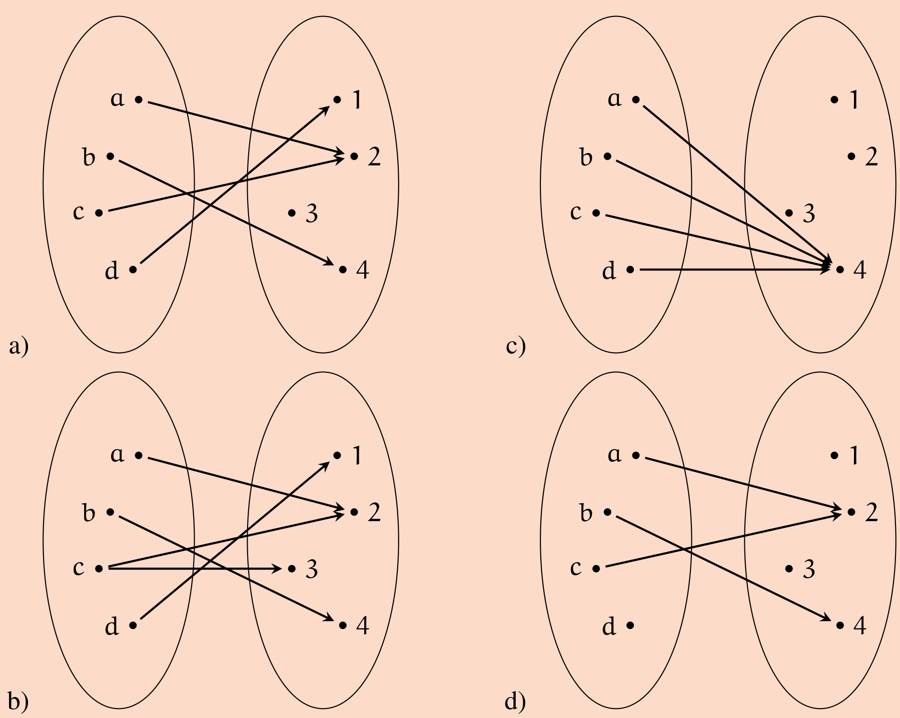
3. Depict all of the functions from $\{\bullet\}$ to $\{a, b, c\}$

4. Depict all of the functions from $\{a, b, c\}$ to $\{\bullet\}$

5. Are there any functions from $\{a, b\}$ to $\emptyset$?

6. Are there any functions from $\emptyset$ to $\{a, b\}$? If there are, how many?

7. For each of the following diagrams, determine whether or not the diagram depicts a function. If not, explain why not.



8. Let $A = \{1, 2, 3\}$. Let $B = \{a, b, c, e\}$ and let $C = \{\perp, \top\}$. Depict some functions $f\colon A \to B$, $g\colon B \to C$, and $g \circ f$.

9. Think about how you might depict a function $h\colon A \to A$ using only one picture of the set A. Describe what you would do, and provide an example.

10. Suppose $f\colon \mathbb{R} \to \mathbb{R}$ is given by the rule $x \mapsto x^2$, suppose $g\colon \mathbb{R} \to \mathbb{R}$ is given by the rule $x \mapsto x - 1$. Write rules for $f \circ g$ and $g \circ f$ without using the symbols f and g. Explain whether or not it is the case that $f \circ g = g \circ f$.

6.2 Extensionality

As with sets, we need a way to say when two functions are equal. Consider an example. Recall that $\mathbb{N}$ denotes the set of natural numbers. Then define $f\colon \mathbb{N} \to \mathbb{N}$ and $g\colon \mathbb{N} \to \mathbb{N}$ by

$$f(n) = n^2 + 2n + 1$$
$$g(n) = (n+1)^2$$

Evidently, for each $n \in \mathbb{N}$, it is true that $f(n) = g(n)$. So even though f and g are defined by different *operations*, the two functions yield the same results. As with sets, this leads to an axiom for equality of functions.

Principle 6.2

[**The Axiom of Function Extensionality**] For functions $f\colon X \to Y$ and $g\colon X \to Y$, if it is the case that $f(x) = g(x)$ for all $x \in X$, then $f = g$. Note that equality of functions only makes sense when the two functions share the same domain and the same codomain.

When we are not concerned about the detailed internals of sets, but only with how functions interact, then an individual function can be depicted very simply as $X \xrightarrow{f} Y$. So a composition of functions can be depicted as in

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[dr, "g \circ f"'] & Y \arrow[d, "g"] \\ & Z \end{tikzcd}$$

We do not really need to draw $g \circ f$ as a separate arrow because the *path* from X to Y to Z is already implicitly a depiction of $g \circ f$. So the simpler diagram

$$\begin{tikzcd} X \arrow[r, "f"] & Y \arrow[d, "g"] \\ & Z \end{tikzcd}$$

shows the same information, namely, that $f\colon X \to Y$ and $g\colon Y \to Z$ are functions and therefore, $g \circ f\colon X \to Z$ is too.

Now a diagram such as this

$$\begin{tikzcd} W \arrow[r, "f"] \arrow[d, "h"'] & X \arrow[d, "g"] \\ Y \arrow[r, "k"'] & Z \end{tikzcd}$$

depicts two composite functions $g \circ f$ and $k \circ h$, but $g \circ f$ and $k \circ k$ may not be equal. We say that the diagram *commutes* or that it is a *commutative diagram* if $g \circ f = k \circ h$. In other words, saying that a certain diagram commutes *is* an assertion that certain functions are equal.

Exercises for Lecture 6

11. For each of the following pairs of functions $\mathbb{N} \to \mathbb{N}$, determine whether they are equal and explain why or why not.

    a) $f(n) = 2n + 3$ and $g(m) = 2m + 3$
    b) $f(n) = 2^{n+1} - 1$ and $g(n) = \sum_{i=0}^{n} 2^i$
    c) $f(n) = n^2 + 5n + 6$ and $g(n) = (n + 3)(n + 2)$
    d) $f(n) = n^4 - 10n^3 + 35n^2 + 50n + 24$ and $g(n) = 24$

12. Let $\mathbb{R}$ denote the set of all real numbers. Let $f(x) = \tan(x)$. Explain why this does *not* define a function from $\mathbb{R}$ to $\mathbb{R}$.

13. Suppose the following functions exist: $f\colon W \to X$, $g\colon X \to Y$, $a\colon W \to Z$, $b\colon Y \to Z$. Draw a commutative diagram asserting that $b \circ g \circ f = a$.

14. Suppose the following functions exist: $f\colon C \to A$, $g\colon C \to B$, $h\colon C \to P$, $p\colon P \to A$ and $q\colon P \to B$. Draw a commutative diagram asserting that $f = p \circ h$ and $g = q \circ h$.

7

Basic Building Blocks

Goals

Lecture

- Characterize and define
  - Pointer and constant functions
  - Solution sets
  - Characteristic functions
  - Products of sets
  - Exponents of sets
- Introduce the idea of a *universal* construction.

Study

- Be able to calculate membership in various constructed sets
- Learn to use universal constructions to define functions.

So far, we have thought mainly about informally defined sets and functions. To fill out our understanding of sets, we need to be able to build sets for specific purposes and with specific structure in mind.

Three finite sets will play particularly important roles. We have already discussed the set $\emptyset$, consisting of no elements. We also need a designated set with one element and a designated set with two elements. We denote these by $\mathbb{1}$ and $\mathbb{2}$. It does not matter at all *what* elements are in these because, as we will soon see, any two sets of the same size are interchangeable. What 'interchangeable' means is discussed later. What 'same size' means is obvious for finite sets. We discuss the general situation later as well.

For the time being, we merely need to agree on a fixed set with one element and a fixed set with two elements. The particular choices we make here will be clearer as we put them to use.

Definition 7.1

Let $\bullet$, $\perp$ and $\top$ be fixed symbols. Then define

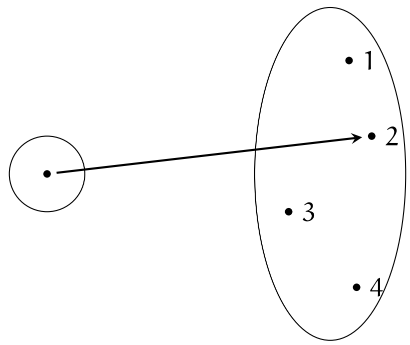
$$\mathbb{1} = \{\bullet\}$$
$$\mathbb{2} = \{\perp, \top\}$$

The single element of $\mathbb{1}$ is intended to look like a generic point in an internal diagram. The element $\top$ is meant to remind you of the letter 'T' (short for 'True') and $\perp$ is meant to be the opposite of $\top$ (that is, 'False').

As mentioned above, the particular choice of elements is not important. Some folks would define $\mathbb{1}$ to be $\{0\}$ and $\mathbb{2}$ to be $\{0, 1\}$. This is convenient because the pattern continues: would be $\{0, 1, 2\}$, and so on. We, however, wish to emphasize other things that have nothing directly to do with the numbers 0 and 1 (as members of $\mathbb{1}$ and $\mathbb{2}$).

7.1 Elements, Pointers and Constant Functions

Suppose we are told that $p \colon \mathbb{1} \longrightarrow X$ is a function. Since $\bullet \in \mathbb{1}$, this function determines an element of X, namely $p(\bullet)$. A picture of the situation might be this:



Since $\mathbb{1}$ has only a single element, it can "point" only to a single element of X. So we might refer to a function $\mathbb{1} \longrightarrow X$ as a *pointer* into X. Each pointer determines an element of X. And conversely, it should be possible to point to any element of X. This leads to our first principle guaranteeing that certain (nearly trivial) functions exist.

Principle 7.1

For any set X, and any $a \in X$, there is a function $\hat{a} \colon \mathbb{1} \longrightarrow X$ given by the rule $x \longmapsto a$.

Thus the function depicted above is $\hat{2}$. In effect, this principle simply asserts that elements of a set X and functions $\mathbb{1} \longrightarrow X$ are interchangible: from $a \in X$ we get $\hat{a} \colon \mathbb{1} \longrightarrow X$; from $p \colon \mathbb{1} \longrightarrow X$. we get the element $p(\bullet)$.

This principle also justifies the drawing of internal diagrams for pointers, like the one above, because it means that any such diagram is guaranteed to depict an actual function.

Suppose $f \colon X \longrightarrow \mathbb{1}$ and $g \colon X \longrightarrow \mathbb{1}$ are functions, that is, their *codomain* is $\mathbb{1}$ instead their *domain*. Then $f(a) = g(a)$ is true for every $a \in X$ because $\bullet$ is the only possible value: $f(a) = \bullet = g(a)$. So $f = g$ by the Principle of Function Extensionality. In other words, there is at most one function from X to $\mathbb{1}$. But the rule $x \longmapsto \bullet$ is as simple a rule as one can imagine. This leads to another definition and another principle.

Definition 7.2

A set T is *terminal* if it is the case that for any set X there is exactly one function from X to T.

Principle 7.2

The set $\mathbb{1}$ is a terminal set. We denote the unique function from X to $\mathbb{1}$ by $\diamond_X \colon X \longrightarrow \mathbb{1}$.

The rule defining $\diamond_X$ must be

$$x \longmapsto \bullet.$$

Using $\Diamond_X$ and $\hat{b}$ for an element $b \in Y$, we can now define a constant function. That is $\hat{b} \circ \Diamond_X$ is a function from X to Y given by the rule $x \mapsto \hat{b}(\Diamond_X(x))$. But since $\hat{b}(\Diamond_X(x)) = \hat{b}(\bullet) = b$, the rule simplifies to $x \mapsto b$. In short, this is the function sending any element of X to the constant b.

Exercises for Lecture 7

1. Show that for any pointer $p\colon \mathbb{1} \to X$, it is the case that $\widehat{p(\bullet)} = p$.
2. Show that any set with exactly one element is a terminal set.
3. Suppose that $f\colon X \to Y$ is a function. Show that for every $a \in A$, $\widehat{f(a)} = f \circ \hat{a}$.

7.2 The Empty Set

For trivial reasons, there is at most one function from $\emptyset$ to A, for any set A. That is, if $f, g\colon \emptyset \to A$ are functions, then for each $x \in \emptyset$, $f(x) = g(x)$ because there are no x's to concern us. Hence by Principle **??**, $f = g$. The empty "rule" that tells us to do nothing actually specifies a function from $\emptyset$ to X. So for any set X, there is exactly one function from $\emptyset$ to X. Let's make that official.

Definition 7.3

An *initial set* is a set I so that for any set X, there is exactly one function from I to X.

Principle 7.3

The emptyset $\emptyset$ is an initial set. For a set X, the unique function from $\emptyset$ to X (given by the empty rule) may be denoted by $\Box_A\colon \emptyset \to X$.

Notice that a function $X \to \emptyset$ is impossible unless X is also empty. So $\emptyset$ is the only initial set.

Exercises for Lecture 7

4. How many functions are there from $\emptyset$ to $\{a, b, c, d\}$? Explain.
5. How many functions are there from $\{a, b, c, d\}$ to $emptyset$? Explain.
6. How many functions are there from $\mathbb{1}$ to $\{a, b, c, d\}$? Explain.
7. How many functions are there from $\{a, b, c, d\}$ to $\mathbb{1}$? Explain.

7.3 Solution Sets, Subsets, Characteristic Functions

Suppose we are given two functions that are "parallel": $f\colon X \to Y$ and $g\colon X \to Y$. To aid readability, we will write this as $X \underset{g}{\overset{f}{\rightrightarrows}} Y$. For some values $a \in X$, it might be the case that $f(a) = g(a)$. We may call such a value a *particular solution to the equation* $f(x) = g(x)$.

It might be the case that there are no particular solutions to an equation $f(x) = g(x)$. For example, there are no natural numbers n such that $n + 1 = n$. On the other hand, there might be many particular solutions. For example, let $f(x) = x^3$ and let $g(x) = 6x^2 - 11x + 6$ both regarded as functions on the natural numbers. Then it is easy to check that 1, 2 and 3 solve the equation $f(x) = g(x)$. In fact, these three are the only particular solutions. We generalize as follows.

Definition 7.4

For two functions $X \underset{g}{\overset{f}{\rightrightarrows}} Y$, a *solution* is a function $S \xrightarrow{s} X$ so that $f \circ s = g \circ s$. Thus for example, if $a \in A$ is a particular solution then the pointer $\hat{a}$ is a solution.

For functions $A \underset{g}{\overset{f}{\rightrightarrows}} B$, an *equalizer* is a solution $E \xrightarrow{e} X$ so that for any solution $S \xrightarrow{s} X$, there is exactly one function $S \xrightarrow{h} E$ so that $e \circ h = k$.

Principle 7.4

For functions $X \underset{g}{\overset{f}{\rightrightarrows}} Y$, the collection of all particular solutions to the equation $f(x) = g(x)$ form a set, denoted by $\{x \in X \mid f(x) = g(x)\}$. The function $\{x \in X \mid f(x) = g(x)\} \xrightarrow{i} A$ given by the rule $x \mapsto x$ (called an *inclusion map*) is an equalizer for f and g.

If $S \xrightarrow{s} X$ is a solution (that is, $f \circ s = g \circ s$), then the function $C \xrightarrow{\check{s}} \{x \in A \mid f(x) = g(x)\}$ given by the rule

$$x \mapsto s(x)$$

is the unique function for which $s = i \circ \check{s}$.

This axiom tells us three main things. First, we can form a subset of X by specifying an equation $f(x) = g(x)$ for any two functions $X \underset{g}{\overset{f}{\rightrightarrows}} Y$, and picking out the particular solutions. Second, a subset formed in this way "embeds" in the given set X by its inclusion map i. Third, for any solution s, the function into the set of particular solutions is defined by the same rule as s.

External diagrams help us understand equalizers. An equalizer is a solution

$$\begin{tikzcd} E \arrow[r, "e"'] & X \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & Y \end{tikzcd}$$

so that if

$$\begin{tikzcd} S \arrow[dr, "s"] & & \\ E \arrow[r, "e"'] & X \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & Y \end{tikzcd}$$

is also a solution ($f \circ s = g \circ s$), then there is exactly one function making

$$\begin{tikzcd} S \arrow[d, "\check{s}"'] \arrow[dr, "s"] & & \\ E \arrow[r, "e"'] & X \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & Y \end{tikzcd}$$

commute.

Inverse Images

Suppose $c \in Y$ and $X \xrightarrow{f} Y$ is a function, then we can form the equalizer of f and the constant function $\hat{c} \circ \diamond_X$. This is more easily written we $\{x \in X \mid f(x) = c\}$. Since it is common to pick out sets like this, special notation is in order.

Definition 7.5

For a function $X \xrightarrow{f} Y$, and a value $c \in Y$,

$$f^{-}(c) = \{x \in X \mid f(x) = c\}.$$

In this case, $f^{-}(c)$ is called the *inverse image of* c *with respect to* f.

A diagram can help understand inverse images as well. Suppose $f \colon X \to Y$ and $c \in C$, then we can arrange a diagram

$$\begin{tikzcd}
 & \mathbb{1} \arrow[d, "\hat{c}"'] \\
X \arrow[r, "f"'] & Y
\end{tikzcd}$$

The inverse image is a subset of X with an inclusion map that makes the following diagram commute:

$$\begin{tikzcd}
f^{-}(c) \arrow[r, "\diamond_{f^{-}(c)}"] \arrow[d, "i"'] & \mathbb{1} \arrow[d, "\hat{c}"] \\
A \arrow[r, "f"'] & C
\end{tikzcd}$$

For any other function $W \xrightarrow{g} X$ that makes following diagram commute:

$$\begin{tikzcd}
W \arrow[drr, "\diamond_W"] \arrow[ddr, "g"'] & & \\
 & f^{-}(c) \arrow[r, "\diamond_{f^{-}(c)}"'] \arrow[d, "i"'] & \mathbb{1} \arrow[d, "\hat{c}"] \\
 & A \arrow[r, "f"'] & C
\end{tikzcd}$$

there is a unique function $B \xrightarrow{\check{g}} f^{-}(c)$ making

$$\begin{tikzcd}
W \arrow[drr, "\diamond_W"] \arrow[ddr, "g"'] \arrow[dr, "\check{g}"] & & \\
 & f^{-}(c) \arrow[r, "\diamond_{f^{-}(c)}"'] \arrow[d, "i"'] & \mathbb{1} \arrow[d, "\hat{c}"] \\
 & X \arrow[r, "f"'] & Y
\end{tikzcd}$$

commute.

In Definition 7.5, the set $f^{-}(c)$ is a subset of X. It would be good to know that any subset of X can be described as an inverse image. This is where the set $\mathbb{2}$ plays a role.

Definition 7.6

A *pointed set* is a set P with a distinguished element $p \in P$.

A *subset classifier* is a set T with a distinguished element $t \in T$ so that for any set X and any subset $A \subseteq X$, there is exactly one function $k\colon X \longrightarrow T$ for which $A = k^{-}(t)$. That is, A is uniquely defined as the inverse image of t with respect to a function into T.

Principle 7.5

The set $\mathbb{2}$ with the distinguished element $\top$ is a subset classifier. For subset $A \subseteq X$, the function corresponding to A, called the *characteristic function of* A, is denoted by κ_A. In other words, κ_A is the unique function for which $A = \kappa_A^{-}(\top)$.

For $A \subseteq X$, the characteristic function is defined by the rule

$$x \mapsto \begin{cases} \top & \text{if } x \in A \\ \bot & \text{otherwise} \end{cases}$$

Just as Principle 7.1 asserts that elements of X and functions $\mathbb{1} \longrightarrow X$ are interchangeable, Principle 7.5 asserts that the subsets of X and the functions $X \longrightarrow \mathbb{2}$ are interchangeable.

Exercises for Lecture 7

8. Draw a depiction of $A = \{a, b, c, d, e, f, g\}$ and its subset $B = \{a, c, e, g\}$ in the same internal diagram. Now depict the characteristic map for B as a subset of A.

9. Define two functions $\mathbb{N} \underset{g}{\overset{f}{\rightrightarrows}} \mathbb{N}$ so that the set of particular solutions fo $f(x) = g(x)$ is $\{1, 5\}$.

10. Consider the functions $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ defined by $f(x) = \sin(x) + \cos(x)$, and $s\colon \mathbb{N} \longrightarrow \mathbb{R}$ defined by $s(n) = 2\pi n^2$. Is s a solution for the equation $f(x) = -1$? What is the set of all particular solutions?

11. Describe what a subset classifier is, using diagrams similar to the diagrams we have used to describe equalizers and inverse images. That is, for a start, we have a diagram

$$\begin{tikzcd} \mathbb{1} \arrow[d, "\hat{t}"] \\ T \end{tikzcd}$$

Now suppose we are given a subset $A \subseteq X$ with its inclusion map:

$$\begin{tikzcd} A \arrow[r, "\diamond_A"] \arrow[d, "i"'] & \mathbb{1} \arrow[d, "\hat{t}"] \\ X & S \end{tikzcd}$$

What additional function is required to exist? What properties is it required to have? [Hint: the result should be that A is an inverse image.]

7.4 Product Sets and Functions of Two Arguments

We should be able to deal with functions of more than one argument, such as a function $f(x, y) = x + y$. To account for these, we take our cue from Descartes.

Descartes studied the geometric plane in terms of a coordinate system consisting of the so-called x-axis and y-axis (what we call cartesian coordinates in his honor). Once we have decided where to place the axes (as long as they do not run in parallel), a pair such as $(2, 3)$ determines a point on the plane, and any point p in the plane determines a pair. So Descartes realized that we might as well just say that the plane actually *is* the collection of all pairs of real numbers. What makes this work is that points in the plane *project* onto the two axes in a universal way. Products of sets generalize this idea.

Definition 7.7

For sets X and Y, a *table* consists of two functions $X \xleftarrow{f} T \xrightarrow{g} Y$. Note that the two functions have the same domain. We may call the two functions *legs* of the table.

For sets X and Y, a *product of* X *and* Y is a table $X \xleftarrow{p} P \xrightarrow{q} Y$ so that for any table $X \xleftarrow{f} T \xrightarrow{g} Y$ there is exactly one function $T \xrightarrow{h} P$ for which $f = p \circ h$ and $g = q \circ h$. For a product, the legs p and q are called the *projections*.

Principle 7.6

For sets X and Y, the collection of all pairs (x, y) where $x \in X$ and $y \in Y$ is a set, denoted by $X \times Y$. The functions $X \xleftarrow{\pi_0} X \times Y \xrightarrow{\pi_1} Y$ given by the rules $(x, y) \mapsto x$ and $(x, y) \mapsto y$ are projections. For $X \xleftarrow{f} T \xrightarrow{g} Y$, the unique function required by the product may be denoted by $\langle f, g \rangle$.

For $X \xleftarrow{f} T \xrightarrow{g} Y$, the function $T \xrightarrow{\langle f,g \rangle} X \times Y$ is given by the rule $t \mapsto (f(t), g(t))$.

As with equalizers and inverse images, products can be described in terms of diagrams.

A product of X and Y is depicted as a diagram

$$\begin{tikzcd} X & P \arrow[l, "p"'] \arrow[r, "q"'] & Y \end{tikzcd}$$

so that for any other table over X and Y:

$$\begin{tikzcd} & T \arrow[dl, "f"'] \arrow[dr, "g"] & \\ X & P \arrow[l, "p"] \arrow[r, "q"'] & Y \end{tikzcd},$$

there is a unique function making

$$\begin{tikzcd} & T \arrow[dl, "f"'] \arrow[dr, "g"] \arrow[d, "\langle f, g \rangle"'] & \\ X & P \arrow[l, "p"] \arrow[r, "q"'] & Y \end{tikzcd},$$

commute.

Suppose we are given two unrelated functions $X \xrightarrow{f} Y$ and $A \xrightarrow{g} B$. We can now form a single function from $X \times A$ to $Y \times B$ by combining f and g "independently". That is, define $f \times g = \langle f \circ \pi_0, g \circ \pi_1 \rangle$. Calculating concretely in terms of elements $(f \times g)(x, y) = (f(x), g(y))$. So $f \times g$ acts on a pair (x, y) by applying f to x and unrelatedly applying g to y.

Products can by generalized to three, four or more sets. For example, given sets X, Y and Z, we might write $X \times Y \times Z$ for the set of triples (x, y, z) where $x \in X$, $y \in Y$ and $z \in Z$. Instead of two projections, this would have three projections $(x, y, z) \mapsto x$, and so on. It turns out, however, that binary products are enough because $X \times (Y \times Z)$ behaves just like $X \times Y \times Z$.

Exercises for Lecture 7

12. For the sets $A = \{a, b, c\}$ and $B = \{1, 2, 3, 4\}$, write out $A \times B$ and $B \times A$
13. What is $\emptyset \times A$?
14. Write out $\{4, a, 0\} \times 2$.
15. Describe in plain English what are the elements of $\mathbb{N} \times \mathbb{N}$.
16. Suppose A is a finite set with m elements and B is a finite set with n elements. How many elements are in $A \times B$?
17. Describe in plain English why it makes sense to refer to $A \times B$ as a "product."
18. For sets $A = \{a, b\}$, $B = \{0, 1, 2\}$ and $C = \{c, d\}$, calculate $A \times B \times C$, $A \times (B \times C)$ and $(A \times B) \times C$. Are these sets equal? If not, how do they differ.
19. Describe a model of the standard fifty-two card poker deck as a product of two sets.

7.5 Function Sets and Parametric Functions

A function from X to Y might depend on a parameter from another set P. For example, the function $sf \colon \mathbb{R} \to \mathbb{R}$ defined by the rule $f(x) = \sin(x + c)$ depends on the constant c. There is a related function $\mathbb{R} \times \mathbb{R} \xrightarrow{g} \mathbb{R}$ defined by $g(c, x) = \sin(x + c)$. Though g describes the same behavior as f, it makes the parameter explicit as another argument. It will be helpful to relate g – a function on pairs – to f a function with a parameter. This leads to the following definition.

Definition 7.8

For sets X and Y, a *parametric function from* X *to* Y is a function $P \times X \xrightarrow{g} Y$ for some set P. The set P may be called the *set of parameters*.

Suppose $Q \times X \xrightarrow{f} Y$ is a parametric function with parameter set Q and $P \xrightarrow{k} Q$ is a function. Then we can form another parametric function with parameters in P by composing: $f \circ (k \times \mathsf{id}_A)$. The function k acts lie a *change of parameters* because it transforms the parametric function with parameters in Q into a parametric function with parameters in Q. Specifically, $f \circ (k \times \mathsf{id}_X)$ is given by the rule $(c, a) \mapsto f(k(c), a)$.

An *evaluation map* for X and Y is a parametric function $F \times X \xrightarrow{a} Y$, so that for any parametric function $P \times X \xrightarrow{g} Y$ there is exactly one change of parameters $h \colon P \to F$ so that $g = a \circ (h \times \mathsf{id}_X)$. In that case, F is called an *exponential with base* Y *and exponent* X.

Principle 7.7

For sets X and Y, the collection of all functions from X to Y, denoted by Y^X, is a set.

The rule $(f, x) \mapsto f(x)$ defines an evaluation map $Y^X \times X \xrightarrow{\mathsf{appl}} Y$.

For a parametric function $f\colon P \times X \to Y$, the unique function from P to Y^X determined by f does not have a completely standard name. Increasingly, mathematicians honor the twentieth century logician, Haskell Curry, by referring to this as 'currying'. For these lectures, we follow that tradition and write $\mathsf{curry}[f]$ for the unique function satisfying $f = \mathsf{appl} \circ (\mathsf{curry}[f] \times \mathsf{id}_X)$.

Calculating how $P \xrightarrow{\mathsf{curry}[f]} Y^X$ must behave, we see that for any parameter $p \in P$, $\mathsf{curry}[f](p)$ is the function from X to Y given by the rule $x \mapsto f(p, x)$. So for any $p \in P$, $\mathsf{curry}[f](p)$ is the function such that for any $x \in X$, $\mathsf{curry}[f](p)(x) = f(p, x)$.

λ Notation

In defining $\mathsf{curry}[f]$, we needed to describe certain elements of Y^X. But elements of Y^X are functions. And a function typically is described by a rule. So it would be convienent to have a notation that permits us to describe the behavior of a function without giving the function a name. The logician Alonzo Church [**?**] was interested in the fundamental idea of just what *is* a function. He proposed a notation for describing functions. He wrote things like $\lambda x.x^2$ to describe the function that squares its input. The Greek letter λ means nothing. It is used only as a marker to introduce a function. This "λ" notation is widely adopted in computer science. Indeed, it appears even in languages such as Python. We could make the λ notation formal (as did Church), but for our purposes informality is enough. We use this notation to describe elements of Y^X. Several examples will help to explain this.

Example 7.1

- For $X \xrightarrow{f} Y$ and $Y \xrightarrow{g} Z$, the composite function $g \circ f \in Z^X$ is $\lambda x.g(f(x))$.
- The element of $\mathbb{N}^{\mathbb{N}}$ defined by $\lambda x.x^{\curvearrowright}$ is the successor function.
- For any $a \in X$, the function $\hat{a} \in X^{\mathbb{1}}$ is $\lambda x.a$.
- For $X \xleftarrow{f} T \xrightarrow{g} Y$, the function $\langle f, g\rangle X \times Y^T$ is $\lambda x.(f(x), g(x))$.
- For a parametric function $P \times X \xrightarrow{f} Y$, the function $P \xrightarrow{\mathsf{curry}[f]} Y^X$ can be defined by the rule $p \mapsto \lambda x.f(p, x)$.
- For any $f \in Y^X$, $f = \lambda x.f(x) = \lambda y.f(y)$.
- The only element of $\mathbb{1}^X$ is $\lambda x.\bullet$.

Exercises for Lecture 7

20. For set $A = \{1, 2, 3\}$ and $B = \{a, b\}$

    a) draw internal diagrams corresponding to each element of B^A (there are eight of them);

    b) draw internal diagrams corresponding to each element of A^B (there are nine of them).

21. If X is a finite set with k elements, Y is a finite set with j elements, how many elements are there in the set Y^X?

22. Consider the function $\mathbb{N} \times \mathbb{N} \xrightarrow{f} \mathbb{N}$ defined by $f(m, n) = m^n$. What element of $\mathbb{N}^{\mathbb{N}}$ is $\mathsf{curry}[f](3)$? Use λ notation to describe it.

23. For the function min from $\mathbb{R} \times \mathbb{R}$ to $\mathbb{R}$ defined to mean the minimum of x and y, define $\mathsf{curry}[\min]$.

24. Use λ notation to describe the element of $\mathbb{N}^{\mathbb{N}}$ that quadruples the square of the input.

7.6 Relations and Function Graphs

Suppose S is a set modelling the students at Chapman and M is a et modelling the academic majors the university offers: Mathematics, Philosophy, HeadScratching, and so on. Then elements of S (students) can be related to elements of M (majors) by 'is majoring in' as in "Jethro is majoring in Biology." Because a student might have a double major, we can not model the situation as a function, at least not in the most obvious way. Instead, we introduce the notion of a *(binary) relation*. The same idea, generalized to higher dimensional relations, is at the heart of what we call *relational databases*. We will investigate the general concept of a *relation* in more detail later. For now, we need them primarily as they relate to functions.

Definition 7.9

A *binary relation from* X *to* Y is subset $R \subseteq X \times Y$. A *binary relation on* X is a relation from X to X. Since we will only be concerned with binary relations, from now on we refer them simply as *relations*

For a relation R from X to Y, we will say "x is R-related to y" and write $R(x, y)$ when $(x, y) \in R$. In many situations, we use "infix" notation, writing $x \; R \; y$ instead of $R(x, y)$.

A relation $R \subseteq X \times Y$ is

- *total* if for every $x \in X$, there is at least one $y \in Y$ for which $x \; R \; y$;
- *deterministic* if for every $x \in X$, there is at most one $y \in Y$ for which $x \; R \; y$;
- *functional* if it is total and deterministic.

Example 7.2

The relation of "less than or equal to" $\leq$ on real numbers can be regarded as a subset $\leq \subseteq \mathbb{R} \times \mathbb{R}$ defined by $(x, y) \in \leq$ when x is actually less than or equal to y and $(x, y) \notin \leq$ otherwise. This is a good example of why we prefer to write $x \leq y$ instead of $\leq(x, y)$ or $(x, y) \in \leq(x, y)$. Note that $\leq$ is a total relation, because for any $x \in \mathbb{R}$ there is a $y \in \mathbb{R}$ for which $x \leq y$.

The relation $=$ on any set is functional because, for any $x \in X$ there is exactly one $y \in X$ so that $x = y$.

Define the relation S on $\mathbb{R}$ by $x \; S \; y$ if and only if $x = y^2$. Then S is not deterministic because $1 \; S \; 1$ and $1 \; S \; -1$. It is not total because there is no y for which $-1 \; S \; y$.

Definition 7.10

A function $f \colon X \to Y$ determines a relation called the *graph of* f defined by $\Gamma_f = \{(x, y) \in X \times Y \mid f(x) = y\}$.

Note: Γ_f is an equalizer $f \circ \pi_0$ and π_1.

Lemma 7.1

For any function $f\colon Y \longrightarrow X$, the graph Γ_f is functional.

Proof: This is pretty obvious from the basic properties of functions. That is, for each $x \in X$, $f(x) \in Y$ and obviously $f(x) = f(x)$. So Γ_f is total. On the other hand if $f(x) = y$ and $f(x) = y'$, then $y = y'$. So Γ_f is deterministic. □

Suppose we have a functional relation $R \subseteq X \times Y$. Then it is reasonable to suppose it actually determines a function.

Principle 7.8

Suppose $R \subseteq X \times Y$ is a functional relation. Then there is a function $F_R\colon X \longrightarrow Y$ so that $\Gamma_{F_R} = R$.

It is easy enough to check that $F_{\Gamma_f} = f$ for any function f. That is, $F_{\Gamma_f}(x) = y$ if and only if $x\ \Gamma_f\ y$ if and only if $f(x) = y$. So functions from $X \xrightarrow{f} Y$ correspond exactly to functional relations from $R \subseteq X \times Y$.

Like functions, relations allow a kind of composition,

Definition 7.11

Suppose $R \subseteq X \times Y$ and $S \subseteq Y \times Z$ are relations. Define $R; S \subseteq X \times Z$ by $x\ R; S\ z$ if and only if there is some $y \in Y$ so that $x\ R\ y$ and $y\ S\ z$.

For set X, let $\Delta_X \subseteq X \times X$ be defined as Γ_{id_X}. That is, $x\ \Delta_X\ y$ if and only if $x = y$.

Exercises for Lecture 7

25. Define $T \subseteq \mathbb{R} \times \mathbb{R}$ by $x\ T\ y$ if and only if $\tan x = y$. Is T deterministic? Is it total?

26. Show that for any relations $R \subseteq X \times Y$, $\Delta_X; R = R = R; \Delta_Y$.

27. Show that for any relations $R \subseteq W \times X$, $S \subseteq X \times Y$ and $T \subseteq Y \times Z$, $R; (S; T) = (R; S); T$.

28. Show that for any functions $X \xrightarrow{f} Y$ and $Y \xrightarrow{g} Z$, $\Gamma_f; \Gamma_g = \Gamma_{g \circ f}$. Show that for any set X, $\Gamma_{\mathsf{id}_X} = \Delta_X$.

8

The Set of Natural Numbers

Goals

Lecture:

- Re-introduce the natural numbers as a set
- Introduce sequences and recursively defined sequences
- Relate recursion to proofs by induction

Study:

- Be able to define simple functions by recursion
- Be able to explain how induction and recursion are related

We have used $\mathbb{N}$ informally to denote the set of natural numbers. It is time that we make the structure of $\mathbb{N}$ explicit within our theory of sets and functions. It will turn out that $\mathbb{N}$ is also a universal construction.

Natural numbers provide a precise picture of counting and of putting things in an order: first, second, third, and so on. Now that we have sets and functions we can consider a function $a\colon \mathbb{N} \to A$ to be an *infinite sequence*: $a(0)$, $a(1)$, $a(2)$, ... When we do that, we sometimes write a_0, a_1, a_2, ... instead. Still a itself is just function from $\mathbb{N}$ to A. To emphasize the notation that a represents an infinite sequence, we sometimes also write $(a_i)_{i \in I}$.

Much of what we discuss in this lecture has about it the feel of computer programming. This is partly because natural numbers are the main objects of calculation. We want to understand, for example, how to define a function like $n \mapsto n!$ (n factorial) as a function from $\mathbb{N}$ to $\mathbb{N}$ by specifying how it behaves. In particular, $0! = 1$ and $(n^{\frown})! = n^{\frown} \cdot n!$ characterize factorial by spelling out how to calculate it. For example,

$$\begin{aligned} 4! &= 4 \cdot 3! \\ &= 4 \cdot (3 \cdot 2!) \\ &= 4 \cdot (3 \cdot (2 \cdot 1!)) \\ &= 4 \cdot (3 \cdot (2 \cdot (1 \cdot 0!))) \\ &= 4 \cdot (3 \cdot (2 \cdot (1 \cdot 1))) && = 24 \end{aligned}$$

It is quite common to think about a sequence in which a_{n+1} is functionally related to a_n. For example, in the sequence $1, 2, 4, 8, \ldots$, the initial entry is 1 and each successive entry is double its predecessor.

Indeed, if we know just those two facts – the initial entry is 1, and each subsequent entry is double its predecessor – then we know how the entire sequence behaves. Also, we know how to calculate the n^{th} entry, by recursion just like factorial.

The most basic sequence, of course, is 0, 1, 2, ... Its initial entry is 0 and each subsequent entry is the successor of its predecessor. So we think of the sequence that comprises $\mathbb{N}$ as a universal recursively defined sequence.

8.1 Sequences and Simple Recurrences

Let us make the informal word *sequence* official.

Definition 8.1

A *sequence in set* X is a function $a \colon \mathbb{N} \to X$.

As we studied in previous lectures, the basic vocabulary of natural numbers is that (i) there is a starting natural number, 0, and (ii) for each natural number n there is a next one, $n^\frown$. To discuss successor in the language of sets and functions, we stipulate that successor is a function $\mathsf{suc} \colon \mathbb{N} \to \mathbb{N}$ given by the rule $n \mapsto n^\frown$. So $\mathbb{N}$ is not just a set. It comes with functions $\mathbb{1} \xrightarrow{\hat{0}} \mathbb{N} \xleftarrow{\mathsf{suc}} \mathbb{N}$.

Suppose $\mathbb{1} \xrightarrow{\hat{b}} X \xleftarrow{r} X$ is a similar structure. Then we ought to be able to define a a sequence in X, so that $a_0 = b$, $a_1 = r(b)$, $a_2 = r(r(b))$, and so on. In general, a_k should be determined by starting with b and repeatedly applying r a total of k times.

Definition 8.2

A *simple recurrence* is a set with functions $\mathbb{1} \xrightarrow{\hat{b}} X \xleftarrow{r} X$. We will say the two functions $\hat{b}$ and r form a *recurrence on* X.

A *natural number set* is a simple recurrence $\mathbb{1} \xrightarrow{\hat{z}} N \xleftarrow{s} N$ so that for any other simple recurrence $\mathbb{1} \xrightarrow{\hat{b}} X \xleftarrow{r} X$, there is exactly one function $N \xrightarrow{f} X$ so that $f \circ \hat{z} = \hat{b}$ and $f \circ s = r \circ f$.

The principle we are interested in here is that simple recurrences determine sequences.

Principle 8.1

The collection of natural numbers is a set, denoted by $\mathbb{N}$. Moreover, $\mathbb{1} \xrightarrow{\hat{0}} \mathbb{N} \xleftarrow{\mathsf{suc}} \mathbb{N}$. makes $\mathbb{N}$ a natural numbers set.

From a simple recurrence $\mathbb{1} \xrightarrow{\hat{b}} X \xleftarrow{r} X$, the corresponding unique sequence in X may be denoted by $\mathsf{s\text{-}rec}[b, r]$. So $\mathbb{N} \xrightarrow{\mathsf{s\text{-}rec}[b,r]} X$ is characterized by

$$\mathsf{s\text{-}rec}[b, r]_0 = b$$
$$\mathsf{s\text{-}rec}[b, r]_{n^\frown} = r(\mathsf{s\text{-}rec}[b, r]_n)$$

So every simple recurrence on a set X determines a sequence in X. On the other hand, it is not the case that every sequence is determined by a simple recurrence. Take for example, the sequence $0, 1, 0, 2, 0, 3, \ldots$. This can not be defined (at least not directly) by giving an initial entry and specifying successive entries based only on the predecessors. After all, the entries 1, 2, 3 and so on all have the same preceding entry.

8.2 Primitive Recursion

Evidently, addition, multiplication, factorial, and other familiar functions should be definable using Principle 8.1. But there are problems to overcome: Addition is not a sequence, at least not in an obvious way. And factorial is not obviously definable by a simple recurrence because we would need a function $r \colon \mathbb{N} \to \mathbb{N}$ so that $n^\frown ! = r(n!)$ for all n. If, in place of r, we could use a function that depends on n as well as on $n!$, we could define factorial recursively the usual way.

Putting things together, we consider a scheme that generalizes simple recursion to permit (i) dependence on a parameter not directly involved in the recursion, and (ii) dependence on n at each stage of the recursion.

Definition 8.3

A *primitive recurrence in* A *(with parameters in* C*)* consists of two functions $C \xrightarrow{b} A \xleftarrow{r} \mathbb{N} \times C \times A$.

A *parametric sequence in* A *with parameters in* C is a function $a\colon C \times \mathbb{N} \to A$. For a parametric sequence, we may write $a_{c,n}$ instead of $a(c, n)$.

Lemma 8.1

For any primitive recurrence $C \xrightarrow{b} A \xleftarrow{r} C \times \mathbb{N} \times A$, there is a unique function $\textsf{p-rec}[b, r]\colon C \times \mathbb{N} \to A$ satisfying:

$$\begin{aligned} \textsf{p-rec}[b, r]_{c,0} &= b(c) \\ \textsf{p-rec}[b, r]_{c,k^\frown} &= r(c, k, \textsf{p-rec}[b, r]_{c,k}) \end{aligned}$$

Proof: The proof of this is intricate enough that we do not give all the details here.

The idea of the proof is to use simple recursion to define a sequence of binary relations R_0, R_1, $\ldots$, from $C \times \mathbb{N}$ to A. In other words, R is a function from $\mathbb{N}$ to $2^{(C \times \mathbb{N}) \times A}$. These relations will satisfy the following conditions for each $i \in \mathbb{N}$:

- $R_i \subseteq R_{i^\frown}$;
- R_i is deterministic;
- R_i is total on $\{0, \ldots, i-1\}$

Then we define R to be the relation $(c, n)\ R\ a$ if and only if $(c, n) R_i a$ holds for some $i \in \mathbb{N}$. It follows from the above conditions that R is functional. So there is a function $f\colon C \times \mathbb{N} \to A$ for which $f(c, n) = a$ if and only if $(c, n)\ R\ a$. By the construction of R, it will follow that f satisfies

$$\begin{aligned} f_{c,0} &= b(c) \\ f_{c,k^\frown} &= r(c, k, f_{c,k}) \end{aligned}$$

Define $R_0 = \emptyset$. This is clearly deterministic and since $C \times \emptyset = \emptyset$, it is total on that $\emptyset$. Define $p\colon 2^{(C \times \mathbb{N}) \times A} \to 2^{(C \times \mathbb{N}) \times A}$ by the following rules (writing p_R for $p(R)$ to eliminate some confusing parentheses:

- $(c, 0)\ p_R\ b(c)$
- $(c, k^\frown)\ p_R\ x$ if and only if either
  - $(c, k^\frown)\ R\ x$ or
  - for some y, $(c, k)\ R\ y$ and $r(c, k, y) = x$.

Now we check that (i) if $R \subseteq S$, then $p_R \subseteq p_S$, (ii) if R is deterministic then p_R is deterministic, and (iii) if R is total on $C \times \{0, \ldots, k-1\}$, then p_R is total on $C \times \{0, \ldots, k\}$. Consequently, defining R_i by the simple recurrence $\textsf{s-rec}[R_0, p]$, we can prove by induction that the conditions stated in the second paragraph hold. Finally, by the way we constructed R, $f_{c,o} = b(c)$ and by induction on k, $f_{c,k^\frown} = r(c, k, f_{c,k})$. $\square$

Example 8.1

The "predecessor" function is defined by the scheme $\mathsf{pred}(0) = 0$ and $\mathsf{pred}(n^\frown) = n$. The primitive recurrence $\mathbb{1} \xrightarrow{\hat{0}} \mathbb{N} \xleftarrow{\pi_1} \mathbb{1} \times \mathbb{N} \times \mathbb{N}$ where π_1 is the projection $(x, y, z) \mapsto y$ determines a function $g\colon \mathbb{1} \times \mathbb{N} \to \mathbb{N}$ given by $g_{\bullet,0} = 0$ and $g_{\bullet,n^\frown} = \pi_2(\bullet, n, g_{\bullet,n}) = n$. Hence, we may define $\mathsf{pred}(n) = g_{\bullet,n}$

Exercises for Lecture 8

1. Define addition $\mathbb{N} \times \mathbb{N} \xrightarrow{+} \mathbb{N}$ by primitive recursion. That is, find functions $\mathbb{N} \xrightarrow{b} \mathbb{N}$ and and $\mathbb{N} \times \mathbb{N} \times \mathbb{N} \xrightarrow{r} \mathbb{N}$ so that

$$m + 0 = b(m)$$
$$m + n^\frown = r(m, n, m + n)$$

   That way, $\mathsf{p\text{-}rec}[b, r]$ is addition.

2. Define multiplication $\mathbb{N} \times \mathbb{N} \xrightarrow{\cdot} \mathbb{N}$ by primitive recursion. You may use addition in defining the primitive recurrence.

3. Define the factorial function by primitive recursion. You may use multiplication in defining the primitive recurrence.

4. For a given function $f\colon \mathbb{N} \to \mathbb{N}$, find a primitive recurrence that defines the function $n \mapsto \sum_{i=0}^{n-1} f(i)$. That is, result will be the sequence $0, f(0), f(0) + f(1), f(0) + f(1) + f(2), \ldots$.

5. The operation of *monus* $m \mathbin{\dot{-}} n$ is defined to be $m - n$ when $m \geq n$ and to be 0 otherwise. Define monus by primitive recursion.

6. Let $\mathbb{N} \xrightarrow{p} \mathbb{2}$ be a characteristic function. Show that bounded existential quantification is primitive recursive. That is, define the function $\exists_p^<\colon \mathbb{N} \to \mathbb{2}$ by $\exists_<^p(n) = \top$ iff there is at least one $k < n$ for which $p(k) = \top$. Show that this can be defined by a primitive recurrence. [Hint: $\exists_<^p(0) = \bot$ because there are no natural numbers below 0. Now ask what value is $\exists_<^p(n^\frown)$ in terms of p and $\exists_<^p(n)$.]

8.3 Primitive Recursion and "for" loops

It is a theorem (that we can not prove here) that there is a technical sense in which primitive recursion is implementable, say in Python, by nested "for" loops. To get a taste of what this means, let us pretend that $P \xrightarrow{b} X$ and $r\colon P \times \mathbb{N} \times X \to X$ are somehow defined by Python functions. Then the following code implements $\mathsf{p\text{-}rec}[b, r]$.

```
def f(p,n):
    result = b(p)
    for i in range(n):
        result = r(p,i,result)
    return result
```

Conversely, suppose a function in Python is defined using only natural number variables and the following parts of Python:

- Assignment statements
- Increment statements: "n+=1"

- loops of the form "**for** i **in** range(n): …"
- evaluations of functions that are defined similarly

The theorem we allude to claims that such a function is definable by primitive recursion. The theorem is not especially difficult, but it does involve a lot of careful checking. For us, the point is that primitive recursion allows us to define a lot of the functions that we expect to be able to program in a standard programming language.

A question arises, however. If "for" loops are sufficient to program any primitive recursive function, why bother with other more complicated loops? One answer is that programming languages are not merely for computing functions. They are used to implement lots of behavior that is not so easily cast in terms of sets and functinos. Another answer, though, is internal to set theory. Namely, there are recursive functions on $\mathbb{N}$ which are not primitive recursive.

For each $n \in \mathbb{N}$, define the sets $P^n \subseteq \mathbb{N}^{\mathbb{N}^n}$ of n-*ary number theoretic primitive recursive functions* to be the smallest sets satisfying:

- $\hat{0} \in P^0$
- $\mathsf{suc} \in P^1$
- for each $k < n$, $\pi^n_k \in P^n$ where $\pi^n_k \in \mathbb{N}^{\mathbb{N}^n}$ is defined by $\pi^n_k(x_0, \ldots, x_{n-1}) = x_k$
- If $g \in P^n$ and for each $k < n$, $f_i \in P^m$, then $g \circ \langle f_0, \ldots, f_{n-1} \rangle \in P^m$
- if $b \in P^m$ and $h \in P^{m+2}$, then $\mathsf{p\text{-}rec}[b, h] \in P^{m+1}$

Consider the following sequence $A_0, A_1, \ldots$ in $\mathbb{N}^{\mathbb{N}}$ defined by

$$A_0 \mathsf{suc}$$
$$A_{k^\frown} = \mathsf{p\text{-}rec}[A_k\widehat{(k)}, A_k \circ \pi^2_1]$$

Putting this in terms of elements

$$\begin{aligned} A_0(m) &= m + 1 \\ A_{k^\frown}(0) &= A_k(k) \\ A_{k^\frown}(m^\frown) &= A_k(A_{k^\frown}(m)). \end{aligned}$$

Evidently, each individual function A_k is a number theoretic primitive recursive unary function. But the function $\mathsf{Ack}\colon \mathbb{N} \to \mathbb{N}$ defined by $\mathsf{Ack}(n) = A_n(n)$ is not. The proof is quite ingeneous. Roughly, one shows that Ack grows faster than any number theoretic primitive recursive function.

To get an idea of how fast this function grows, $\mathsf{Ack}(0) = 1$, $\mathsf{Ack}(1) = 3$, $\mathsf{Ack}(3) = 61$, $\mathsf{Ack}(4) = 2^{2^{65536}} - 3$ a number vastly larger than the number of electrons in the visible universe.

8.4 Lists

For a set A, the lists consisting of elements from A also form a set. The structure of this set is similar to $\mathbb{N}$.

Definition 8.4

For a set A, a *simple* A-*recurrrence* is a set with two functions $\mathbb{1} \xrightarrow{\hat{b}} X \xleftarrow{r} A \times X$.

For a set A, an A-*list set* is a simple A-recurrence $\mathbb{1} \xrightarrow{\hat{e}} L \xleftarrow{c} A \times L$ so that for any simple A-recurrence $\mathbb{1} \xrightarrow{\hat{b}} X \xleftarrow{r} A \times X$ there is exactly one function $h\colon L \to X$ so that

$$\begin{aligned} h \circ \hat{e} &= \hat{b} \\ h \circ c &= r \circ (\mathsf{id}_A \times h). \end{aligned}$$

Principle 8.2

The collection of lists with items drawn from A is a set, denote by $\mathsf{List}[A]$. The functions $\mathbb{1} \xrightarrow{\widehat{[]}} \mathsf{List}[A] \xleftarrow{:} A \times \mathsf{List}[A]$ constitute an A-list set, where the function $A \times \mathsf{List}[A] \xrightarrow{\mathtt{mathord:}} \mathsf{List}[A]$ is the function given by the rule $(a, L) \mapsto a : L$.

For $\mathbb{1} \xrightarrow{\hat{b}} X \xleftarrow{r} A \times X$, we reuse our notation for recursive functions defined from $\mathbb{N}$ and write $\mathsf{s\text{-}rec}[\hat{b}, r]$ for the unique function satisfying

$$\mathsf{s\text{-}rec}[\hat{b}, r]([]) = b$$
$$\mathsf{s\text{-}rec}[\hat{b}, r](a : L) = r(a, \mathsf{s\text{-}rec}[\hat{b}, r](L))$$

Example 8.2

$\mathbb{1} \xrightarrow{\hat{0}} \mathbb{N} \xleftarrow{+} \mathbb{N} \times \mathbb{N}$ determine a function $\mathsf{s\text{-}rec}[\hat{0}, +]$ from $\mathsf{List}(\mathbb{N})$ to $\mathbb{N}$. The function satifies $\mathsf{s\text{-}rec}[\hat{0}, +]([]) = 0$ and $\mathsf{s\text{-}rec}[\hat{0}, +](n : L) = n + \mathsf{s\text{-}rec}[\hat{0}, +](L)$. So $\mathsf{s\text{-}rec}[0, +]$ returns the sum of items in a list natural numbers. Earlier, we wrote this as $\sum L$. In other words, we *defined* $\Sigma = \mathsf{s\text{-}rec}[\hat{0}, +]$.

Exercises for Lecture 8

7. For a function $f\colon X \to Y$, specify a simple X recurrence that defines the function

$$\mathsf{map}[f]\colon \mathsf{List}[X] \to \mathsf{List}[Y]$$

so that
$$\mathsf{map}[f]([a_0, a_1, \ldots, a_{n-1}]) = [f(a_0), f(a_1), \ldots, f(a_{n-1})]$$

8. For $\mathsf{map}[\cdot]$ as defined in the previous exercise, show that for any $f\colon X \to Y$ and $g\colon Y \to Z$, $\mathsf{map}[g \circ f] = \mathsf{map}[g] \circ \mathsf{map}[f]$.

9. Define concatenation in $\mathsf{List}[A]$ by a simple A recurrence. [Hint: I am asking for a function from $\mathsf{List}[A] \times \mathsf{List}[A]$ to $\mathsf{List}[A]$, but simple A recurrence alone will not do the job, because that can only define a function $\mathsf{List}[A] \to X$ for some X. Do this instead: (i) specify a simple A-recurrence to define a function $c\colon \mathsf{List}[A] \to \mathsf{List}[A]^{\mathsf{List}[A]}$ for which $c(M)(L)$ is the concatetation of L followed by M, (ii) define $L \star M$ to be $c(M)(L)$.

10. Write a scheme for "primitive A recursion", specified by functions $P \xrightarrow{b} X \xleftarrow{r} P \times \mathsf{List}[A] \times X$. Write the equations involving b and r that should determine a unique function $P \times \mathsf{List}[A] \to X$. You do not need to prove that your scheme actually defines a function.

9

Powersets

Goals

The exponential 2^X for a set X is the set of all characteristic maps on X. Each $k \in 2^X$ corresponds to a subset of X by $k^{-1}(\top) = \{x \in X \mid k(x) = \top\}$. Vice versa, a subset $A \subseteq X$ determines a characteristic function $\kappa_A \colon X \to 2$ by the rule

$$x \mapsto \begin{cases} \top & \text{if } x \in A \\ \bot & \text{otherwise} \end{cases}$$

So 2^X is, in this sense, *representative* of the collection of all subsets of X. It is convenient also to suppose that the actual collection of subsets of a set X forms a set. This should "behave" exactly like 2^X, but concretely, consist of subsets rather than characteristic functions.

Principle 9.1

For any set X, the collection of subsets of X is a set, denoted by $\mathcal{P}(X)$, called the *power set of* X. Moreover, there is a function $\ni_X \colon \mathcal{P}(X) \times X \to 2$ defined by

$$\ni_X(A, x) = \begin{cases} \top, & \text{if } x \in A \\ \bot & \text{otherwise} \end{cases}$$

The function $\ni_X$ is an evaluation map, meaning that for any function $f \colon P \times X \to 2$, there is a unique function $f^\dagger \colon P \to \mathcal{P}(X)$ for which $f = \ni_X \circ (f^\dagger \times \mathsf{id}_X)$. The function $f^\dagger$ is given by the rule $p \mapsto \{x \in X \mid f(p, x) = \top\}$.

For a function $f \colon X \to Y$, $\ni_Y \circ (\mathsf{id}_{\mathcal{P}(X)} \times f)$ is a function from $\mathcal{P}(Y) \times X$ to 2. So there is a unique function from $\mathcal{P}(Y)$ to $\mathcal{P}(X)$ determined by f according to the above princple. This is called *inverse image*.

Definition 9.1

For $X \xrightarrow{f} Y$, let $\mathcal{P}(Y) \xrightarrow{f^-} \mathcal{P}(X)$ denote the unique function for which $\ni_B \circ (\mathsf{id}_{\mathcal{P}(B)} \times f) = \ni_A \circ (f^- \times \mathsf{id}_A)$. In terms of elements, f^- satsifies

$$x \in f^-(B) \text{if and only if } f(x) \in B$$

for every $B \subseteq Y$. So f^- is given by the rule $B \mapsto \{x \in X \mid f(x) \in B\}$

This notation clashes slightly with our earlier definition of inverse image of an element, $f^-(b) = \{x \in X \mid f(x) = b\}$. But this is harmless because $f^-(b)$ in the earlier usage is the same as $f^-(\{b\})$ in the new usage.

Exercises for Lecture 9

1. For $f\colon \mathbb{N} \to \mathbb{N}$ defined by $f(n) = x^2$, what is $f^{-}(\{2,3,4,5,6,7,8\})$?
2. For $\sin\colon \mathbb{R} \to \mathbb{R}$, what is $\sin^{-}(\{-1,1\})$?
3. Show that for any two functions $f\colon X \to Y$ and $g\colon Y \to Z$, $(g \circ f)^{-} = f^{-} \circ g^{-}$.

9.1 Intersections, Unions, Residuals, Differences and Negations

Suppose we are currently interested in a particular set, which we will consider to be the "universe of discourse." For example, for now we might only be interested in natural numbers. So our universe of discourse is $\mathbb{N}$. Or we might be interested in poker. So our universe of discourse is Deck, the set we defined earlier in these lectures. For this discussion, let us refer to this set as U (for "universe"). Then the structure of $\mathcal{P}(U)$ is of particular interest.

Suppose A and B are subsets of U. Then it makes sense to consider the elements that A and B have in common. For example, for $A \subseteq \mathbb{N}$ being the set of even natural numbers and $B \subseteq \mathbb{N}$ the set of perfect squares, we might want to concentrate on the set of even, perfect squares, which is another subset of $\mathbb{N}$. In general, for $A \subseteq U$ and $B \subseteq U$, the elements in common constitute another subset of U. This is called the *intersection* and is denoted by $A \cap B$.

Likewise, we might consider merging A and B into a single set (in our example, the set of numbers that are either even or perfect squares). This is called the *union*. Related to intersection, there is a largest set C so that $C \cap A \subseteq B$. This is called the *residual*, and is denoted by $A \Rightarrow B$. Flipping this the other way, there is also a smallest set C so that $A \subseteq B \cup C$. This is called the *set difference* and is denoted by $A \setminus B$.

These operations on subsets of X are closely related to the logic of propositions. Imagine that U consists of a "universe" of possible worlds. Then subsets of U are collections of worlds where certain things are true. For example, perhaps $P \subseteq U$ is the set of worlds in which pigs fly; $K \subseteq X$ is the set of worlds in which kittens smoke cigars. So $P \cap K$ is the set of worlds in which pigs fly *and* kittens smoke cigars. Likewise, $P \cup K$ is the set of worlds in which *Either* pigs fly *or* kittens smoke cigars. And $P \Rightarrow K$ is the set of worlds in which it is true that if pigs fly, *then* kittens smoke cigars.

Understanding the interaction of $\cap$, $\cup$ and $\Rightarrow$ is closely related to the logic of "and", "or" and "implies". If we add a sentence "False" that is never true and another one "True" that is always true, then the logic of "and", "or" and "implies" form what is called a *Heyting algebra*.

In fact, "implies" interacts with "False" in a useful way. It turns out that "P implies False" is essentially the same as saying "P is not true". And if "P is not true" is not true, then "P" must be true. This oservation indicates that the Heyting algebra of subsets is actually a *Boolean algebra*.

The operation of set difference is not as familiar in a logical setting. It corresponds to"but not", so $P \setminus K$ is the set of worlds in which pigs fly, but kittens do not smoke cigars.

Lemma 9.1

In the following, let U be a set, and $A, B \subseteq X$ be subsets.

- There is a subset of U, denoted by $A \cap B$, so that for every $C \subseteq U$, $C \subseteq A \cap B$ if and only if $C \subseteq A$ and $C \subseteq B$.
- There is a subset of U, denoted by $A \cup B$, so that for every $C \subseteq U$, $A \cup B \subseteq C$ if and only if $A \subseteq C$ and $B \subseteq C$.
- There is a subset of U, denoted by $A \Rightarrow B$, so that for every $C \subseteq U$, $C \subseteq A \Rightarrow B$ if and only if $C \cap A \subseteq B$.

- There is a subset of U, denoted by $A \setminus B$, so that for every $C \subseteq U$, $A \setminus B \subseteq C$ if and only if $A \subseteq B \cup C$.

Proof: A and B are determined by characteristic functions $\kappa_A \colon X \to 2$ and $\kappa_B \colon X \to 2$. So $\rangle \kappa_A, \kappa_B \langle$ is a function from X to 2×2. If we compose with a function $h \colon 2 \times 2 \to 2$, we have another characteristic function on X. So this determines another subset of X. So all of the above constructions amount to defining suitable functions $2 \times 2 \to 2$.

The four functions corresponding to the subset operations can be given by tables. In these tables, we read the first argument on the left, and the second argument on the top.

$\wedge$	$\bot$	$\top$
$\bot$	$\bot$	$\bot$
$\top$	$\bot$	$\top$

$\vee$	$\bot$	$\top$
$\bot$	$\bot$	$\top$
$\top$	$\top$	$\top$

$\to$	$\bot$	$\top$
$\bot$	$\top$	$\top$
$\top$	$\bot$	$\top$

$-$	$\bot$	$\top$
$\bot$	$\bot$	$\bot$
$\top$	$\top$	$\bot$

So for each $h \in \{\wedge, \vee, \to, -\}$, there is a subset $(h \circ \langle \kappa_A, \kappa_B \rangle)^{-}(\top)$.

It is routine to check that for $h = \wedge$, the result is $A \cap B$; for $h = \vee$, the result is $A \cup B$; for $h = \to$, the result is $A \Rightarrow B$; and for $h = -$, the result is $A \setminus B$. □

This lemma, together with the fact that $\emptyset$ is the smallest element of $\mathcal{P}(X)$ and X is the largest, can be summarized by saying that $\cap$, $\cup$, $\Rightarrow$, $\emptyset$ and X make $\mathcal{P}(X)$ into what is known as a *Heyting algebra*. We spell out the axioms for Heyting algebras in the next section, but generally speaking these are the structures that correspond to a sort of minimal version of propositional logic in which "and", "or", "implies", "true" and "false" interact in natural ways. Likewise, "or", "and", "but not", "false" and "true" interact in natural ways to determine a co-Heyting algebra.

Additionally, "implies" and "false" interact in a stronger way, as do "but not" and "true". In particular, $\mathcal{P}(U)$ is a *Boolean algebra*. Let us abbreviate $A \Rightarrow \emptyset$ by writing A^* (read this informally as "not A). Then the Law of Double Negation asserts that double negations do not change anything: $A^{**} = A$.

Lemma 9.2

In $\mathcal{P}(X)$, $A^{**} = A$.

Proof: Calculating the members of A^*, it is easy to check that for every element $x \in X$, either $x \in A$ or $x \in A^*$, but not both. So $x \in A^{**}$ if and only if $x \notin A^*$ if and only if $x \in A$.

Put differently, define the function $\neg \colon 2 \to 2$ by $\neg\top = \bot$ and $\neg\bot = \top$. Then A^* is defined by $(\neg \circ \kappa_A)^{-}(\top)$. Obviously, $\neg \circ \neg = \mathsf{id}_2$. As with the other operations, it is now routine to check that $A^{**} = A$. □

The Law of Double Negation is precisely the property that distinguishes Boolean algebras from general Heyting algebras. That is, one can define the term "Boolean algebra" to mean "Heyting algebra that satisfies the Law of Double Negation." So $\mathcal{P}(X)$ is indeed a Boolean algebra.

Exercises for Lecture 9

4. An easy consequence of Double Negation is that $A \Rightarrow \emptyset = U \setminus A$. Prove it.

9.2 Laws of Finitary Set Operations

As we mentioned, $\mathcal{P}(X)$ forms a Boolean algebra. This means that $\cup$, cup, $\Rightarrow$, $\emptyset$ and X satisfy various laws. We spell the most important out here.

Laws

For any set U and any subsets A, B and C:

Semilattice Laws

Associativity	$A \cap (B \cap C) = (A \cap B) \cap C$
	$A \cup (B \cup C) = (A \cup B) \cup C$
Commutativity	$A \cap B = B \cap A$
	$A \cup B = B \cup A$
Idempotency	$A \cap A = A$
	$A \cup A = A$

Lattice Laws

Absorptivity	$A = (A \cap B) \cup A$
	$A = (A \cup B) \cap A)$
Ordering	$A = B \cap A$ if and only if $A \cup B = A$

Bounded Lattice Laws

Identity	$A \subseteq A \cap U$
	$A \cup \emptyset \subseteq A$

Heyting Algebra Law

Residuation	$A \cap B \subseteq C$ if and only if $A \subseteq B \Rightarrow C$

co-Heyting Algebra Law

Co-Residuation	$A \setminus B \subseteq C$ if and only if $A \subseteq B \cup C$

Boolean Algebra Law

Double Negation	$A^{**} \subseteq A$

Distributive Lattice Laws

Distributivity	$A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$
	$(A \cup B) \cap (A \cup C) \subseteq A \cup (B \cap C)$

Other Boolean Laws

de Morgan's Laws	$(A \cap B)^* = A^* \cup B^*$
	$(A \cup B)^* = A^* \cap B^*$

Several remarks are in order.

- The Semilattice Laws describe how $\cap$ and $\cup$ behave without any interaction between the two. In fact, any binary operation that is associative, commutative and idempotent is called a *semilattice operation*.

- The Lattice Laws describe how $\cap$ and $\cup$ interact. The two Absorption Laws together are equivalent to the Ordering Law. For suppose $A = A \cup (A \cap B)$ holds for all A and B. Now suppose $X = Y \cap X$. Then $Y \cup X = Y \cup (X \cap Y) = Y$. Conversely, suppose $A = B \cap A$ implies $B \cup A = B$ for all A and B. Then $(X \cap Y) = (X \cap Y) \cap X$. So $X \cup (X \cap Y) = X$.

- The remaining laws are stated in terms of $\subseteq$ instead of equality. Becaause of the Ordering Laws, $A \subseteq B$ is equivalent to $A = B \cap A$ and also equivalent to $B = A \cup B$.

- The Bounded Lattice Laws indicate that U is the unit element for $\cap$ and $\emptyset$ is the unit element for $\cup$. It follows that $\emptyset$ is the smallest element of $\mathcal{P}(U)$ and U is the largest.
- The Residuation and Co-residuation Laws indicate that $A \Rightarrow B$ and $A \setminus B$ are defined as *duals* of one another.
- In the Double Negation Law recall that A^* is defined to be $A \Rightarrow \emptyset$. Since $A \cap \emptyset \subseteq A \Rightarrow \emptyset$, it follows from Residuation that $A \subseteq A^{**}$. So in fact, $A = A^{*}*$.
- If we defined $A^{\dagger} = U \setminus A$, then in any co-Heyting algebra, $A^{\dagger\dagger} \subseteq A$. In a Boolean algebra $A^{\dagger} = A^*$.
- With respect to Distributivity, in any lattice, the opposite inclusions hold: $A \cap B \subseteq A \cap (B \cup C)$ and $A \cap B \subseteq A \cap (B \cup C)$, so $(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$. Similarly, for the opposite inclusion for the second Distributive Law.
- The Distributivity laws are equivalent to each other. Suppose $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$ is true for all A, B and C. Then $(X \cup Y) \cap (X \cup Z) \subseteq ((X \cup Y) \cap X) \cup ((X \cup Y) \cap Z) = X \cup ((X \cup Y) \cap Z) \subseteq X \cup ((X \cap Z) \cup (Y \cap Z)) = X \cup (Y \cap Z)$
- The Heyting (or co-Heyting) Law implies Distributivity. Obviously, $A \cap B \subseteq (A \cap B) \cup (A \cap C)$, so $B \subseteq A \Rightarrow ((A \cap B) \cup (A \cap C))$. Likewise $C \subseteq A \Rightarrow ((A \cap B) \cup (A \cap C))$. So $B \cup C \subseteq A \Rightarrow ((A \cap B) \cup (A \cap C))$. And again using Residuation, $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$.

Exercises for Lecture 9

In the following, assume that U is a set, and all other sets are subsets of U.

5. Prove, using only the Semilattice Laws, the Absorption Laws and the Bounded Lattice Laws, that $A \cap \emptyset = \emptyset$. Likewise, show that $A \cup U = U$.
6. Prove, using only the Semilattice and Lattice Laws, that the two Distribution Laws are equivalent. That is, show that if $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$ holds for all A, B, C, then so does $(A \cup B) \cap (A \cup C) \subseteq A \cup (B \cap C)$. Hint: Let $X = A \cup B$, $Y = A$ and $Z = C$. So $(A \cup B) \cap (A \cup C) = X \cap (Y \cup Z)$. Now use the first distributivity law, followed by absorption, then a second use of distributivity, and association and finally a second use of absorption.
7. Prove, using only the Semilattice, Lattice and Heyting Algebra Laws, the Distributivity Law (either one).
8. Prove, using only the Semilattice, Lattice Laws, Heyting Algebra and Boolean Algebra Laws, that the two de Morgan's Laws hold.
9. Prove that $A \Rightarrow B = A^* \cup B$, using any method.
10. Prove that $A \setminus B = A \cap B^*$, using any method.

9.3 Quantifiers and Completeness

Consider how we might try to define the set of perfect square natural numbers. These are the natural numbers of the form m^2 for a natural number m. So the first few are $0, 1, 4, 9, 16$ and so on. We would be right to define this set by $\{n \in \mathbb{N} \mid \text{for some } m \in \mathbb{N}, m^2 = n\}$. We need to make sense of this, since "for some $m \in \mathbb{N}, \ldots$" does not look like anything we have encountered yet. Just as we abbreviated "and" with symbol $\bigwedge$ and "or" with $\bigvee$, we will abbreviate "for some $n \in \mathbb{N}, \ldots$" as $\exists n \in \mathbb{N}, \ldots$.

Suppose $P\colon X \to 2$ is a characteristic function on X. We will write $\{x \in X \mid P(x)\}$ instead of $\{x \in X \mid P(x) = \top\}$, mostly to avoid some clutter, but also because this allows for more informal descriptions. Using this notation, we can form subsets by writing things like $\{x \in X \mid P(x) \wedge Q(x)\}$, to denote the subset of elements satisfying two conditions (this is an intersection). Similarly, we can also write $\{x \in X \mid P(x) \vee Q(x)\}$ for a union, $\{x \in X \mid P(x) \to Q(x)\}$ for a residual, and $\{x \in X \mid \neg P(x)\}$ for a complement.

Definition 9.2

For sets W and X and subset $C \subseteq X$, write $\{(w, x) \in W \times X \mid x \in C\}$ for the subset of $W \times X$ defined by $pi_1^-(C)$.

Lemma 9.3

For sets W and X and subset $R \subseteq W \times X$, there is a subset of X, denoted by $\{x \in X \mid \exists w \in W.R(w, x)\}$ so that for all $C \subseteq X$,

$$\{x \in X \mid \exists w \in W.R(w, x)\} \subseteq C \text{ if and only if } R \subseteq \{(w, x) \in W \times X \mid x \in C\}.$$

Also, there is a subset of X, denoted by $\{x \in X \mid \forall w \in W.R(w, x)\}$ so that for all $P \colon X \to 2$,

$$C \subseteq \{x \in X \mid \forall w \in W.R(w, x)\} \text{ if and only if } \{(w, x) \in W \times X \mid x \in C\} \subseteq R.$$

Proof: The proof of this is fairly technical and not especially infomrative. A sketch will suffice.
The relation R determines a function r from X to $\mathcal{P}(W)$ by $x \mapsto \{w \in W \mid R(w, x)\}$. Then

$$\{x \in X \mid \forall w \in W.R(w, x)\} = r^-(\{W\})$$
$$\{x \in X \mid \exists w \in W.R(w, x)\} = r^-(\mathcal{P}(W) \setminus \{\emptyset\})$$

Proving that these sets satisfy the desired conditions is technical, but routine.

Concretely, $\{x \in X \mid \exists w \in W.R(w, x)\}$ consists of those $x \in X$ so that $R(w, x)$ for some $w \in W$; $\{x \in X \mid \forall w \in W.R(w, x)\}$ consists of those $x \in X$ so that $R(w, x)$ for all $w \in W$. This justifies our notation: $\exists w \in W, R(w, x)$ is read as “there *exists* $w \in W$ so that $R(w, x)$;” $\forall w \in W, R(w, x)$” is read as “for *all* $w \in W$, $R(w, x)$.” □

We can use this lemma to generalize $\cup$ and $\cap$ to arbitrary sets of subsets.

Definition 9.3

Let $A \colon I \to \mathcal{P}(X)$ be a function into the powerset of X. We may write A_i instead of $A(i)$ to emphasize that each A_i is a set. Then define

$$\bigcup_{i \in I} A_i = \{x \in X \mid \exists i \in I. x \in A_i\}$$

According to Lemma **??**, $\bigcup_{i \in I} A_i$ is again a subset of X, generalizing $\cup$ to the union of arbitrary families of subsets, rather than just two. In this sense, $\mathcal{P}(X)$ is *complete*. That is, the union of any family of subsets of X exists. It is this fact that justifes saying that $\mathcal{P}(X)$ is a *complete* Boolean algebra.

Exercises for Lecture 9

11. Show that $\bigcup_{i\in I} A_i \subseteq C$ if and only if $A_k \subseteq C$ holds every $k \in I$.

12. Define $\bigcap_{i\in I} A_i$ in analogy with $\bigcup_{i\in I} A_i$.

13. For $I = \emptyset$, what is $\bigcup_{i\in I} A_i$?

9.4 Atomicity

The complete Boolean algebras $\mathcal{P}(X)$ have one more feature that characterizes powersets among other complete Boolean algebras. They are *atomic*. This means, roughly, that all subsets are built from the simplest ones.

An *atom* of $\mathcal{P}(X)$ is a subset A with the property that $\emptyset \subseteq B \subseteq A$ implies $B = \emptyset$ or $B = A$. That is, there is nothing strictly between $\emptyset$ and A. Clearly, the singleton subsets of X are the atoms $\mathcal{P}(X)$.

Lemma 9.4

For any set X, the rule $x \mapsto \{x\}$ determines a function from X to $\mathcal{P}(X)$.

Proof: Define the *diagonal* subset of $X \times X$ as $\Delta_X = \{(x, y) \in X \times X \mid x = y\}$. This is the equalizer of the functions π_0 and π_1. Let $\delta_X \colon X \times X \to 2$ be the characteristic map for Δ_X. So $\delta_X(x, y) = \top$ if and only if $x = y$. Let $s\colon X \to \mathcal{P}(X)$ be the unique function for which $\in_X \circ (s \circ \mathsf{id}_X) = \delta_X$. In other words $\ni_X(s(x), y) = \delta_X(x, y)$. Since $\ni_X(s(x), y) = \top$ if and only if $y \in s(x)$, it is the case that $y \in s(x)$ if and only if $x = y$. So $s(x) = \{x\}$. □

Now every subset of X is obtained as a union of singletons: $A = \bigcup_{x\in A}\{x\}$. For a complete Boolean algebra, this is what is meant by saying that $\mathcal{P}(X)$ is a complete *atomic* Boolean algebra. Although we do not investigate this here, any complete atomic Boolean algebra has the same structure as $\mathcal{P}(X)$ for some X. The structure of $\mathcal{P}(X)$ is even preserved by inverse images.

Lemma 9.5

For any function $f\colon X \to Y$, any $B\colon I \to \mathcal{P}(Y)$, it is the case that $f^{-}(\bigcup_{i\in I} B_i) = \bigcup_{i\in I} f^{-1}(B_i)$ and $f^{-1}(\bigcap_{i\in I} B_i) = \bigcap_{i\in I} f^{-1}(B_i)$.

Proof: $x \in f^{-}(\bigcup_{i\in I} B_i)$ if and only if $f(x) \in \bigcup_{i\in I} B_i$ if and only if $f(x) \in B_k$ for some $k \in I$ if and only if $x \in f^{-}(B_k)$ for some $k \in I$ if and only if $x \in \bigcup_{i\in I} f^{-}(B_i)$. The proof for $\bigcap$ is similar with "for some ..." replaced by "for all ...". □

Exercises for Lecture 9

14. Write out $\mathcal{P}(\{a, b, c\})$

15. Write out $\mathcal{P}(\emptyset)$

16. Is it the case that $\emptyset \in \mathcal{P}(A)$ for any set A? Explain.

17. Write out $\mathcal{P}(2 \times 2)$ and $\mathcal{P}(\mathcal{P}(2))$. Pay attention to writing them in a systematic way, so that it is clear you have actually listed everything.

18. I claim that $\mathcal{P}(\emptyset)$ is a terminal set (Definition 7.2). Justify the claim.

19. I claim that $\mathcal{P}(\emptyset) \in \mathcal{P}(\mathcal{P}(\emptyset))$ is a subset classifier (Definition 7.6). Justify the claim.

9.5 Forward images

For a function $f \colon A \to B$ the function $f^- \colon \mathcal{P}(C) \to \mathcal{P}(B)$ has what is known as an *upper adjoint*.

Definition 9.4

For $f \colon X \to Y$, define $f^+ \colon \mathcal{P}(X) \to \mathcal{P}(Y)$ by the rule $A \mapsto \{y \in B \mid \exists x \in A . f(x) = y\}$. The subset $f^+(A) \subseteq Y$ is called the *forward image of* A *with respect to* f.

The important fact about f^+ is that is related to f^-.

Lemma 9.6

For any $A \subseteq X$ and $B \subseteq Y$, $f^+(A) \subseteq B$ if and only if $A \subseteq f^-(B)$.

Proof: Suppose $f^+(A) \subseteq B$. For $x \in A$, $f(x) \in f^+(A)$. So $f(x) \in B$. By definition, this means $x \in f^-(B)$. This shows that $A \subseteq f^-(B)$. Conversely, suppose $A \subseteq f^-(B)$. For $y \in f^+(A)$, there is some $x \in A$ so that $f(x) = y$. So there is some $x \in f^-(B)$ so that $f(x) = y$. Thus $y = f(x) \in B$. This shows that $f^+(A) \subseteq B$ □

There is also a lower adjoint of f^-, but as it is less commonly used, we do not investigate it here.

The important features of f^+ are easily checked. First, f^+ preserves atoms. That is, $f^+(\{x\}) = \{f(x)\}$. Second, f^- preserves all unions So $f^+(\bigcup_{i \in I} A_i) = \bigcup_{i \in I} f^+(A_i)$. Note that f does not necessarily preserve intersections.

Exercises for Lecture 9

20. Define a function $f \colon X \to Y$ and two subsets $A, B \subseteq X$ so that $f(X) \cap f(Y) \neq f(X \cap Y)$. Try to find the smallest example you can.

21. Prove that for any function $f \colon X \to Y$ and any $A \subseteq X$, the inclusion $A \subseteq f^-(f^+(A))$ holds.

22. Prove that for any function $f \colon X \to Y$ and any $B \subseteq Y$, the inclusion $f^+(f^-(B)) \subseteq B$ holds.

10

Additional Constructions

Goals

content

Many other constructions can be built up using the principles we have discussed. In this lecture, we consider some of the most useful and most general.

10.1 Unions

Although not strictly needed for most applications, mathematicians generally agree that sets, no matter how they are related, can be merged into a single set with elements taken form the originals. That is, union ($\bigcup$) is meaningful for any set of sets. We take this as an additional principle.

Principle 10.1

Suppose $\mathcal{X}$ is a set and each element of $\mathcal{X}$ is a set (so $\mathcal{X}$ is a set of sets). Then there is a set U so that $\mathcal{X} \subseteq \mathcal{P}(U)$ and for any sets U, if $\mathcal{X} \subseteq \mathcal{P}(Z)$, then $U \subseteq Z$.

Using this principle, the set U is actually the union $\bigcup_{X \in \mathcal{X}} X$. That is, $x \in \bigcup_{X \in \mathcal{X}} X$ holds if any only if $x \in X$ for some $X \in \mathcal{X}$.

For a sets of sets $\{A, B\}$, we write $A \cup B$ instead of $\bigcup_{X \in \{A,B\}} X$.

10.2 Co-products

The sets $\emptyset$ and $\mathbb{1}$ behave in opposite ways, in the sense that for any set X, (i) there is exactly one function from $\emptyset$ to X and (ii) there is exactly one function from X to $\mathbb{1}$. We signalled this by refering to $\emptyset$ as *initial* and $\mathbb{1}$ as *terminal*. We say that the notion of an initial set is *dual* to the notion of a terminal set. The constructions of equalizers and products have duals called *co-equalizers* and *co-products*. Co-equalizers are indeed definable using the principles we already have, but it turns out that a related notion of *quotients* is more commonly used. We discuss quotients below, and build co-products first.

Definition 10.1

For sets X and Y, a *co-table* is a set with a pair of functions $X \xrightarrow{f} C \xleftarrow{g} Y$.

A *co-product* is a co-table $X \xrightarrow{i} S \xleftarrow{j} Y$ so that for any co-table $X \xrightarrow{f} C \xleftarrow{g} Y$ there is exactly one function $c\colon S \to C$ so that $f = c \circ i$ and $g = c \circ j$.

Lemma 10.1

Any two sets X and Y have a co-product.

Proof: Define $X \uplus Y = \{0\} \times X \cup \{1\} \times Y$. So the set $X \uplus Y$ consists of pairs (i, w) where $i = 0$ and $w \in X$ or $i = 1$ and $w \in Y$. Think of $X \uplus Y$ has consisting if a copy of X and a completely separate copy of Y.

Define the two functions $\mathsf{inj}_0 \colon X \to X \uplus Y$ and $\mathsf{inj}_1 \colon Y \to X \uplus Y$ by $\mathsf{inj}_0(x) = (0, x)$ and $\mathsf{inj}_1(y) = (1, y)$. Now, for any pair of functions $X \xrightarrow{f} C \xleftarrow{g} Y$ define $[f, g] \colon X \uplus Y \to C$ by cases:

$$[f, g](i, w) = \begin{cases} f(w) & \text{if } i = 0 \\ g(w) & \text{if } i = 1. \end{cases}$$

Evidently, $[f, g] \circ \mathsf{inj}_0 = f$ and $[f, g] \circ \mathsf{inj}_1 = g$. And clearly, no other function than $[f, g]$ has this property. □

The set $X \uplus Y$ is sometimes called the *disjoint union* of X and Y. It is a union, preceded by "marking" each $x \in X$ by putting it into a pair $(0, x)$ and marking each $y \in Y$ differently by putting it into a pair $(1, y)$. Hence it is the union of disjoint copies of X and Y.

Exercises for Lecture 10

Let $A = \{a, b, c, d, e\}$ and $B = \{w, x, y, z, a, b, c, \}$

1. Calculate $A \uplus B$
2. Calculate $A \uplus \emptyset$
3. Find a bijection between $\mathbb{N}$ and $\mathbb{1} \uplus \mathbb{N}$.
4. Find a bijection between $\mathbb{1} \uplus \mathbb{1}$ and $\mathbb{2}$.
5. For any sets X, Y and Z, define a bijection betweeen $X \times (Y \uplus Z)$ and $(X \times Y) \uplus (X \times Z)$.

10.3 Quotients

Sometimes, elements of a set X will be classified into "like" kinds. For example, we might classify the natural numbers into even and odd. We might classify poker cards according to suit, ignoring the rank – or by rank, ignoring suit. Or, if C is set modelling a Discrete Mathematics class, we might classify the elements (students) according to their grade: A, B, etc. Situations like this is modelled by what is known as a *partion* of a set. If we also wish to think of all A students as being "equivalent", all B students as being "equivalent", we model this by what is known as an *equivalence relation*.

Definition 10.2

For a set X, a *partition of* X is a set $P \subseteq \mathcal{P}(X)$ so that $\bigcup_{A \in P} A = X$ and for every $A, B \in P$, if $A \neq B$ then $A \cap B = \emptyset$. The sets in P are called *blocks*.

For a set X, an *equivalence relation on* X is a binary relation satisfying

- $x \mathrel{E} x$ for all $x \in X$,
- $x \mathrel{E} y$ and $y \mathrel{E} z$ implies $z \mathrel{E} z$ for all $x, y, z \in X$, and
- $x \mathrel{E} y$ imples $y \mathrel{E} x$

For a partition P define the relation $\equiv_P \subseteq X \times X$ by $x \equiv_P y$ if and only if for some $A \in P$, $x \in A$ and $y \in A$.

For an equivalence relation E and an element $x \in X$, let $[x]_E = \{y \in X \mid x \mathrel{E} y\}$. So $x \mapsto [x]_E$ defines a function from X to $\mathcal{P}(X)$. Let $X/E = \{[x]_E \mid x \in X\}$. That is, X/E is the range of the function $[\cdot]_E$.

The two notions, partition and equivalence relation, are essentially the same in the sense that there is a natural bijection between partitions and equivalence relations.

Lemma 10.2

Let X be any set. Then

- For any partition P of X, the relation $\equiv_P$ is an equivalence relation on X;
- For any equivalence relation E on X, the collection $\mathcal{P}_E$ forms a partition of X;
- For any partition P of X, $P = X/{\equiv_P}$
- For any equivalence relation E on X, $E = \equiv_{X/E}$;
- the rule $x \mapsto [x]_E$ determines an onto function from X to X/E.

Proof: Exercise. □

Suppose $E \subseteq X \times X$ is an equivalance relation and $f \colon X \to Y$ is a function so that $x \mathrel{E} x'$ implies $f(x) = f(x')$. Then we can define a function from X/E to Y by the "rule" $[x]_E \mapsto f(x)$. We can not call this a rule in the usual way because the left side is not a variable or a tuple of variables. But we can define a relation $F \subseteq (X/E) \times Y$ by stipulating that for $B \in X/E$ and $y \in Y$, $B \mathrel{F} y$ if and only if $f(x) = y$ for some $x \in B$. This relation is total because every block $B \in X/E$ non-empty. So there is some $x \in B$, and thus $B \mathrel{F} f(x)$. The relation F is also deterministic because if $B \mathrel{F} y$ and $B \mathrel{F} y'$, then there is some $x \in B$ for which $f(x) = y$ and there is some $x' \in B$ for which $f(x') = y'$. But $x, x' \in B$ implies $x \mathrel{E} x'$. Hence $f(x) = f(x')$.

Definition 10.3

Suppose $E \subseteq X \times X$ is an equivalence relation. A function $f \colon X \to Y$ is E-*invariant* if $x \mathrel{E} x'$ implies $f(x) = f(x')$. For an E-invariant function $f \colon X \to Y$, let f/E denote the function from X/E to Y defined by $(f/E)([x]_E) = f(x)$.

Example 10.1

Let $E \subseteq \mathbb{R} \times \mathbb{R}$ be the relation xEy if and only if $x - y = k2\pi$ for some integer k. This is an equivalence relation: xEx is true because $x - x = 0\dot{2}\pi$. If $x - y = k2\pi$ then $y - x = -k2\pi$, so E is symmetric. And if $x - y = k2\pi$ and $y - z = j2\pi$, then $x - z = (x - y) + (y - z) = (k + j)2\pi$. So E is transitive. The functions sin and cos are E invariant because $\sin(x + k2pi) = \sin(x)$ and $\cos(x + k2\pi) = \cos(x)$ for any x and any k.

Exercises for Lecture 10

On the integers $\mathbb{Z}$, define a relation $\equiv_7$ by $i \equiv_7 j$ if and only if there is some integer k so that $i + 7k = j$.

6. Show that $\equiv_7$ is an equivalence relation.
7. Describe the set $[5]_{\equiv_7}$.
8. Show that the function $f(n) = n + 3$ is $\equiv_7$-invariant.
9. Show that the function $g(n) = 2n$ is $\equiv_7$-invariant.
10. Determine whether the function $h(n) = 2^n$ is $\equiv_7$-invariant.

10.4 The Axioms of Choice, Replacement and Beyond

There are other axioms of set theory that mathematicians sometimes use, but that do not fit easily into our way of thinking about *constructions*. One such axiom is called the Axiom of Choice. Roughly is says that some how, one make arbitrarily many choices all at once. A famous example, due to Russel, is having an infinite collection of identical socks. One needs the Axiom of Choice to say that the socks can be put into pairs. The Axiom of Choice is quite useful in mathematics, sometimes to prove something that genuinely needs the axoim, other times to simplify a proof that could have been proved without it. Because of the special role the Axiom of Choice plays in mathematics, we return to discuss the Axiom of Choice in detail after we have some other concepts in place.

Another axiom says, roughly, that if a collection is the same size as a set, then it is a set. This is the Axiom of Replacement (think of replacing the elements of a set with elements of the collection, thus forming a new set). This is an important axiom for dealing with set theory, *per se*, but it does not come up much in practice. We can state a form of the axiom that is technical, but occasionally useful.

Principle 10.2

Suppose I is a set, and for each $i \in I$, a_i is a mathematical entity. Then there is a set, denoted by $\{a_i\}_{i \in I}$,satisfying $x \in \{a_i\}_{i \in I}$ if and only if $x = a_i$ for some $i \in I$. The set $\{a_i\}_{i \in I}$ is sometimes called an *indexed set*, or *indexed family*. For indexed set $\{a_i\}_{i \in I}$, the rule $i \mapsto a_i$ determines a function from I to $\{a_i\}_{i \in I}$.

For example, let $V_i = \mathbb{N}$ and let $V_{k+1} = \mathcal{P}(V_k)$. Then the principle tells us that $\{V_i\}_{i \in \mathbb{N}}$ is a set. Now define $V_\omega = \bigcup_{i \in \mathbb{N}} V_i$. This is a set that can not be constructed without Replacement. This princple can be made formal by introducing some additional logical machinery, but we will not make use of it in these lectures enough to warrant that attention.

Other stronger and stronger axioms of sets are possible. For these lectures, though, our goal is to have enough set theoretic machinery to do “ordinary” mathematics. We can stop here (remembering the promise to come back to the Axiom of Choice later). The interested student should follow up with a course in Set Theory.

Part III

Applications

11

Counting Finite Sets

Goals

Lecture

- Use the basic set and function constructions to discover techniques of counting finite systems.
- Introduce recurrence equations.

Study

- Learn the standard counting techniques.
- Be able to apply these techniques to common practical situations.

Consider the following questions:

- How many ways are there to choose a president, secretary and treasurer from among the 25 members of a student club?
- How many different five card poker hands are there?
- How many different ways can you be dealt a five card poker hand, one card at a time.
- How many of those poker hands are four-of-a-kind?
- How many three-letter words contain the letter "a"? [A "word" is any three letter sequence. So 'azx' counts even though it is not an English word.]

These are examples of *counting problems*. It turns out that there are actually a very small handful of techniques that allow us to solve a wide range of these sorts of problems.

11.1 Finite Sets

Recall that a finite set is a set the elements of which, in principle, can be listed. So a set modelling the collection of five card poker hands is finite. Not that we would actual wish to compile a list of all of them, but it could be done in principle. Far worse, we could consider listing the elements of $\mathcal{P}(\mathcal{P}(\mathcal{P}(\mathcal{P}(\mathcal{P}(\mathcal{P}(\mathcal{P}(\emptyset)))))))$. Actually writing these in a list would vastly exceeed using up all paper ever produced. Nevertheless, the set is finite because, *in principle*, its elements can be listed.

Recall that for any list $L = [a_0, \ldots, a_{n-1}]$, there is a set $\{a_0, \ldots, a_{n-1}\}$. So a set is finite if a corresponding L exists. Simple as that. Also, recall that we already defined the length of a list, inductively. So we can define *finite set* and the *size* of a finite set formally.

Definition 11.1

A set A is *finite* if there is a list $L = [a_0, \ldots, a_{n-1}]$ so that $A = \{a_0, \ldots, a_{n-1}\}$.

For a finite set A, let $|A|$ (the *size* of A) be the smallest natural number n for which there is a $L = \{a_0, \ldots, a_{n-1}\}$ satisfying $A = \{a_0, \ldots, a_{n-1}\}$.

A fancy synonym for *size* is *cardinality*.

This definition is a bit unsatisfying because calculating $|A|$ seems to require that we compare A to all of the ways we could list A's elements, and pick a shortest one. There should be a more direct method. Let us define standard sets of each finite size to use for comparison.

Definition 11.2

For natural number n, let $\tilde{n} = \{x \in \mathbb{N} \mid x < n\}$. So $\tilde{0} = \{\}$, $\tilde{1} = \{0\}$, $\tilde{2} = \{0, 1\}$, and so on.

Lemma 11.1

A set A is finite if and only if for some $n \in \mathbb{N}$, there are functions $A \xrightarrow{f} \tilde{n}$ and $\tilde{n} \xrightarrow{g} A$ so that $g \circ f = \mathsf{id}_A$.

For a finite set A, the value n can be chosen to be $|A|$. In that case, the functions f and g also $f \circ g = \mathsf{id}_{\tilde{n}}$

Proof: Suppose n and the indicated functions exist. Then the list $[g(0), \ldots, g(n-1)]$ lists elements of A. So $\{g(0), \ldots, g(n-1)\} \subseteq A$. For $a \in A$, $f(a) \in \tilde{n}$, so $g(f(a))$ is on the list. But $a = g(f(a))$. Thus $A \subseteq \{g(0), \ldots, g(n-1)\}$.

Suppose A is finite, then $A = \{a_0, \ldots, a_{n-1}\}$ for some list of length n. So we define $g(i) = a_i$ and $f(x) =$ the smallest i for which $x = a_i$.

The last part of the claim is obvious, but you should at least sketch why you believe it. □

This lemma provides a useful tool for determining the size of a finite set because the sets $\tilde{n}$ act as "measuring sticks": $|A| = n$ if and only if we can match up the elements of A with the elements of $\{0, \ldots, n-1\}$. So we are free to write $A = \{a_0, \ldots, a_{|A|-1}\}$, and prove facts abount finite sets by induction on $|A|$.

One detail needs to be taken into account in proofs by induction. Suppose $\{a_0, \ldots, a_{|A|-1}\} = A = \{b_0, \ldots, b_{|A|-1}\}$. That is, the same set A is listed in two different ways. We will need to be careful that a proof does not depend on which listing we use. This will become clear as we go along.

11.2 The Basic Principles

The basic fact about counting elements in finite sets is shat we may call *additivity*.

Lemma 11.2

For finite sets A and B, $\|A \uplus B\| = \|A\| + \|B\|$. So if $A \cap B = \emptyset$, then $\|A \cup B\| = \|A\| + \|B\|$.

Proof: Too obvious to bother writing the details. □

Lemma 11.3

For finite sets A and B, $\|A \times B\| = \|A\| \cdot \|B\|$.

Proof: This is also fairly obvious, but let's tackle a proof to illustrate the technique involved.
The proof is by inducion on the size of B. Fix a finite set A with $\|A\| = m$.

- [Basis] $A \times \emptyset = \emptyset$, and obviously, $\|\emptyset\| = 0$.
- [Inductive Hypothesis] Assume that for any finite set X, if $\|X\| = k$, and $\|A \times X\| = m \cdot \|X\|$.
- [Inductive Step] Suppose $\|B\| = k + 1$. Then B can be written as $\{b_0, \ldots, b_k\}$. Let $X = \{b_0, \ldots, b_{k-1}\}$ and let $A' = \{(a, b_k)\}_{a \in A}$. Clearly, $\|A\| = \|A\|'$. By the inductive hypothesis, and $\|A \times X\| = m \cdot k$. Now $A \times B = A \times X \cup A'$ and $A \times X \cap A' = \emptyset$. So $\|A \times B\| = \|A \times X\| + \|A\|' = m \cdot k + m = m \cdot (k + 1)$.

  [Note that the proof does not depend on how b_k is chosen. So it is valid for every way we might have written B.]

$\square$

Now that we know how to handle cartesian products of finite sets, exponents can not be far behind.

Lemma 11.4

For finite sets A and B, $\|B^A\| = \|B\|^{\|A\|}$.

Proof: The proof is by induction on $\|A\|$. We leave it as an exercise below. $\square$

Exercises for Lecture 11

In these exercises, you will prove the foregoing lemma. Suppose B is a finite set and $\|B\| = m$.

1. Show that $\|B^\emptyset\| = 1$. [This is really just an observation about functions with domain $\emptyset$.]
2. State the inductive hypothesis: "Assume that for any finite set X with $\|X\| = k$, ldots"
3. Prove the inductive step: For A of size $k + 1$, $\|B^A\| = \|B\|^{\|A\|}$.

11.3 Arrangements

Suppose we are given a set B and wish to form a list of k items drawn from B. This is called an k-*arrangement from* B. If the items on the list are allow to appear more than once, it is a k-arrangement *with replacement*. Otherwise it is *without replacement*.

Evidently, a k-arrangement from B, $[b_0, \mathtt{dotsc}, b_{k-1}]$ corresponds to a function $b\colon \tilde{k} \to B$. With replacement (items are permitted to appear more than once), there are no restrictions on the function. Without replacement (items can not be repeated), the function must be *one-to-one*, meaning that $b_i = b_j$ implies $i = j$.

The number of k-arrangements with replacement is $\|B\|^k$. The number of k-arrangements without replacement is the number of one-to-one functions.

Definition 11.3

For sets A and B, let $B^{\underline{A}}$ denote the set of one-to-one functions from A to B.

Definition 11.4

For $x \in RR$ and $k \in \mathbb{N}$, let $x^{\underline{k}}$ by the following recurrence.

$$x^{\underline{0}} = 1$$
$$x^{\underline{k^\frown}} = x \cdot (x-1)^{k}$$

Lemma 11.5

For finite sets A and B, $\|B^{\underline{A}}\| = \|B\|^{\underline{\|A\|}}$.

To count the number of poker hands, we can think about a set-theoretic *model* of the situation. We have a deck consisting of 52 cards. We can represent it as a set D. In a *hand*, the order of the cards does not matter. So we can model a hand as a subset of D of size 5. Similar situations, of course, come up for other sets and other sizes.

Definition 11.5

For a set X and natural number n, let $X^{[n]}$ denote the collection of subsets of X of cardinality n. So for example, $X^{[0]} = \{\emptyset\}$, and $X^{[1]}$ consists of all the singleton subsets of X.

So the question of how many five card poker hands there are amounts to asking for $\|D^{[5]}\|$.

To be dealt a five card poker hand one at a time, you are dealt a first card, then a second, then a third, and so on. This can be modelled as a function d from the set $\tilde{5} = \{0, 1, 2, 3, 4\}$ into D. That is, your deal is $d(0)$, $d(1)$, $d(2)$, $d(3)$, then $d(4)$. But you can not be dealt the same card twice (if your dealer is honest). So if $d(i) = d(j)$, then $i = j$. In other words, d must be a one-to-one function. Thus to calculate the number of ways to be dealt a five card hand, we need to calculate the number of one-to-one functions from $\{0, 1, 2, 3, 4\}$ to D. Again, we generalize.

Definition 11.6

For sets X and Y, let $Y^{\underline{X}}$ denote the set of one-to-one functions from X to Y. So

$$Y^{\underline{X}} = \{f \in Y^{X} \mid \forall x_0, x_1 \in X. f(x_0) = f(x_1) \to x_0 = x_1\}.$$

So to count the number of ways to be dealt a five card hand we need to calculate $\|D^{\underline{\tilde{5}}}\|$.

The simplest counting principle is the one you have known since you learned to count.

Principle 11.1

For sets A and B, if A is finite and $A \sim B$, then B is finite and $\|A\| = \|B\|$.

This seems so trivial that it hardly needs stating. But in fact we will use it count the size of a set B by counting A and setting up a suitable bijection. we illustrate by establishing another very simple fact.

Lemma 11.6

For finite sets A and B, $\|A \uplus B\| = \|A\| + |B\|$. So if $A \cap B = \emptyset$, then $\|A \cup B\| = \|A\| + \|B\|$.

Proof: This hardly needs a proof, but here goes. Suppose $\|A\| = m$ and $\|B\| = n$. That is, there are bijections $f\colon A \to \tilde{m}$ and $g\colon B \to \tilde{n}$. Define $h\colon A \uplus B \to \widetilde{m+n}$ by

$$h(i, x) = \begin{cases} f(x) & \text{if } i = 0 \\ m + g(x) & \text{otherwise.} \end{cases}$$

It is routine to check that h is indeed a bijection.

If $A \cap B = \emptyset$, then there is a bijection between $A \cup B$ and $A \uplus B$. □

When we declare a set A to be *finite*, we mean that it is equipotent with some $\tilde{n}$. So letting $a\colon \tilde{n} \to A$ be a bijection, we can list the elements of $A = \{a_0, a_1, \ldots, a_{n-1}\}$. This allows us to prove facts about finite sets using induction on their sizes. Our next fact is easy to prove and illustrates how induction helps.

Surely, we could have said that this lemma is "obvious" and skipped the proof. But for the purpose of learning how to prove claims, we spelled the proof out.

Lemma 11.7

For finite sets A and B, $\|B^A\| = \|B\|^{\|A\|}$.

Proof: Exercise. Hint: Use induction on $\|A\|$. The key to the inductive step is to show that for $a_0 \notin X$, $B^{X \cup \{a_0\}} \sim B \times B^X$. □

11.4 Permutations and Combinations

A permutation is an arrangement of elements in which their order matters. A combination is an arrangement of elements in which their order does not matter.

A student club with 20 members must pick a president, a secretary and a treasurer. Which club member is picked for president versus secretary versus treasurer matters. Also, club by-laws forbid anyon from serving in two posts at the same time. This is a permutation of 3 items from 20.

In a five card poker hand, the order in which the five cards are dealt does not matter. This is a combination of 5 items chosen from 52.

These two kinds of counting problems are not the same, but they are closely related. To understand them, we need to think about how they may be modelled in sets.

The leadership of the student club can be modelled as assigning to each of the three leadership positions, a member of the club. So let $L = \{\mathsf{p}, \mathsf{s}, \mathsf{t}\}$, and let M be a set modelling the membership of the club. Thus a leadership assignment is a function $a\colon L \to M$. But suppose $a(\mathsf{m}) = a(\mathsf{s})$. This would signify that the same person is assigned to be the president and the secretary, violoating the by-laws. To avoid this, a should be a one-to-one function. Hence, to count the number of leadership assignments is to count the number of one-to-one functions from L to M.

For five car poker hand, let D be the set modelling a dick of cards. Then a hand is modelled as a subset $H \subseteq D$ for which $\|H\| = 5$.

Definition 11.7

For finite sets A and B, let $B^{\underline{A}}$ denote the collection of all one-to-one functions from A to B.

For finite set A and natural number n, an n-*subset* is a subset of A of size n. Let $A^{[n]}$ denote the collection of all n-subsets of A.

A set modelling all possible leadership assignments for the student club is $M^{\underline{L}}$. A set modelling all possible poker hands is $D^{[5]}$. So we solve our two problem specifically by figuring how to count $B^{\underline{A}}$ and $B^{[n]}$ generally.

Consider the number of one-to-one functions from L to M. Evidently, we can chose any student from M to be president. So there are 20 possibilities. After that, we chose a secretary from the remaining 19, then a treasurer from the remaining 18. So $20 \cdot 19 \cdot 18$ is our answer. Pretty obviously, the pattern is to multiply $20 \cdot (20-1) \cdot (20-2)$. This generalizes.

Definition 11.8

For an integer a and natural number m, define the *falling exponent* $a^{\underline{m}}$ recursively by

$$
\begin{aligned}
a^{\underline{0}} &= 1 \\
a^{\underline{m^{\frown}}} &= a \cdot (a-1)^{\underline{m}}
\end{aligned}
$$

So $a^{\underline{m}} = (a-0) \cdot (a-1) \cdots (a-2) \cdots (a-(m-1))$. So there are m terms in this expression.

We can now prove that $\|B^{\underline{A}}\|$ is indeed calculated by a falling exponent.

Lemma 11.8

For finite sets A and B, $\|B^{\underline{A}}\| = \|B\|^{\underline{\|A\|}}$.

Proof: Fix a finite set B. Now by induction on $\|A\|$,

- [Basis] The set $B^{\underline{\emptyset}}$ consists of the single function from $\emptyset$ to B. So $\|B^{\underline{\emptyset}} = 1$.
- [Inductive Hypothesis] Assume that $\|B^{\underline{X}}\| = \|B\|^{\underline{\|X\|}}$ for any set X for which $\|X\| = k$.
- [Inductive Step] Suppose $\|A\| = k + 1$. Let a_0 be an element of A.

  If $B = \emptyset$, then there are no functions from A to B. So $\|B^{\underline{A}}\| = 0$. But also, by the inductive definition of falling exponents, $0^{\underline{k+1}} = 0 \cdot (-1)^{\underline{k}} = 0$.

  If $B \neq emptyset$, let $b_0 \in B$. We define a bijection between $B^{\underline{A}}$ and $B \times (B \setminus \{b_0\})^{\underline{A \setminus \{a_0\}}}$. Then the result follows directly from the recursive definition of falling exponents.

  For the needed bijection, first note that for any $b, b' \in B$ there is a bijection $e_{b,b'}$ between $B \setminus \{b\}$ and $B \setminus \{b'\}$ given by

  $$e_{b,b'}(x) = \begin{cases} x & \text{if } x \neq b' \\ b & \text{otherwise} \end{cases}$$

  It is easily checked that $e_{b,b'} \circ e_{b',b''} = e_{b,b''}$, and $e_{b,b}$ is identity on $B \setminus \{b\}$. For any one-to-one function $f\colon A \to B$, define $h_f\colon A \setminus \{a_0\} \to B \setminus \{b_0\}$ by $h_f(a) = e_{f(a_0),b_0}(f(a))$. Then $f \mapsto (f(a_0), h_f)$ is a function from $B^{\underline{A}}$ to $B \times (B \setminus \{b_0\})^{\underline{A \setminus \{a_0\}}}$.

  If $(f(a_0), h_f) = (g(a_0), h_g)$, then $f(a_0) = g(a_0)$ and for every $a \in A \setminus \{a_0\}$, $e_{f(a_0),b_0}(f(a)) = e_{g(a_0),b_0}(g(a))$. If $f(a) \neq b_0$ and $g(a) \neq b_0$ for some a, then $f(a) = e_{f(a_0),b_0}(f(a)) = f(a) = e_{g(a_0),b_0}(g(a)) = g(a)$. If $f(a) \neq b_0$ and $g(a) = b_0$, then $f(a) = e_{f(a_0),b_0}(f(a)) = b_0 = g(a)$. Since this is a contradiction, we rule out $f(a) \neq b_0 = g(a)$. Likewise, $f(a) = b_0 \neq g(a)$ is impossible. So $f(a) = g(a)$ for all a. This shows that $f \mapsto (f(a_0), h_f)$ is one-to-one.

  Consider $(b, k) \in B \times (B \setminus \{b_0\})^{\underline{A \setminus \{a_0\}}}$. Define $f\colon A \to B$ by

  $$f(a) = \begin{cases} b & \text{if } a = a_0 \\ e_{b_0,b}(k(a)) & \text{otherwise} \end{cases}$$

  Then f is indeed one-to-one, for $e_{b_0,b}(k(a)) \neq b$ by definition. And $f(a_0) = b$ and for any $a \in A \setminus \{a_0\}$,

  $$\begin{aligned} h_f(a) &= e_{f(a_0),b_0)}(f(a)) \\ &= e_{f(a_0),b_0}(e_{b_0,b}(k(a))) \\ &= e_{f(a_0),b}(k(a)) \\ &= k(a) \end{aligned}$$

  Hence the function $f \mapsto (f(a_0), h_f)$ is also onto.

□

The law of exponents that $a^m \cdot a^n = a^{m+n}$ has an analogue for falling exponents.

Lemma 11.9

For any integer a and natural numbers m and n, $a^{\underline{m+n}} = a^{\underline{m}} \cdot (a-m)^{\underline{n}}$.

Proof: Exercise. Use induction on m. □

Notice that $m^{\underline{m}} = m!$, and $m^{\underline{n}} = m!/(m-n)!$.

To determine a formula for $B^{[n]}$ for a finite set B and natural number number n, we note that a way to pick an element of $B^{\underline{\tilde{n}}}$ (one-to-one function from $\{0, \ldots, n-1\}$ to B) is to pick an n-subset of $A \subseteq B$, then pick an ordering for that set (one-to-one function from $\{0, \ldots, n-1\}$ to A). In other words, $B^{\underline{\tilde{n}}}$ is in a bijection with $B^{[n]} \times \tilde{n}^{\underline{\tilde{n}}}$.We have formulas for $B^{\underline{\tilde{n}}}$ and for $\tilde{n}^{\underline{\tilde{n}}}$. So

$$\|B\|^{\underline{n}} = \|B^{[n]}\| \cdot n!$$

Definition 11.9

For natural numbers m and k so that $k \leq m$, define $\binom{m}{k}$ to be $\frac{m!}{(m-k)!k!}$.

Lemma 11.10

For finite set B and natural number k so thst $k \leq \|B\|$, $\|B^{[n]}\| = \binom{\|B\|}{k}$.

Proof: From the obervations we made in the previous paragraph, $\|B^{\underline{\tilde{n}}} = \|B^{[n]}\| \cdot \|\tilde{n}^{\underline{\tilde{n}}}$. Since $\|B^{\underline{\tilde{n}}} = \|B\|^{\underline{n}}$ and $\|\tilde{n}^{\underline{\tilde{n}}}\| = n!$, the result follows by solving for $\|B^{[n]}\|$. □

Example 11.1

The number of five card hands from a 52 card deck is $\binom{52}{5}$. That is, $\frac{52!}{47!5!} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 52 \cdot 51 \cdot 10 \cdot 49 \cdot 2 = 2,598,960$.

Example 11.2

At a pub, you and a group of 10 friends order beer and, as usual, clink your glasses before quaffing. If everyone clinks every one else's glass exactly once, how many separate clink sounds should you hear?

We can model the situation by taking F to be aset modelling you and your friends. A "clink" corresponds to a 2-subset because (i) you need two glasses to make a noise and (ii) it does not matter if you think of Tadeusz clinking Suenyung's glass or the other way around. So we merely need to count $F^{[2]}$. Since $\|F\| = 11$, the answer is $\binom{11}{2} = 55$.

A Useful Relation: Pascal's Triangle

It is easy to see that $\binom{n}{0} = 1$. Combinatorially, this signifies that there is exactly one 0-subset of any set of size n. Likewise, $\binom{n}{n} = 1$, signifying that there is exactly one n-subset of a set of size n. For both reasons $\binom{0}{0} = 1$. More interesting things occur in between, when $n > 0$ and $0 < k < n$.

Suppose $A = X \cup \{a_0\}$ where $a_0 \notin X$ and $k \leq \|X\|$. Each $k+1$-subset $B \subseteq A$ either contains a_0 or not. If $a_0 \in B$, then $B \cap X$ is a k-subset of X. Otherwise $B \cap X$ is a $k+1$-subset of X. Moreover, any $k+1$ subset of X is also a $k+1$ subset of A. And for any k-subset of X, $X \cup \{a_0\}$ is a $k+1$-subset of A. So there is a bijection between $A^{[k+1]}$ and $X^{[k]} \cup X^{[k+1]}$. But $X^{[k]} \cap X^{k+1]} = \emptyset$. So

$$\|A^{[k+1]}\| = \|X^{[k]}\| + \|X^{[k+1]}\|.$$

Translating this back to combinations,

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

This simple relationship is the basis of a famous mathematical figure, known as Pascal's triangle. Written in combination symbols, the first few rows are

$$\begin{array}{c}
\binom{0}{0} \\
\binom{1}{0} \quad \binom{1}{1} \\
\binom{2}{0} \quad \binom{2}{1} \quad \binom{2}{2} \\
\binom{3}{0} \quad \binom{3}{1} \quad \binom{3}{2} \quad \binom{3}{3} \\
\binom{4}{0} \quad \binom{4}{1} \quad \binom{4}{2} \quad \binom{4}{3} \quad \binom{4}{4} \\
\binom{5}{0} \quad \binom{5}{1} \quad \binom{5}{2} \quad \binom{5}{3} \quad \binom{5}{4} \quad \binom{5}{5} \\
\binom{6}{0} \quad \binom{6}{1} \quad \binom{6}{2} \quad \binom{6}{3} \quad \binom{6}{4} \quad \binom{6}{5} \quad \binom{6}{6} \\
\binom{7}{0} \quad \binom{7}{1} \quad \binom{7}{2} \quad \binom{7}{3} \quad \binom{7}{4} \quad \binom{7}{5} \quad \binom{7}{6} \quad \binom{7}{7} \\
\binom{8}{0} \quad \binom{8}{1} \quad \binom{8}{2} \quad \binom{8}{3} \quad \binom{8}{4} \quad \binom{8}{5} \quad \binom{8}{6} \quad \binom{8}{7} \quad \binom{8}{8} \\
\binom{9}{0} \quad \binom{9}{1} \quad \binom{9}{2} \quad \binom{9}{3} \quad \binom{9}{4} \quad \binom{9}{5} \quad \binom{9}{6} \quad \binom{9}{7} \quad \binom{9}{8} \quad \binom{9}{9} \\
\binom{10}{0} \quad \binom{10}{1} \quad \binom{10}{2} \quad \binom{10}{3} \quad \binom{10}{4} \quad \binom{10}{5} \quad \binom{10}{6} \quad \binom{10}{7} \quad \binom{10}{8} \quad \binom{10}{9} \quad \binom{10}{10}
\end{array}$$

Written again as numerals, the first few rows

$$\begin{array}{c}
1 \\
1 \quad 1 \\
1 \quad 2 \quad 1 \\
1 \quad 3 \quad 3 \quad 1 \\
1 \quad 4 \quad 6 \quad 4 \quad 1 \\
1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1 \\
1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1 \\
1 \quad 7 \quad 21 \quad 35 \quad 35 \quad 21 \quad 7 \quad 1 \\
1 \quad 8 \quad 28 \quad 56 \quad 70 \quad 56 \quad 28 \quad 8 \quad 1 \\
1 \quad 9 \quad 36 \quad 84 \quad 126 \quad 126 \quad 84 \quad 36 \quad 9 \quad 1 \\
1 \quad 10 \quad 45 \quad 120 \quad 210 \quad 252 \quad 210 \quad 120 \quad 45 \quad 10 \quad 1
\end{array}$$

Notice that the numbers at the edges are all 1's and all others are the sum of the two numbers just above. Many other useful relations are part of Pascal's triangle. You will explore some of them in the following exercises.

Exercises for Lecture 11

1. The Chapman chapter of Students United for Caring has 20 members. They have decided to form a 5 person outreach committee. They want to pick the committee by lottery. How many different ways could the committee be formed.

2. The Department of Unrepeatible Experiments gives three awrads to it top graduating students: the "Outstanding Student" award, the "Unremarkable Student" award and the "Least Likely to Succeed" award. This year there are 100 graduating students. How many ways can the awards be given?

3. Suppose you own 25 pigeons and have constructed a pigeon coup that is a 10-by-10 grid of boxes. Each box is big enough to fit one pigeon. How many ways can the pigeons nest in the pigeon coup? Pigeons are individuals. So it matters which pigeon nests in which box.

4. Suppose you own 25 chickens instead of pigeons. Unlike pigeons, all chickens are identical. How many ways can the chickens nest?

5. In Pascal's triangle, add the entries in a row. Notice that the result is always 2^n where n is the row number (the apex of the triangle is row 0). Explain this fact.

6. In Pascal's triangle, for any row except row 0, add every other entry. For example in row 5, we add $1+10+5$. In row 6, we add $1+15+15+1$. In every case, the result is 2^{n-1} where n is the row number. Explain this fact.

11.5 Without and With Replacement

Suppose we play a lottery. There are 50 balls numbered 1 to 50. Then 6 balls are selected at random. We know that this means there are $\binom{50}{6}$ combinations. But suppose we change the game just a little. After each ball is selected, its number is written down, then it is put back in the hopper before the next ball is selected. So it is possible to get the same ball more than once. Still, the order does not matter. This is called choosing a combination *with replacement*. The usual lottery method is choosing combination *without replacement*.

For permutations, a similar distinction is useful. What we have called a permutation of k items from n is *without replacement*. This is modelled mathematically by a one-to-one function. *With replacement* would mean that we pick items $a_0, a_1, \ldots, a_{k-1}$ from a set of n elements and lift the one-to-one requirement so that a_i may equal a_j without any restriction. We know how to count permutations without replacement ($n^{\underline{k}}$). It is obvious that permutations with replacement are unrestricted functions, so we know how to count them too (n^k).

For combinations with replacement the analysis is more subtle. At first, we might think that we can merely count the permuttions with replacement (n^k) nd divide by the number of permutations of k items from k elements. So $\frac{n^k}{k!}$ seems like a reasonable conjecture. But checking a few easy examples shows this can not be right.

Suppose we have a set of two elements $\{a, b\}$ and wish to choose a combination of 3 items with replacement. We can simply list them: aaa, aab, abb, and bbb. so the result is 4. But $2^3/3!/neq4$. So much for our conjecture. Notice that $2^3/3!$ is not even a natural number. So our formula does not make any sense at all.

Let us try a different tac. Consider an equation $3 = x + y$. How many ways can this be solved with natural numbers? The solutions are eas Notice that this is the same as choosing a combination of 3 items from 2 elements with replacement. In fact, the same idea holds up in general. A combination of k items from n elements with replacement is the number of ways to solve an equation $k = x_0 + x_1 + \cdots + x_{n-1}$ with natural numbers.

Lemma 11.11

For any natural numbers $n > 0$ and k, the number of combinations of k items from n elements is $\binom{n+k-1}{k}$.

Proof: Before we attempt a formal proof, consider an example of solving $6 = x_0 + x_1 + x_2 + x_3$ in natural numbers, using our successor notation. For example, $0^{\frown\frown} + 0^{\frown} + 0^{\frown} + 0^{\frown\frown}$ is a solution. Notationally, the 0's are not telling us anything because the initial 0 is always there, and all subsequent 0's follow a $+$ sign. So we can abbreviate our solution by erasing the 0's: $\frown\frown + \frown + \frown + \frown\frown +$. Every solution can be written this way, and every different placemnt of the $+$'s and $\curvearrowright$'s corresponds to a different solution. This amounts to choosing 6 $\curvearrowright$ symbols and 3 $+$ symbols for the total of nine positions, so there are $\binom{9}{6}$.

The general situation can be proved by induction on n.

- [Basis] For $n = 1$, there is exactly one solution of the equation $k = x_0$. And $\binom{n+k-1}{k} = 1$.
- [Inductive Hypothesis] Assume that for some $0 < n$, it is the case that for all k, the number of solutions to $k = x_0 + \cdots + x_{n-1}$ is $\binom{n+k-1}{k}$.
- [Inductive Step] To count the solutions of $k = x_0 + \cdots + x_n$, we proceed by induction on k.
  - [Basis] There is one solution of $0 = x_0 + \cdots + x_n$ and $\binom{(n+1)-0-1}{0} = 1$.
  - [Inductive Hypothesis] Suppose the number of solutions of $k = x_0 + \cdots + x_n$ is $\binom{(n+1)+k-1}{k}$.
  - [Inductive Step] There are two sorts of solution to $k + 1 = x_0 + \cdots + x_n$: either $x_n = 0$ or $x_n > 0$. Solutions for which $x_n = 0$ are equivalent to solutions of $k+1 = x_0 + \cdots + x_{n-1}$. The number of such solutions is $\binom{n+(k+1)-1}{k+1}$ by the "outer" inductive hypothesis. The solutions for which $x_n > 0$ are equivalent to solutions of $k = x_0 + \cdots + x_n$. The number of such solutions is $\binom{(n+1)+k-1}{k}$ by the "inner" inductive hypothesis. Since these two sorts of solution cover all cases (and they are disjoint), the total number of solutions of $k + 1 = x_0 + \cdots + x_n$ is $\binom{n+k}{k} + \binom{n+k}{k+1} = \binom{(n+1)+(k+1)-1}{k+1}$ by Pascal's identity $\binom{m}{j} + \binom{m}{j+1} = \binom{m+1}{j+1}$.

  The inner induction is complete, so the number of solutions of $k = x_0 + \cdots + x_n$ is $\binom{(n+1)+k-1}{k}$ for all k. This is the needed conclusion for the outer inductive step.

□

Exercises for Lecture 11

1. Calculate the number of ways to order dozen donuts in a shop that carries 7 varieties. You may order any number of donuts of any variety.
2. Calculate the number of ways to order a half dozen tacos at a stand that carries thirteen varieties.
3. For natural numbers m and n, the number of natural number solutions of $m = x_0 + \cdots + x_n$ equals the number of natural number solutions of $n = x_0 + \cdots + x_m$. Proof this.

12

Monomorphisms, Epimorphisms and Isomorphisms

Goals

Lecture

- Introduce special kinds of functions: monomorphism, epimorphism, injection, surjection.
- Investigate the relations between these.

Study

- Learn to recognize injections and surjections by internal behavior
- Be able to illustrate simple, small examples of injections, non-injections, surjections, non-surjections.

12.1 Basics

Recall that we required that if a natural number has a predecessor, it has exactly one. We wrote that as an axiom: $m^\curvearrowright = n^\curvearrowright$ implies $m = n$. Putting this in terms of functions: $\mathtt{suc}(m) = \mathtt{suc}(n)$ implies $m = n$. Likewise, the singleton function from X to $\mathcal{P}(X)$ given by $x \mapsto \{x\}$ has a similar property that $\{x\} = \{y\}$ implies $x = y$.

Also, recall that we proved that addition is *cancellative*: $m + p = n + p$ implies $m = n$, and that multiplication by a positive natural number is cancellative: $m \cdot p^\curvearrowright = n \cdot p^\curvearrowright = m = n$.

This leads to some definitions.

Definition 12.1

For a function $g\colon B \to C$, say that

- g is an *injection* if $g(x_0) = g(x_1)$ implies $x_0 = x_1$ for every $x_0, x_1 \in B$;
- g is a *monomorphism* if $g \circ f_0 = g \circ f_1$ implies $f_0 = f_1$ for every $f_0, f_1 : A \to B$; and
- g is an *epimorphism* if $h_0 \circ g = h_1 \circ g$ implies $h_0 = h_1$ for every $h_0, h_1 : C \to D$.

So an injection is a function with the property that application of it is cancellative. A monomorphism is a function with the property that composition with it on the left is cancellative. An epimorphism is a function with property that composition with it on the right is cancellative.

An injection is also said to be *injective* or *one-to-one*. A monomorphism is also said to be *left-cancellable*. An epimorphism is also said to be *right-cancellable*.

We will look at epimorphisms later. For now we concentrate on monomorphisms and injections.

Example 12.1

1. Consider the function $f\colon \mathbb{N} \to \mathbb{N}$ defined as $f(n) = 2 \cdot n + 2$. Suppose $f(m) = f(n)$, then $2 \cdot m + 2 = 2 \cdot n + 2$. Addition is cancellative, so $2 \cdot m = 2 \cdot n$. And multiplication by a positive natural number is cancellative, so $m = n$. Thus f is an injection.

2. Consider $g\colon \mathbb{R} \to \mathbb{R}$ defined as $f(x) = x^2$. This is not injective because $f(1) = f(-1)$, but $1 \neq -1$.

Suppose $g\colon B \to C$ is a monomorphism. Then it clearly is also an injection. That is, suppose g is a monomorphism and that $g(b_0) = g(b_1)$. We already know that $\widehat{g(b_0)} = g \circ \widehat{b_0}$. So $g \circ \widehat{b_0} = g \circ \widehat{b_1}$. Because g cancels on the left, $\widehat{b_0} = \widehat{b_1}$. But this implies that $b_0 = b_1$. The converse of this is also true.

Lemma 12.1

A function $g\colon B \to C$ is a monomorphism if and only if it is an injection.

Proof: The previous paragraph proves that every monomorphism is an injection. To prove the converse, assume that g is an injection. That is, $g(b_0) = g(b_1)$ implies $b_0 = b_1$. For two functions $f_0, f_1\colon A \to B$, suppose $g \circ f_0 = g \circ f_1$. For any $x \in A$, then $g(f_0(x)) = g(f_1(x))$. So $f_0(x) = f_1(x)$. Hence by the Principle of Function Extensionality, $f_0 = f_1$. □

As simple as this result seems, it is valuable because it provides an internal characterization (injections deal with elements) for an external property (monomorphisms deal with composition).

Inclusion maps of subsets constitute an important class of monomorphisms. That is, if $A \subseteq X$, then the inclusion map $i\colon A \to X$ is obviously one-to-one, so it is a monomorphism.

Epimorphisms are also defined externally, but we do not yet have an internal characterization. Taking a closer look at injectivity gives a clue about how we might characterize epimorphisms internally, too.

Lemma 12.2

A function $g\colon B \to C$ is an injection if and only if for any $y \in C$ there is *at most one* $x \in B$ for which $g(x) = y$.

Proof: There is nothing to prove here, really. This "at most one" formulation is essentially just a rephrasing of the definition of injectivity. □

This suggests a sort of "dual" definition. To make the connection clear, we repeat the definition of injections.

Definition 12.2

For a function $g\colon B \to C$, say that

- g is an *injection* if for every $y \in C$, there is *at most one* $x \in B$ for which $g(x) = y$;
- g is a *surjection* if for every $y \in C$, there is *at least one* $x \in B$ for which $g(x) = y$;
- g is a *bijection* if it is both an injection and a surjection. That is, for every $y \in C$, there is *exactly one* $x \in B$ for which $g(x) = y$.

We already know that injections and monomorphisms are the same things.

Lemma 12.3

A function $g\colon B \to C$ is an surjection if and only if it is an epimorphism.

Proof: Assume that g is a surjection. Suppose $h_0, h_1\colon B \to D$ are functions for which $h_0 \circ g = h_1 \circ g$. For $y \in C$, pick some $x \in B$ for which $g(x) = y$. This can be done because g is a surjection. So $h_0(y) = h_0(g(x)) = h_1(g(x)) = h_1(y)$. This works for every $y \in C$, so $h_0 = h_1$. This proves g is an epimorphism.

Assume that g is *not* a surjection. We will prove that it is not an epimorphism. Supposing that g is not a surjection means there is an element $c \in C$ so that $g(x) \neq c$ for all $x \in B$. Consider the functions $h_0\colon C \to 2$ defined by $h_0(y) = \top$ and $h_1\colon C \to 2$ defined by

$$h_1(y) = \begin{cases} \top & \text{if } y = c \\ \bot & \text{otherwise} \end{cases}$$

For every $x \in B$, $h_0(g(x)) = \top = h_1(g(x))$. So $h_0 \circ g = h_1 \circ g$, but obviously $h_0(c) \neq h_1(c)$. So $h_0 \neq h_1$. □

Exercises for Lecture 12

1. Prove that for any set A, id_A is a monomorphism.
2. Prove that if $f\colon A \to B$ and $g\colon B \to C$ are monomorphisms, then $g \circ f$ is a monomorphism.
3. Formulate and prove analogous claims for epimorphisms.

12.2 Split Monomorphisms and Split Epimorphisms, the Axiom of Choice

Suppose $f\colon A \to XY$ and $g\colon Y \to X$ are functions so that $g \circ f = \mathsf{id}_X$. That is, $g(f(x)) = x$ for all $x \in X$. Then $f(x) = f(x')$ implies $x = g(f(x)) = g(f(x')) = x'$. So f is one-to-one. Also, for any $x \in X$, $g(f(x)) = x$ means that g is onto.

Definition 12.3

A function $f\colon X \to Y$ is a *split monomorphism* if there is a function $g\colon Y \to X$ for which $g \circ f = \mathsf{id}_X$. In that case, g is called a *left inverse* of f. A function $f\colon X \to Y$ is a *split epimorphism* if there is a function $g\colon Y \to X$ for which $f \circ g = \mathsf{id}_Y$. In that case, g is called a *right inverse* of f.

Except for in one special case, every monomorphism is split.

Lemma 12.4

If $f\colon X \to Y$ is a monomorphism, then either $X = \emptyset$ or f is split.

Proof: Suppose $f\colon X \to Y$ is a monomorphism and X is not empty. Let x_0 be an element of X. Then define the relation $G \subseteq Y \times X$ by $y\ G\ x$ if and only if either (i) $y = f(x)$ or (ii) $y \neq f(x')$ for all $x' \in X$ and $x = x_0$. Evidently, G is a total relation: for each $y \in Y$, either $y = f(x)$ for some x, or $y \; G \; x_0$. Also it is deterministic: If $y \; G \; x$ and $y \; G \; x'$, then either $f(x) = y = f(x')$, hence $x = x'$, or $x = x_0 = x'$. So G determines a function, $g \colon Y \longrightarrow X$. Evidently, $g(f(x)) = x$ by definition of g. □

We might hope that the dual fact is also true, that every epimorphism is split. But in fact, this is does not follow from the principles to which we have agreed so far. This requires a new principle, called the Axiom of Choice (AC).

Principle 12.1

Every epimorphism is split. For an epimorphism $f \colon X \longrightarrow Y$, a function $g \colon Y \longrightarrow X$ for which $f \circ g = \mathsf{id}_Y$ is called a *choice function* for f. This name makes sense, because for each $y \in Y$, $g(y) \in f^{-}(y)$. So g choices an element of $f^{-}(y)$ for each y.

It is probably not obvious to you why the Axiom of Choice is even necessary. But remember, we intend to be explicit about how sets and functions are built. Without AC, there may not be any other principle by which an epimorphism can be split. The principle that epimorphisms are split has many useful consequences, most of which are difficult to state without further development. In general, we will not use AC without explicitly discussing how we use it.

Exercises for Lecture 12

1. The function $f \colon \mathbb{N} \longrightarrow \mathbb{N}$ given by $x \mapsto x^2$ is a monomorphism. Give an epimorphism that splits it.
2. Let $i \colon \mathbb{Z} \longrightarrow \mathbb{R}$ be the obvious inclusion function. Let $\lfloor \, \rfloor \colon \mathbb{R} \longrightarrow \mathbb{N}$ be the floor function: $\lfloor x \rfloor$ is the largest integer less than or equal to x. Is i a left or right inverse of $\lfloor \, \rfloor$?
3. The function $x \mapsto e^x$ is a monomorphism. Define a function that splits it.

12.3 Bijections, Inverses and Cardinality

Definition 12.4

A function $f \colon X \longrightarrow Y$ is an *isomorphism* if there is a function $g \colon Y \longrightarrow X$ so that $g \circ f = \mathsf{id}_X$ and $f \circ g = \mathsf{id}_Y$. In that case, g is called an *inverse* of f. Note that g is both a left and right inverse.

Note that if f is an isomorphism, then it is both a split monomorphism and a split epimorphism. Hence it is one-to-one and onto. The term we have used for this is *bijection*. Conversely, if $f \colon X \longrightarrow Y$ is a bijection, then the relation $F^{-1} \subseteq Y \times X$ given by $y \; F^{-1} \; x$ if and only if $y = f(x)$ is total and deterministic. So it determines a function f^{-1}. It is routine to check that f^{-1} is an inverse of f. In other words, any bijection is invertible.

It is also easy to see that if f has an inverse, that inverse is unique. So we may refer to *the* inverse of f, and write f^{-1}.

Exercises for Lecture 12

1. Suppose $f \colon X \longrightarrow Y$ and $g \colon Y \longrightarrow Z$ are isomorphisms. Show that $g \circ f$ is an isomorphism. [Hint: Ask what the inverse of $g \circ f$ is.]

2. Show that $\mathsf{id}_X \colon X \to X$ is an isomorphism. [Hint: Ask what the inverse of id_A is.]

3. Suppose $f \colon X \to Y$ is an isomorphism. Show that f^{-1} is an isomorphism. [This is quite easy.]

The problem of judging which sets are the *same size* seems easy enough when we look at finite sets. But consider, for example, how to compare $\mathbb{N}$ and $\mathbb{N} \times \mathbb{N}$. These are both infinite sets, but are they the same size? Cantor realized that bijections are the key to understanding this. That is, he *defined* two sets to be the same size if they are bijective. This gives us a precise criterion for comparing sets by their size. We introduce the idea formally.

Definition 12.5

For sets A and B, say that A and B are *equipotent*, written $A \sim B$, if A and B are bijective.

Thus "equipotent," "bijective" and "isomorphic" mean the same thing. The usage differs mainly in terms of emphasis. When we are interested in merely comparing sets by size, we mostly use the word "equipotent".

12.4 Countable Sets

We are often interested in sets that are not too big, i.e., that ar no bigger than $\mathbb{N}$.

Definition 12.6

A set A is *countable* if A is empty, or there is a surjection $a \colon \mathbb{N} \to A$. That is, a_0, a_1, $\ldots$, exhaustively "lists" the element of A.

Notice that any non-empty finite set A is countable. For if we can list the elements, $A = \{a_0, \ldots a_n\}$, then we define a function $r \colon A \to A$ by

$$r(a_i) = \begin{cases} a_{i+1} & \text{if } i < n \\ a_0 & \text{otherwise} \end{cases}$$

So the recurrence $\mathbb{1} \xrightarrow{hata_0} A \xleftarrow{r} A$ determines a function $\mathsf{s\text{-}rec}[a_0, r]$. It is easy to see that this is surjective.

Definition 12.7

For a set X, let $\mathcal{P}_{\text{ne}}(X)$ denote the set of non-empty subsets of X. That is,

$$\mathcal{P}_{\text{ne}}(X) = \{A \in \mathcal{P}(X) \mid \exists x \in X. x \in A\}$$

Lemma 12.5

There is a function $\mu \colon \mathcal{P}_{\text{ne}}(\mathbb{N}) \to \mathbb{N}$ so that for each $A \in \mathcal{P}_{\text{ne}}(\mathbb{N})$,

- $\mu(A) \in A$.
- if $n \in A$, then $\mu(A) \leq n$.

Proof: We define a relation $M\colon \mathcal{P}_{ne}(\mathbb{N}) \times \mathbb{N} \to 2$ by letting $A\ M\ n$ hold if and only if $n \in A$ and for all $m \in A$, $n \leq m$. That is, $A\ M\ n$ holds only for the least element of A, if such an element exists. Clearly, M is deterministic, because no set can have two distinct smallest elements.

By induction, we show that if $A \subseteq \mathbb{N}$ has no least element, then A must be empty, i.e., $k \notin A$ for all $k \in \mathbb{N}$.

- [Strong Inductive Hypothesis] Assume that for some k, it is the case that for each $j < k$, $j \notin A$.
- [Inductive Step] We claim that $k \notin A$. For otherwise, by the inductive hypothsis, all natural numbers less than k do not belong to A. So k would be the least element of A.

Hence for any non empty $A \subseteq \mathbb{N}$, there is a least element n. Thus there is an element n for which $A\ M\ n$. This shows that M is a functional relation. So it determines the function μ. $\square$

Lemma 12.6

A set A is countable if and only if there is a monomorphism $m\colon A \to \mathbb{N}$.

Proof: If $m\colon A \to \mathbb{N}$ is one-to-one, then either $A = \emptyset$, or there is an onto function $\mathbb{N} \to A$.

Suppose $A = \emptyset$, then the unique function from A to $\mathbb{N}$ is always monomorphism. Suppose $f\colon \mathbb{N} \to A$ is a surjection. Then f^{-} is a function from A to $\mathcal{P}_{ne}(\mathbb{N})$. Define $m = \mu \circ f^{-}$. That is, $m(a)$ is the smallest natural number for which $f(n) = a$. This is obviously a one-to-one function because in fact, $f(m(a)) = a$ for each a. $\square$

Lemma 12.7

If A and B are countable, then so is $A \times B$. If A and B are countable, then so is $A \cup B$. If I is countable and for each $i \in I$, A_i is countable, then $\bigcup_{i \in I} A_i$ is countable. If A is countable, then so is $\mathsf{List}[A]$.

Proof: The assumption that I is countable means there is a surjection $f\colon \mathbb{N} \to I$. Likewise, the assumption that each A_i is countable means there is a surjection $g_i\colon \mathbb{N} \to A_i$ for each $i \in I$. It suffices now to show that there is a surjection fromt $\mathbb{N} \times \mathbb{N}$ to $\bigcup_{i \in I} A_i$.

Let $h\colon \mathbb{N} \times \mathbb{N} \to \bigcup_{i \in I} A_i$ y defined by $h(m, n) = g_{f(m}(n)$. For $x \in \bigcup_{i \in I} A_i$, $x \in A_k$ for some $k \in I$. Since g_k is a surjection, $x = g_k(n)$ for some $n \in \mathbb{N}$. But likewise, f is a surjection, so $k = f(m)$ for some $m \in \mathbb{N}$. Hence $x = g_{f(m)}(n) = h(m, n)$. $\square$