Model Theory and First-order Logic Minicourse

Franklin

July 29, 2022

1 My "Reduced Inventory"

At the beginning of the minicourse I shared my own personal set of axioms for the integers:

(A1)	$\forall x\ \forall y\ x + y = y + x$	Addition is commutative
(A2)	$\forall x\ \forall y\ x \cdot y = y \cdot x$	Multiplication is commutative
(A3)	$\forall x\ \forall y\ \forall z\ (x + y) + z = x + (y + z)$	Addition is associative
(A4)	$\forall x\ \forall y\ \forall z\ (x \cdot y) \cdot z = x \cdot (y \cdot z)$	Multiplication is associative
(A5)	$\forall x\ x + 0 = x$	Zero element

But it was pretty clear the audience did not approve of my set of axioms. Someone pointed out that I'm missing an axiom expressing the fact that multiplication on $\mathbb{Z}$ is *distributive* over addition. I really didn't want to add this to my set of axioms, though - after all, maybe it could be proven from these five axioms, if only I was clever enough to come up with the right proof! And I certainly wouldn't want to add a sentence that was already provable to my inventory, in the interest of keeping it as "reduced" as possible.

Our prospects of proving distributivity were quickly shattered, though. If we define a kind of "funny multiplication function" $*$ on $\mathbb{Z}$ by letting

$$x * y = 1$$

for all $x, y \in \mathbb{Z}$, notice that this "alternative multiplication" satisfies all of our axioms that concern multiplication, namely (A2) and (A4). In other words, the binary operation $*$ is both commutative and associative. However, it does not distribute over addition: for instance,

$$1 * (2 + 2) = 1$$

$$1 * 2 + 1 * 2 = 2$$

meaning that $1 * (2 + 2) \neq 1 * 2 + 1 * 2$. If it were possible to prove distributivity just using the axioms (A1) through (A5), then it would be possible to prove that $*$ is distributive as well, since the integers with ordinary addition and "funny multiplication" also satisfy (A1) through (A5). But in this system, "funny multiplication" *definitively does not* distribute over addition! So it must not be the case that distributivity can be proven from (A1) through (A5). How tragic.

Perhaps this can be remedied by just adding another axiom to our system. For instance, maybe if we add an extra axiom for distributivity, we'll have a set of axioms which is capable of proving all the interesting facts about $\mathbb{Z}$ which we'd like to prove:

(A1)	$\forall x\ \forall y\ x + y = y + x$	Addition is commutative
(A2)	$\forall x\ \forall y\ x \cdot y = y \cdot x$	Multiplication is commutative
(A3)	$\forall x\ \forall y\ \forall z\ (x + y) + z = x + (y + z)$	Addition is associative
(A4)	$\forall x\ \forall y\ \forall z\ (x \cdot y) \cdot z = x \cdot (y \cdot z)$	Multiplication is associative
(A5)	$\forall x\ x + 0 = x$	Zero element
(A6)	$\forall x\ \forall y\ \forall z\ x \cdot (y + z) = (x \cdot y) + (x \cdot z)$	Multiplication distributes over addition

Sadly, it seems there exist systems satisfying these axioms as well which are very different from the integers we know and love. In particular, if we consider the trivial ring $\{0\}$ with addition and multiplication defined as usual, we can check these axioms one by one and verify that each of (A1) through (A6) holds true in this ring! But clearly there are some very important facts about $\mathbb{Z}$ that aren't true in the trivial ring, such as the existence of anything other than the zero element, not to mention any of the nontrivial things that we'd like to prove about $\mathbb{Z}$. So it seems a few more additions to our reduced inventory are in order. Let's postulate the existence of a multiplicative identity, and that this multiplicative identity is distinct from zero:

(A1)	$\forall x\ \forall y\ x + y = y + x$	Addition is commutative
(A2)	$\forall x\ \forall y\ x \cdot y = y \cdot x$	Multiplication is commutative
(A3)	$\forall x\ \forall y\ \forall z\ (x + y) + z = x + (y + z)$	Addition is associative
(A4)	$\forall x\ \forall y\ \forall z\ (x \cdot y) \cdot z = x \cdot (y \cdot z)$	Multiplication is associative
(A5)	$\forall x\ x + 0 = x$	Zero element
(A6)	$\forall x\ \forall y\ \forall z\ x \cdot (y + z) = (x \cdot y) + (x \cdot z)$	Multiplication distributes over addition
(A7)	$\forall x\ x \cdot 1 = x$	Multiplicative identity
(A8)	$\neg(1 = 0)$	Zero and one are distinct

But alas! It seems that these properties are also insufficient to characterize the integers uniquely. For instance, the set of real numbers $\mathbb{R}$ with addition and multiplication as it is usually defined also satisfies all each of these axioms. So if, say, we wanted to prove from these axioms that there exist nonzero elements in the integers without multiplicative inverses:

$$\exists x\ \neg(x = 0) \wedge \neg(\exists y\ x \cdot y = 1)$$

we would know instantly that proving such a thing is impossible: for this isn't true at all in the system $\mathbb{R}$, even though $\mathbb{R}$ satisfies axioms (A1) through (A8).

In general, it we assemble a set of sentences ("axioms") which we think encapsulate all the essential properties of a system that we care about, such as $\mathbb{Z}$, we might ask ourselves the questions "What theorems are we capable of proving with these axioms? Can we prove all of the sentences that are true about $\mathbb{Z}$? Do these axioms distinguish $\mathbb{Z}$ from all other rings?" The line of attack that a model theorist takes to answering these questions is to search for other algebraic systems that satisfy the same axioms, which are called **models** of those axioms. Then, if we find some alternative system satisfying all the axioms from our inventory, but which fails to satisfy a sentence ψ that is satisfied in $\mathbb{Z}$, then we know that our axioms are incapable of proving ψ, and therefore incapable of proving everything there is to prove about $\mathbb{Z}$ and distinguish it from other models. The techniques developed by model theorists lead to some profound philosophical insights, as well as some *really bizarre* algebraic structures.

2 Syntax: First-order Logic

In model theory, when we refer to a "theorem", we are really referring to a sentence in **first-order logic**, which is very different from writing theorems in plain English. In particular, there is a very limited set of symbols that we are allowed to use to formulate a sentence in first-order logic. These symbols are:

1. Variable names, like $x_1, x_2, x_3, \cdots$ or $x, y, z, \cdots$
2. Logical operators, like $\wedge$ "and", $\vee$ "or", $\neg$ "not", and $\rightarrow$ "implies"
3. Quantifiers, i.e. the **universal quantifier** "for all" $\forall$ and the **existential quantifier** "there exists" $\exists$
4. The equality sign $=$
5. Parentheses () used for disambiguation

Not just any combination of these symbols is allowed. For instance, something like $\forall\forall\exists\forall\neg\neg()$ is not a "grammatically correct" sentence in first-order logic. There's a very specific set of rules used to describe how one is allowed to build a "grammatically correct" sentence, but we won't go into it here. (If you want to see a detailed treatment, see Chang and Kiesler's *Model Theory: Third Edition*, which is where I learned about all this.) Notice that there is a difference between "nonsense" like $\forall\forall\exists\forall\neg\neg()$, which is not even a sentence, and sentences that are *always false*, like the following:

$$\exists x\ \neg(x = x)$$

We can make sense of the above sentence: intuitively, it claims that there exists an x that is not equal to itself. Of course, this is false no matter what kind of structure we consider, but as a sentence, it does *make sense* - it's just not true. Similarly, there are sentences that are always true, such as

$$\forall x\ x = x$$

Such sentences are called **valid**, whereas sentences that are always false are called **refutable**.

Of course, we cannot say very many interesting things using this bare-bones set of symbols given to us by first-order logic. When doing model theory, we usually consider an additional **language** which introduces a few more symbols. These symbols can be of one of three different types, which determine how they are used syntactically:

1. **Function symbols**, each of which has an **arity** describing how many arguments it takes. For instance, $+$ is a 2-ary function symbol in our language for arithmetic, since it represents a function of two arguments. The negative sign $-$ could be a 1-ary function symbol, since it represents a function of one arguments that maps a number to its additive inverse.

2. **Relation symbols**, each of which also has an arity describing the number of arguments it takes. For example, $\leq$ is a 2-ary relation symbol in our language for arithmetic, since it represents a relation that takes two numbers and determines whether or not the first number is less than or equal to the second number. In the language of set theory, the symbol $\in$ is also a 2-ary (or, binary) relation symbol.

3. **Constant symbols**, which represent specific elements of the system under consideration. For example, 0 and 1 might be constant symbols in our language for arithmetic, which we interpret as the additive and multiplicative identities respectively.

There is a lot of technical detail that goes into defining how these symbols interact with each other and with the "general purpose" first-order logic symbols. We won't go into the tedious inductive definition, but here's a small example. If our language is given by $\mathcal{L} = \{f, g, R, S, c_1, c_2\}$, where f is a unary function symbol, g is a ternary function symbol, R is a ternary relation symbol, S is a binary relation symbol, and c_1, c_2 are constant symbols, then the following would be an example of a well-formed sentence in this language:

$$\forall x\ (\exists y\ S(g(x, y, c_1), g(x)) \to S(x, c_2) \wedge R(y, c_1, c_2))$$

What on earth is this sentence saying? Who knows - it could mean many things, depending on what exactly the symbols f, g, R, G, c_1, c_2 are chosen to represent, or how they're "interpreted". Syntactically, this is just a sequence of symbols. In the next section, we'll see how they interact with the "meaning" that is assigned to them using models.

For a few actual examples of languages that might be used to formalize certain mathematical theories, consider the following:

- $\mathcal{L} = \{+, \cdot, -, \leq, 0, 1\}$ is a language that might be used to talk about arithmetic. It has two binary function symbols, one unary function symbol, one binary relation symbol, and two constant symbols.

- $\mathcal{L} = \{\leq\}$ is a language that might be used to describe the ordering on an ordered set.

- $\mathcal{L} = \{\in\}$ is the language that is often used for doing formal set theory, for instance *Zermelo-Fraenkel Set Theory with Choice*, sometimes abbreviated *ZFC*, in which $\in$ is used to represent the "membership" relation.
- $\mathcal{L} = \{\}$ is the **pure identity language**. It has no function symbols and no constant symbols, and the only relation symbol that we have access to in this language is the equals sign $=$ (which is part of the "standard language" for doing first-order logic anyways).

In addition to the long list of rules describing how "grammatically correct" sentences of the language $\mathcal{L}$ can be formed, there is also a set of rules describing how sentences can be formally manipulated to write proofs. The process of logical inference can be described purely algorithmically using this formal language, so that you could be given a set of symbols that mean nothing to you whatsoever, and yet you could nevertheless write a proof of an elaborate sentence simply by mechanically applying the rules of inference. Once again, I won't list all of the inference rules of first-order logic (see Chang and Kiesler if you'd like more detail).

3 Semantics: Models

Moving beyond just first-order logic, where a sentence is merely a sequence of symbols, we may begin to look at *models*, where sentences take on "meaning". Given a language $\mathcal{L}$, consisting of some function, relation and constant symbols, a **model** for that language consists of the following data:

1. An underlying set A, called the **universe**, which will contain the *elements* of the model
2. For each n-ary function symbol of $\mathcal{L}$, an *actual function* $A^n \to A$ corresponding to that symbol
3. For each m-ary relation symbol of $\mathcal{L}$, an *actual relation* $\subset A^m$ corresponding to that symbol
4. For each constant symbol of $\mathcal{L}$, an element $\in A$ corresponding to that symbol

For each function symbol of the language $\mathcal{L}$, the actual function corresponding to it is called the **interpretation** of that symbol in the given model. Similarly, the relation corresponding to a given relation symbol is called its *interpretation*, and the same goes for the element of the universe corresponding to a certain constant symbol. If we have a language

$$\mathcal{L} = \{f_1, f_2, \cdots, R_1, R_2, \cdots, c_1, c_2, \cdots\}$$

where $f_1, f_2, \cdots$ are function symbols (each with their own arity), and $R_1, R_2, \cdots$ are relation symbols (also with their own arity), and $c_1, c_2, \cdots$ are constant symbols, then we might denote a model of this language as follows:

$$\mathfrak{A} = \langle A, g_1, g_2, \cdots, S_1, S_2, \cdots, a_1, a_2, \cdots \rangle$$

where A is a universe of the model, $g_1, g_2, \cdots$ are the functions corresponding to the symbols $f_1, f_2, \cdots$, and $S_1, S_2, \cdots$ are the relations corresponding to the symbols $R_1, R_2, \cdots$, and finally $a_1, a_2, \cdots \in A$ are the elements of the universe corresponding to the constant symbols $c_1, c_2, \cdots$.

To continue a previous example, if we consider the familiar language

$$\mathcal{L} = \{+, \cdot, \leq, 0, 1\}$$

our "typical model" for this language is the integers with traditional integer arithmetic:

$$\langle \mathbb{Z}, +, \cdot, \leq, 0, 1 \rangle$$

However, there are other models of the same language. For instance, the following is also a model of the same language:

$$\langle \mathbb{Z}, +, f, \leq, 0, 1 \rangle$$

where $f : \mathbb{Z}^2 \to \mathbb{Z}$ is defined by $f(x, y) = 1$. This is our "pathological multiplication" function that we defined earlier to show that distributivity was not provable from (A1) through (A5). We also have the following model:

$$\langle \{0\}, +, \cdot, \leq, 0, 1 \rangle$$

This is the trivial ring, which we used earlier to show that the distinctness of the multiplicative identity and the additive identity does not follow from (A1) through (A7). And finally, the following is yet another model of the same language:

$$\langle \mathbb{R}, +, \cdot, \leq, 0, 1 \rangle$$

which is the ring of real numbers, used to prove that many important properties of $\mathbb{Z}$ cannot be proven using merely (A1) through (A8). Of course, there are many more models of this language. We could use something as strange as

$$\langle \mathbb{N}, \cdot, \cdot, \mid, 49, 50 \rangle$$

Of course, this model would not satisfy very many of the axioms from our list earlier, but it would nevertheless be a model of the language $\mathcal{L}$.

One of the most significant problems we're concerned with as model theorists is how to distinguish between two different mathematical structures using our language. That is, given two *models* of a language, can we come up with some sentence which holds true in one of the models, but not in the other, thereby setting them apart from a first-order logic standpoint? For instance, consider the following two models of the language considered above:

$$\mathfrak{A} = \langle \mathbb{Z}, +, \cdot, \leq, 0, 1 \rangle$$

$$\mathfrak{B} = \langle \mathbb{Q}, +, \cdot, \leq, 0, 1 \rangle$$

Can we find a first-order sentence that discriminates between integer arithmetic and rational arithmetic? The answer is yes, and it may not take you long to come up with such a sentence. An example is the following sentence:

$$\varphi = \big(\forall x \ \exists y \ y + y = x\big)$$

In plain english, ϕ says that for all x, there exists y such that $y + y = x$. In other words, we can "halve" any element of the universe. Notice that φ is satisfied in $\mathfrak{B}$, the rational numbers, but it is *not satisfied* in $\mathfrak{A}$, the integers. The model-theoretic way of saying this would be that $\mathfrak{A}$ **is not a model of** φ, whereas $\mathfrak{B}$ **is a model of** φ. As another example, consider the following pair of models:

$$\mathfrak{A} = \langle \mathbb{Q}, +, \cdot, \leq, 0, 1 \rangle$$

$$\mathfrak{B} = \langle \mathbb{R}, +, \cdot, \leq, 0, 1 \rangle$$

How can we distinguish between the rationals and the reals in terms of their first-order properties? One way is to use the fact that every nonnegative real number has a real square root, but it is not the case that every nonnegative rational number has a rational square root. We can formulate this in first-order logic as follows:

$$\psi = \big(\forall x \ (0 \leq x) \to (\exists y \ y \cdot y = x)\big)$$

so that $\mathfrak{A}$ is not a model of ψ, but $\mathfrak{B}$ *is* a model of ψ.

We might also consider trying to distinguish between these algebraic structures in a simplified language. For instance, we might consider the smaller language $\mathcal{L} = \{+, \leq\}$, and try to distinguish between some of the above algebraic structures just in terms of how their addition and inequalities behave. (That is, we are treating them as *ordered groups* rather than *ordered rings*.) If we now consider

$$\mathfrak{A} = \langle \mathbb{Z}, +, \leq \rangle$$

$$\mathfrak{B} = \langle \mathbb{Q}, +, \leq \rangle$$

and try to distinguish between these two models, it appears that we don't have much work to do, since the sentence φ that we used to distinguish between these models before only used the symbol $+$

anyways. Hence, φ is also a model in this reduced language, so we may again use it to distinguish $\mathfrak{A}$ from $\mathfrak{B}$. What about our second example?

$$\mathfrak{A} = \langle \mathbb{Q}, +, \leq \rangle$$

$$\mathfrak{B} = \langle \mathbb{R}, +, \leq \rangle$$

Can we find a first-order sentence that holds in one of these models, but not the other?

We might simplify our language even further to consider just $\mathcal{L} = \{\leq\}$, and try to use this language to distinguish between different *ordered sets* in terms of their first-order properties. For instance, using this language, can we distinguish between the ordering of the natural numbers and the integers?

$$\mathfrak{A} = \langle \mathbb{N}, \leq \rangle$$

$$\mathfrak{B} = \langle \mathbb{Z}, \leq \rangle$$

Certainly, for the orderings of these two sets are very different. For instance, $\mathbb{N}$ has a smallest element and $\mathbb{Z}$ does not, so we can use the sentence

$$\zeta = \big(\exists x \ \forall y \ x \leq y\big)$$

to discriminate between these models, since $\mathfrak{A}$ is a model of ζ but $\mathfrak{B}$ is not. What about the ordering of the integers and the ordering of the rational numbers?

$$\mathfrak{A} = \langle \mathbb{Z}, \leq \rangle$$

$$\mathfrak{B} = \langle \mathbb{Q}, \leq \rangle$$

This one is a little bit trickier, but after a bit of investigation, we might use the fact that $\mathbb{Q}$ is a *dense* ordering, that is, between any two elements of $\mathbb{Q}$ lies another element of $\mathbb{Q}$. This is not the case for $\mathbb{Z}$, since, for instance, there are no integers between 0 and 1 (as you very well know by now). Let's write this as a first-order sentence:

$$\xi = \big(\forall x \ \forall y \ \neg(y \leq x) \to (\exists z \ \neg(z \leq x) \wedge \neg(y \leq z))\big)$$

Note that we're using $\neg(y \leq x)$ as a kind of "workaround" to represent the strict inequality $x < y$ using only the one symbol $\leq$ that is available to us.

4 Completeness and compactness

And now, here's a brief treatment of two of my favorite theorems from model theory. Both of their statements seem kind of mundane at first glance, but they can be used to come up with some strange and bizarre examples of models that "masquerade" as models that we are very familiar with, while actually having huge structural differences that are untouchable by first-order logic.

Before introducing these theorems, we first need to know what it means for a set of sentences to be *consistent*. Given a fixed language $\mathcal{L}$ and a set of sentences Σ (possibly infinite) of that language, we say that Σ is **consistent** if it is not possible to prove a contradiction from the sentences in Σ. That is, if we take the sentences in Σ as assumptions, it is not possible to construct a proof of a sentence taking the form $\varphi \wedge \neg\varphi$. If it *is possible* to prove a contradiction using premises from Σ, then we say that Σ is **inconsistent**. It makes sense that we would care about studying which sets of sentences are consistent and which ones are inconsistent, because if we are trying to construct a set of axioms for some mathematical system like integer arithmetic or number theory, we would not want our theory to contain any contradictions, so it is important that our set of axioms be *consistent*.

If you pick a sentence φ over $\mathcal{L}$, the set $\Sigma = \{\varphi, \neg\varphi\}$ is an obvious example of an inconsistent set of sentences, and a slightly less obvious example is the set

$$\Sigma = \{(\varphi \to \varphi) \to \varphi, \ \varphi \to (\neg\varphi)\}$$

However, it is not always so easy to tell whether a set of sentences over some language is inconsistent just by inspection. Consider, for instance, the following set of senteces over the language $\mathcal{L} = \{f, g, c, d, e\}$ where f, g are binary function symbols and c, d, e are constant symbols:

$$\begin{array}{l|l}
\varphi_1 & \forall x\ \forall y\ (f(x,y) = f(y,x)) \wedge (g(x,y) = g(y,x)) \\
\varphi_2 & \forall x\ \forall y\ \forall z\ (f(x,f(y,z)) = f(f(x,y),z)) \wedge (g(x,g(y,z)) = g(g(x,y),z)) \\
\varphi_3 & \forall x\ (f(x,c) = x) \wedge (g(x,d) = x) \wedge (f(x,e) = x) \wedge \neg(c = e) \\
\varphi_4 & \forall x\ \forall y\ \forall z\ g(x,f(y,z)) = f(g(x,y),g(x,z)) \\
\varphi_5 & \forall x\ \exists y\ f(x,y) = c
\end{array}$$

This set of sentences is actually inconsistent. In fact, some pair of them φ_i, φ_j is inconsistent. Can you figure out which two sentences are inconsistent? And is there something familiar about these sentences, even though the function symbols f, g and the constant symbols c, d, e are unfamiliar?

Of course, if a set of sentences Σ is inconsistent, then it cannot have any models, for if $\varphi \wedge \neg\varphi$ is a contradiction implied by the sentences in Σ, then any model $\mathfrak{A}$ of Σ would have to satisfy $\varphi \wedge \neg\varphi$. But this is impossible, because a model cannot satisfy both φ and $\neg\varphi$ simultaneously. Hence, inconsistent sets of sentences have no models. But is the converse true? That is, if a set of sentences Σ *does not* give rise to a logical contradiction, does there necessarily exist a model that satisfies all of the sentences $\sigma \in \Sigma$? The **Completeness Theorem** answers this question in the affirmative.

Theorem (Completeness Theorem)**.** *If Σ is a consistent set of first-order sentences in some language $\mathcal{L}$, then it has a model.*

This theorem seems rather tame, but as it turns out, the proof is rather technical. For details, you can find a full proof at the beginning of Chapter 2 of Chang and Kiesler's *Model Theory: Third Edition.* Although it seems almost obvious, this theorem has an interesting philosophical interpretation as well:

"Logical consistency is equivalent to metaphysical possibility."

That is, if some set of assertions isn't logically contradictory, then it is "metaphysically possible", that is, there is some "possible world/universe" (i.e. model) in which those sentences all hold true. Given a consistent set of sentences, it's not at all obvious how to actually construct a set of functions, relations and constants corresponding to the symbols of the language that simultaneously satisfies all of them - hence the technicality of the proof.

The true power of the Completeness Theorem is unlocked through its partnership with the **Compactness Theorem**:

Theorem (Compactness Theorem)**.** *If Σ is a set of first-order sentences such that every finite subset $\Sigma' \subset \Sigma$ is consistent, then Σ is consistent.*

This one is far less intuitive. If we have an infinite set of sentences Σ and we want to prove that it is consistent, it suffices to prove that each finite subset of Σ is consistent. But why? Well, suppose for the sake of contradiction that Σ were *inconsistent.* Then there would have to exist some proof of a contradiction $\varphi \wedge \neg\varphi$ using the sentences from Σ as hypotheses. But every proof consists of a finite sequence of logical inferences - so any proof using sentences from Σ as hypotheses could only make use of at most finitely many sentences from Σ. If we let Σ' be the set of sentences of Σ that are used in a given proof of $\varphi \wedge \neg\varphi$, then we have that Σ' is finite, and *also inconsistent*, since the sentences of Σ' can be used to prove $\varphi \wedge \neg\varphi$. Therefore any inconsistent set of sentences Σ has an inconsistent finite subset $\Sigma' \subset \Sigma$. By contrapositive, if every finite subset $\Sigma' \subset \Sigma$ is consistent, then Σ must be consistent!

Okay, so how are these theorems used together? Here's an illuminating example. We have already considered the model of the language $\mathcal{L} = \{+, \cdot, \leq, 0, 1\}$ consisting of the set of integers with the ordinary interpretations of each of these symbols, i.e. the "standard model":

$$\mathfrak{Z} = \langle \mathbb{Z}, +, \cdot, \leq, 0, 1 \rangle$$

Let us now consider a slightly modified language, in which we add one additional constant symbol: $\mathcal{L}' = \{+, \cdot, \leq, 0, 1, c\}$. Notice that if we pick any integer $n \in \mathbb{Z}$, then

$$\langle \mathbb{Z}, +, \cdot, \leq, 0, 1, n \rangle$$

is a model of this modified language, in which the *interpretation* of the symbol c is the element $n \in \mathbb{Z}$. Now, let us take all of the sentences σ of the original language $\mathcal{L}$ that are true in the model $\mathfrak{Z}$ and

collect them together into a set of sentences Σ. Certainly Σ is a *consistent* set of sentences of $\mathcal{L}$, since $\mathfrak{Z}$ is a model of this set of sentences, and no inconsistent set of sentences can have a model. Notice that Σ is also a set of sentences of the modified language $\mathcal{L}'$ - it just happens that none of these sentences uses the newly added constant symbol c. Now, let us define another set of sentences $\Phi = \{\varphi_1, \varphi_2, \varphi_3, \cdots\}$ of the language $\mathcal{L}'$ as follows:

φ_1	$1 \leq c$
φ_2	$1 + 1 \leq c$
φ_3	$1 + 1 + 1 \leq c$
φ_4	$1 + 1 + 1 + 1 \leq c$
$\cdots$	$\cdots$

The first sentence φ_1 states that the interpretation of c is greater than one, the second sentence φ_2 states that it is greater than two, and so on. Notice that if we choose some $n \in \mathbb{Z}$, the model

$$\langle \mathbb{Z}, +, \cdot, \leq, 0, 1, n \rangle$$

is *never* a model of all of the sentences of Φ, since there is no integer that is simultaneously greater than 1, and greater than $1 + 1$, and greater than $1 + 1 + 1$, and so on.

Finally, let us consider the set of sentences $\Sigma \cup \Phi$. This consists of all sentences of $\mathcal{L}'$ that do not use the symbol c and are satisfied in $\langle \mathbb{Z}, +, \cdot, \leq, 0, 1 \rangle$, together with all of the sentences φ_i described above. It's not immediately obvious whether $\Sigma \cup \Phi$ is consistent or inconsistent - after all, the models $\langle \mathbb{Z}, +, \cdot, \leq, 0, 1, n \rangle$ don't satisfy all of their sentences, so we don't immediately know that $\Sigma \cup \Phi$ is consistent. This is where the Compactness Theorem comes in. By Compactness, if every finite subset of $\Sigma \cup \Phi$ is consistent, then the whole set is consistent. So let us suppose that $\Psi \subset \Sigma \cup \Phi$ is a finite subset of this set of sentences, so that $\Psi = \{\psi_1, \cdots, \psi_k\}$. Since at most finitely many of the sentences φ_i appears in Ψ, there must exist a maximum N such that $\varphi_N \in \Psi$. But if N is this maximum value, then we have that

$$\langle \mathbb{Z}, +, \cdot, \leq, 0, 1, N \rangle$$

is a model of Ψ: all of the sentences of Ψ taken from Σ must be true in this model because they are true in $\langle \mathbb{Z}, +, \cdot, \leq, 0, 1 \rangle$; and all of the sentences of Ψ taken from Φ must be true in this model because they all take the form φ_i with $i \leq N$ by the definition of N, and so the integer $N \in \mathbb{Z}$ is greater than or equal to i for all $\varphi_i \in \Psi$. Therefore, since every finite subset $\Psi \subset \Sigma \cup \Phi$ is consistent, we have that all of $\Sigma \cup \Phi$ is consistent by Compactness.

Now we are ready to apply Completeness. Since $\Sigma \cup \Phi$ is consistent, we can say that there exists a model satisfying all of the sentences of $\Sigma \cup \Phi$. In this model, *every theorem of first-order logic that is true about the ordinary integers* also holds, yet *there exists an element of this model that is greater than* 1*,* 2*,* 3*, and so on.* If $\langle A, f, g, R, 0_A, 1_A, a \rangle$ is this model, we can throw away the extra constant $a \in A$ to obtain another model of our original language $\mathcal{L}$:

$$\mathfrak{Z}' = \langle A, f, g, R, 0_A, 1_A \rangle$$

Every sentence that is true of $\mathfrak{Z}$ (i.e. the sentences of Σ) is also true of $\mathfrak{Z}'$ and vice versa. (In model-theoretic terminology, we say that $\mathfrak{Z}$ and $\mathfrak{Z}'$ are **elementarily equivalent** and write $\mathfrak{Z} \equiv \mathfrak{Z}'$ to denote this relationship.) Yet somehow $\mathfrak{Z}'$ also contains an "infinite integer" that is greater than $1, 2, 3$, and so on, despite being indistinguishable from $\mathfrak{Z}$ by any first-order sentence. If we are concerned with being able to distinguish between mathematical systems with different structures using first-order sentences, this is a nightmare. How could two structures that are so different have precisely the same theorems?

5 Exercises and teasers

1. Can you distinguish between the models $\langle \{1, 2\} \rangle$ and $\langle \{1, 2, 3\} \rangle$ in the pure identity language? How about $\langle 1, 2, 3 \rangle$ and $\langle \mathbb{N} \rangle$? How about $\langle \mathbb{N} \rangle$ and $\langle \mathbb{R} \rangle$?

2. In the language $\mathcal{L} = \{\leq\}$, can you distinguish between the following pairs of models?

   $\mathfrak{A} = \langle [0,1], \leq \rangle \quad \mathfrak{B} = \langle [0,2], \leq \rangle$
   $\mathfrak{A} = \langle [0,1], \leq \rangle \quad \mathfrak{B} = \langle [0,1] \cup [2,3], \leq \rangle$
   $\mathfrak{A} = \langle [0,1], \leq \rangle \quad \mathfrak{B} = \langle [0,1) \cup (2,3], \leq \rangle$
   $\mathfrak{A} = \langle \mathbb{N}, \leq \rangle \quad \mathfrak{B} = \langle \mathbb{N} + \mathbb{Z}, \leq \rangle$

3. A **formula** in a language $\mathcal{L}$ is similar to a sentence, except that it may have **free variables**. An example in the language of a ring is the formula $\varphi(x) = (\exists y\ y + y = x)$, which in plain English says "x is the double of some number". A formula is not *satisfied* or *not satisfied* in a given model, since its truth value may depend on the input x. For instance, in $\langle \mathbb{Z}, +, \cdot, 0, 1 \rangle$, we have that $\varphi(2)$ is true but $\varphi(3)$ is false, and we say that 2 **satisfies** φ but 3 **fails to satisfy** φ. In the model $\langle \mathbb{Q}, +, \cdot, 0, 1 \rangle$, both 2 and 3 satisfy φ.

   Given a model $\mathfrak{A}$ with universe A, and a subset $S \subset A$, we say that S is **definable** if there exists a formula φ with one free variable such that S is the set of all $a \in A$ satisfying φ. Similarly, we say that an element $a \in A$ is *definable* if there exists a formula φ such that a is the unique element of A satisfying φ, i.e. if the set $\{a\} \subset A$ is a definable subset. For instance, in the model $\langle \mathbb{Z}, +, \cdot, 0, 1 \rangle$, the element 2 is definable because it is defined by $\varphi(x) = (x = 1 + 1)$, and the subset $S \subset \mathbb{Z}$ of even numbers is definable because it is the set of all integers satisfying $\psi(x) = (\exists y\ y + y = x)$.

   Here are some questions to explore on your own:

   (a) Using the language $\mathcal{L} = \{+, \cdot, 0, 1\}$, what subsets of the model $\langle \mathbb{Z}, +, \cdot, 0, 1 \rangle$ can you define? We have already defined the set of even numbers... can you define the set of integers divisible by 3? The set of integers $\equiv 1 \bmod 7$? The set of perfect squares? The set of primes? The set of powers of 2?

   (b) Using the smaller language $\mathcal{L} = \{+\}$, what subsets of the model $\langle \mathbb{Z}, + \rangle$ can you define? Can you define the set of perfect squares? Can you define the subset $\mathbb{N} \subset \mathbb{Z}$? Can you define the element $0 \in \mathbb{Z}$? Can you define the element $1 \in \mathbb{Z}$?

   (c) Given a model $\langle A \rangle$ of the pure identity language, can you determine all of the definable subsets of A?

   (d) Can you prove that if $\mathcal{L}$ is finite and $\mathfrak{A} = \langle A, \cdots \rangle$ is a model whose universe is a countably infinite set, then there exists an undefinable subset of A? Does there necessarily exist an undefinable element of A?

   (e) Can you find a finite language and a model of that language where all elements are undefinable? Where no elements are undefinable? Where exactly one element is undefinable? Where exactly two elements are undefinable?

4. Consider the language $\mathcal{L} = \{+, \cdot, \leq, 0, 1\}$ and the model $\mathfrak{R} = \langle \mathbb{R}, +, \cdot, \leq, 0, 1 \rangle$. The real numbers are an *Archimedean Field*, meaning that for any real number $\epsilon > 0$, there exists a real number $N \in \mathbb{R}$ in the form $N = 1 + 1 + \cdots + 1$ such that $N\epsilon > 1$. Can you prove using Compactness and Completeness that there exists a model $\mathfrak{R}'$ of the same language which has all the same true sentences as $\mathfrak{R}$ (i.e. $\mathfrak{R}' \equiv \mathfrak{R}$) and yet *fails to be Archimedean*?

5. Consider the language $\mathcal{L} = \{\leq\}$ and the model $\mathfrak{A} = \langle \mathbb{N}, \leq \rangle$. You have learned that the natural numbers are *well-ordered*, meaning that every subset of $\mathbb{N}$ has a least element. Can you prove using Compactness and Completeness that there is a model $\mathfrak{A}'$ having all the same true sentences as $\mathfrak{A}$ (i.e. $\mathfrak{A}' \equiv \mathfrak{A}$) such that the universe of $\mathfrak{A}'$ is *not well-ordered*? Can you actually construct such a model, in addition to showing that it exists?

If you solve any of these exercises, come tell me about your solution! If you fail to solve any of these exercises, tell me about your failure!

If you enjoyed the minicourse and notes, and are interested in learning more about model theory, here are some more things you might like to read about:

- I learned model theory from Chang and Kiesler's *Model Theory: Third Edition*, which I would highly recommend to anyone who has some proof background and enough patience to slowly work through pages of dense proofs. For context, it took me about a month and a half to work through the first one-and-a-half chapters of this book - it's very dense! But in my opinion, it's well-written and the exercises are really interesting. Also, you can find all of my model theory notes and some of my solutions to exercises on my webiste!

- Another pair of theorems from model theory with profound philosophical implications are the **Löwenheim-Skolem Theorems**, which address the possible cardinalities that the universe of a model satisfying certain sentences can have. In particular, they say that given a model of a finite language $\mathcal{L}$, there always exists an elementarily equivalent model with *any given cardinality*. For instance, there exists a model with all the same true sentences as $\langle \mathbb{Z}, +, \cdot, \leq, 0, 1 \rangle$, but with an uncountably infinite universe! This is pretty unbelievable.

- In ZFC (a common set of axioms for set theory), the **Continuum Hypothesis (CH)** states that there does not exist a set A whose cardinality is strictly between the cardinality of $\mathbb{N}$ and the cardinality of $\mathbb{R}$, so that $|\mathbb{N}| < |A| < |\mathbb{R}|$. That is, $|\mathbb{R}|$ is the "next biggest" cardinality after $|\mathbb{N}|$. It turns out that this is unprovable in ZFC, and in fact there exist both models of ZFC where CH is true, and models where CH is false. This can be proven using an advanced model-theoretic technique called **forcing**.

- There is a really cool technique for explicitly defining "weird counterexample models" using a mathematical construction called an **ultraproduct**, which makes use of a set-theoretic structure called an **ultrafilter**. These are really advanced topics, but in my opinion one of the most amazing and mind-boggling ideas in model theory.