Encoding Mathematics in First-Order Logic

Over the past few lectures, we have seen the syntax and the semantics of first-order logic, along with a proof system based on tableaux. It is of course straightforward to adapt the tableaux procedure into a Gentzen sequent system and develop a proof system in the style of refinement logic (see Smullyan XI.2).

In this lecture, we look at how to use first-order logic to reason formally about a particular domain of interest. Because it provides for a number of almost ready-made examples, we shall mostly look at how to formalize mathematical reasoning, with the understanding that what I say here applies equally well to other domains.

In order to do this right, we need two minor extensions to the first-order logic framework we have studied until now, both rather standard. The first extension consists of adding *function symbols* to the logic, and the second extension consists of adding an *equality* predicate.

- Functions symbols capture functions over individuals in the universe. Currently, atomic formulas are of the form $P(c_1, \ldots, c_n)$, where P is a predicate symbol, and $c_1, \ldots, c_n$ are variables or parameters. Now, we consider atomic formulas to be of the form $P(t_1, \ldots, t_n)$, where P is a predicate symbol of arity n, and $t_1, \ldots, t_n$ are *terms*. The set of terms is defined inductively as follows:

  1. A parameter a is a term;
  2. A variable x is a term;
  3. if $t_1, \ldots, t_n$ are terms and f is a function symbol of arity n, then $f(t_1, \ldots, t_n)$ is a term.

  Recall that the semantics of first-order logic is given in terms of an interpretation $I = (U, \varphi, \iota)$, where U is the universe of individuals, φ assigns to every parameter a an element $\varphi(a) \in U$ of the universe, and ι assigns to every n-ary predicate symbol P a relation $\iota(P) \subseteq U^n$. To account for function symbols, we extend φ so that it maps every n-ary function symbol f to a function $\varphi(f)$ from U^n to U. With this, it is easy to see that every closed term t represents a unique element $[t]$ of U: $[a]$ is $\varphi(a)$, while $[f(t_1, \ldots, t_n)]$ is $(\varphi(f))([t_1], \ldots, [t_n])$. With this, an atomic sentence $P(t_1, \ldots, t_n)$ is true under I if $([t_1], \ldots, [t_n]) \in \iota(P)$. The rest of the semantics of first-order logic is unchanged.

- An equality predicate is a binary predicate symbol $=$, typically written $t_1 = t_2$ rather than $= (t_1, t_2)$. First-order logic with equality is an extension of first-order logic with such an equality predicate, where the interpretation of $=$ is fixed as follow: it is the identity relation on U, the universe of the interpretation. In other words, the atomic formula $t_1 = t_2$ is true under an intepretation if $[t_1]$ and $[t_2]$ are the same individual. While this seems rather innocuous, we need to add new proof rules corresponding to this new predicate to our tableau procedure, since there are new valid formulas in first-order logic with equality. For instance, the following formula is valid:

  $$(x = x') \wedge (y = y') \Rightarrow (P(x, y) \Leftrightarrow P(x', y'))$$

  This embodies the principle that we can substitute equals for equals.

So how do we go about reason about a particular domain? The idea is rather simple. We start with a number of function symbols and predicates for dealing with the domain at hand. This is often called the *vocabulary*. We then write down a number of formulas capturing the properties of the domain in question using the vocabulary we have defined, and use those formulas to derive new properties.

For instance, consider how we can reason about arithmetic? A reasonable vocabulary is to take as functions $+$ and $\cdot$ (addition and multiplication), and as predicate the symbol $<$. (We will be able to derive subtraction and division.) Following standard mathematical convention, I will write these functions and predicates as $t_1 + t_2, t_1 \cdot t_2, t_1 < t_2$, rather than $+(t_1, t_2), \cdot(t_1, t_2), < (t_1, t_2)$. I also write $t_1 \neq t_2$ as an abbreviation for $\neg(t_1 = t_2)$. Furthermore, I assume constants 0 and 1, representing the numbers zero and one, respectively. The following formulas capture the properties of $+, \cdot, <, 0, 1$, as typically used in arithmetic:

$$(\forall x)(\forall y)(x + y = y + x)$$
$$(\forall x)(\forall y)(x \cdot y = y \cdot x)$$
$$(\forall x)(\forall y)(\forall z)[(x + y) + z = x + (y + z)]$$
$$(\forall x)(\forall y)(\forall z)[(x \cdot y) \cdot z = x \cdot (y \cdot z)]$$
$$(\forall x)(\forall y)(\forall z)[x \cdot (y + z) = (x \cdot y) + (x \cdot z)]$$
$$(\forall x)(x + 0 = x)$$
$$(\forall x)(x \cdot 1)$$
$$(\forall x)(\exists y)(x + y = 0)$$
$$(\forall x)(\forall y)[y \neq 0 \Rightarrow (\exists z)(x = y \cdot z)]$$
$$(\forall x)(\forall y)[x < y \Rightarrow \neg(y < x)]$$
$$(\forall x)(\forall y)(\forall z)[(x < y \wedge y < z) \Rightarrow x < z]$$
$$(\forall x)(\forall y)[x \neq y \Rightarrow (x < y \vee y < z)]$$
$$(\forall x)(\forall y)(\forall z)[y < z \Rightarrow (x + y < x + z)]$$

$$(\forall x)(\forall y)(\forall z)[(0 < x \wedge y < z) \Rightarrow x \cdot y < x \cdot z]$$
$$0 \neq 1$$

These formulas are often called *axioms* of the theory. Let AR be the set of all formulas above. Since AR is finite, I will often use AR as though it were a conjunction of the axioms above. Now, if AR truly captures the properties of arithmetic, then these properties of arithmetic should be the logical consequences of AR. In other words, if φ is a formula such that $AR \Rightarrow \varphi$ is valid, then φ is a formal property of arithmetic. We can understand this by going back to the definition of validity. Recall that $AR \Rightarrow \varphi$ is valid if for all interpretations I, $AR \Rightarrow \varphi$ is true under I, that is, if for all interpretations I, AR true under I implies φ true under I. In other words, φ is true in all the models that "obey the rules" specified by the axioms.

(Of course, if the axioms are inconsistent, then there is no model that satisfies them, and our axiomatization is essentially useless. Are the axioms above consistent?)

If the above axioms are to mean anything at all, they should be true in the model we had in mind when coming up with the axioms in the first place. Consider the real numbers, under the "standard" interpretation of the symbols $+, \cdot, <, 0, 1$, that is, consider the interpretation $I_{\mathbb{R}} = (\mathbb{R}, \varphi_{\mathbb{R}}, \iota_{\mathbb{R}})$, where $\mathbb{R}$ is the set of real numbers, $\varphi_{\mathbb{R}}$ assigns to $+, \cdot$ the standard operations $+, \cdot$ on the reals, and assigns to $0, 1$ the real numbers zero and one, and $\iota_{\mathbb{R}}$ assigns to $<$ the standard ordering on the reals. We can verify that AR is true under $I_{\mathbb{R}}$, so if $AR \Rightarrow \varphi$ is valid, it is certainly the case that φ is true under $I_{\mathbb{R}}$. In other words, if we can prove something follows from the axioms AR, it will be true of the reals.

For examples, the following formulas are logical consequences of AR (check them!):

$$(\forall x)[(\forall y)(y + x = y) \Rightarrow x = 0]$$
$$(\forall x)[(\forall y)(y \cdot x = y) \Rightarrow x = 1]$$
$$(\forall x)(\forall y)(\forall z)(x + y = x + z \Rightarrow y = z)$$

In summary, the approach for reasoning about a particular domain is to figure out an appropriate vocabulary to talk about the domain, and capture the basic properties of the domain using a set of formulas, and use logical consequence to study consequences of these properties.

Some Limitations

Interestingly, the axioms AR does not completely capture reasoning about the real numbers. For instance, let $\mathbb{Q}$ be the set of rational numbers, with their standard interpretation $I_{\mathbb{Q}}$. (This is similar to the standard interpretation of the reals.) It is straightforward to check that AR is true under $I_{\mathbb{Q}}$. Thus, among other things, anything we can derive using AR is true of the rational numbers

as well. But, clearly, the real numbers and the rational numbers are different beasts. Could we come up with axioms that exactly capture the real, using the vocabulary given above? To help us out, consider why the reals and the rational are different. The real numbers satisfy the following property, called the *supremum property*: for every bounded nonempty set S of real numbers (that is, every number in S is less than a certain real number), there is a real number r such that

1. $s \leq r$ for all $s \in S$, and
2. if $s \leq r'$ for all $s \in S$, then $r \leq r'$.

The r in the property is called the supremum of S. It is a classical result of analysis that the supremum property holds of the reals. It is an equally classical results that this property does not hold of the rationals (that is, the property where we replace every mention of "real" by "rational"), by considering, say, the set $\{x \in \mathbb{Q} \mid x^2 < 2\}$. This set is bounded above, but the only supremum of this set would be $\sqrt{2}$, which is not a rational number.

Now, one would hope that we can simply add this property as a new axioms to AR. But there is a problem doing that: the property is not expressible in first-order logic! It requires quantification not over individuals, but over sets of individuals, something that one cannot write in first-order logic. In fact, the following result holds:

Proposition 1 *$\mathbb{R}$ and $\mathbb{Q}$ satisfy exactly the same first-order formulas over the vocabulary $+, \cdot, <, 0, 1$.*

Proof. This can be proved using a technique known as *Ehrenfeucht-Fraissé games*, which we will not study, but is actually not that difficult. □

So what if we want to reason more generally about mathematics? At the very least, we should be able to distinguish the reals from the rationals. Two approaches have been taken in the history of the subject.

The first approach starts from the observation that the supremum property needs quantification over sets of reals, in other words, over unary relations over reals. We can write this in second-order predicate logic, which allows quantifying over arbitrary relations over individuals in the domain. In general, other mathematical properties might require quantifying over relations over relations (differentiation), and so on. So we can look at higher-order logic, which allows quantifying over arbitrary relations. This is (believed to be) able to express most of mathematics, but is highly undecidable. The higher-order logic approach is used by many formal verification systems that aim at formalizing mathematics (such as HOL, Isabelle, etc)

The approach that is the most popular in mathematics itself is to change the vocabulary! Take as a basis a number of axioms that let you reason about sets; that is, define an operation $\in$, and give

axioms that characterize what properties $\in$ has. Now, once you have sets, you can define much of mathematics, and write down properties using sets. (The supremum property can be expressed as a first-order formulas over the vocabulary of *sets*, but not a first-order formula over the vocabulary of $+, \cdot, <, 0, 1$.) This seems like a reasonable approach, but it has its drawbacks. The biggest one is the following one: what axioms do you take as axioms of the theory of sets? The field of *set theory* is concerned with figuring out what axioms are reasonable axioms to take as those for sets, and thereby which axioms lie at the bottom of the mathematics infrastructure (when built using first-order logic and sets).

As an example of the kind of subtlety that arises, consider the following example. At the end of the 19th century, the axioms for sets (an infinite number of them) included axioms such as : for every property $\varphi(x)$, a first-order logic formula over the language of sets with one free variable, there exists a set with exactly those elements that make the property true; for $\varphi(x)$, there is an axiom

$$(\exists S)[s \in S \Leftrightarrow \varphi(s)].$$

Now, if you have such axioms, you have a big problem: consider the property $x \notin x$. By the axiom for this property, you can form the set $R = \{x \mid x \notin x\}$. Now, is it the case that $R \in R$?. You can prove that $R \in R$ if and only if $R \notin R$! In other words, the axioms are inconsistent. This turns out to be a pain to get around; most modern axiom systems for set theory (Russell's theory of types, Bernays and Godel's theory of sets, Zermelo and Frankel's theory of sets) are complex exactly in order to ensure that they are expressive while not allowing you to construct a set such as R.