Intro Logic for Computer Scientists and Philosophers

Owain West

July 31, 2017

Contents

To the Reader

This book is meant to be used for a one-semester short course serving as an introduction to formal logic for technically-minded and philosophicall-minded students alike. As such, it aims to cover material which is pertinent to both audiences. Students who are comfortable with the introductory material should feel free to skip ahead.

Part I

Introduction

This conviction of the solvability of every mathematical problem is a powerful incentive to the worker. We hear within us the perpetual call: There is the problem. Seek its solution. You can find it by pure reason, for in mathematics there is no *ignoramibus*.

David Hilbert, *Mathematical Problems*, 1900

---

1 What?

A single definition of logic would be hard to give. To some, logic is rigorous the study of truth. To others, it is the study of proof. Others approach logic as a source for mathematical foundations, and other still approach logic as a means of understanding computation.

In a way, all of these characterizations are fair. As it stands now, there are four main branches of study in logic: *model theory* ("the study of truth"), *proof theory* ("the study of proof"), *set theory* ("the study of foundations"), and *computability theory* ("the study of computation").

Mathematical Logic is distinguished from Mathematics proper in that it is concerned primarily with *metamathematics*, that is, treating statements about mathematics as objects of mathematics themselves. In particular, notions like "truth" and "provability" are formalized, allowing meta-properties to be studied in a rigorous way.

2 Why?

3 Mathematical Preliminaries

The only hard prerequisite for learning about mathematical logic is a willingness to think. The actual mathematical preliminaries necessary are simple enough that we can cover them quickly here; readers who think they need more background knowledge before beginning Chapter 1 should check the list of Suggested Readings.

3.1 Sets

There is an intuitive, naive way to talk about sets as well as a rigorous, detailed way to define them. We take the naive approach here, as it will be (mostly) sufficient for our needs.

A *set* is, intuitively, simply unordered collection of objects (which are called the *elements* of the set). For example, we might have a set $WeekDays$ whose members are *Monday, Tuesday, Wednesday, Thursday, Friday*, or a set $EvenPrimes$ whose only member is the integer 2. To say that an object a is an element of a set A, we write $a \in A$, which can be read as "a is an element of A" or "a is in A".

Sets can be specified *extensionally* (by listing the elements of the set) or *intensionally* (by specifying some property which defines the members of the set). Consider the set (call it P) of all prime numbers less than 20. Extensionally, this is

$$P = \{2, 3, 5, 7, 11, 13, 17, 19\}$$

The $\{, \}$ brackets indicate that the elements listed between are the members of the set in question.

Intensionally, P is

$$P = \{n \mid 0 < n < 20 \textbf{ and } n \text{ is prime.}\}$$

To specify a set S in this manner, we write $S = \{\text{placeholder} \mid \text{statement about placeholder}\}$, and then S is the collection of all objects such that the "statement about placeholder" holds of that object[1].

Sets are *unordered* and *disregard multiplicity*. For example

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

and

$$\{1\} = \{1, 1\}$$

are both correct assertions; in both cases, the sets actually *are equal* because neither order nor multiplicity matters for a set. Sets A, B are equal if and only (henceforth abbreviated iff) if they have the same elements (multiplicity notwithstanding).

For a set A, we write $|A|$ for the number of elements of A (called the *cardinality* of A). For finite sets, $|A|$ can be easily computed by counting how many elements A has. For example

$$A = \{a, b, c\} \implies |A| = 3$$

[1]This creation pattern for sets is called the *naive comprehension principle*. The fact that "naive" is part of its name might set alarm bells ringing in your head, and for good reason. It turns out that this principle is actually contradictory - allowing sets to be defined by arbitrary properties leads to contradiction, and actually precipitated the biggest crisis of foundation in the history of mathematics, which itself resulted in rapid development in Logic. For our current purposes, this principle is fine. We'll discuss it in more detail later.

$$B = \{a, a, b, c\} \implies |B| = 3$$
$$C = \{n \mid n \text{ is a positive integer no greater than } 20\} \implies |C| = 20$$

For sets A, B, the *intersection* of A and B, written $A \cap B$, is the collection of elements x such that $x \in A$ and $x \in B$. For example

$$A = \{1, 2, 3\}, B = \{3, 4, 5\} \implies A \cap B = \{3\}$$

The *union* of sets A, B (written $A \cup B$) is the collection of elements x such that $x \in A$ or $x \in B$ (or both; in general, I will use "p or q" to mean "either p or q or both p and q", which is the standard interpretation of "or" in logic. I will say "p exclusive or q" or "either p or q" when I wish to exclude the case that both p, q are true). For example

$$A = \{1, 2, 3\}, B = \{3, 4, 5\} \implies A \cup B = \{1, 2, 3, 4, 5\}$$

Sets can be subtracted as well. For A, B, we write $A - B$ for the set of elements which are in A but not in B. For example

$$A = \{1, 2, 3\}, B = \{3, 4, 5\} \implies A - B = \{1, 2\}$$

The *symmetric difference*[2] $A \triangle B$ is defined as $A \triangle B := (A - B) \cup (B - A)$. For example,

$$A = \{1, 2, 3\}, B = \{3, 4, 5\} \implies A \triangle B = \{1, 2, 4, 5\}$$

We can also take the product of sets. For sets A, B we have

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

which says that the elements of $A \times B$ are ordered pairs (a, b) with $a \in A$ and $b \in B$.

We say that A is a *subset* of B iff every element of A is also an element of B. In this case, we write $A \subseteq B$. Note that for every set X, we have $X \subseteq X$. If A is a subset of B but we want to specify additionally that A is not equal to B, we write $A \subset B$ and say that "A is a proper subset of B". For example, let

$$A = \{1, 2, 3\}, B = \{1, 2, 3\}, C = \{1, 2, 3, 4\}$$

Then we have

$$A \subseteq B, A \subseteq C, A \subset C$$

but we do not have

$$A \subset B$$

because $A = B$.

There are some special sets which deserve special attention. Some among them are

[2]The := symbol is used to mean "defined to be".

- The emptyset, which is the unique set containing no elements

$$\emptyset := \{\}$$

- The natural numbers

$$\mathbb{N} := \{0, 1, 2, 3, ...\}$$

- The integers

$$\mathbb{Z} := \{..., -2, -1, 0, 1, 2, ...\}$$

- The rational numbers

$$\mathbb{Q} := \{\frac{a}{b} \mid a, b \in \mathbb{Z}, b \neq 0\}$$

- The real numbers $\mathbb{R}$

3.2 Functions and Relations

Functions

A function is a map from one set (called the domain) to another set (called the codomain) which maps every element in the domain to a single element of the codomain. For example, we might define the function f as follows

$$f : \mathbb{R} \to \mathbb{R}$$

$$x \mapsto x^2$$

This says that f is a function from domain $\mathbb{R}$ to codomain $\mathbb{R}$ which sends an element $x \in \mathbb{R}$ to $x^2 \in \mathbb{R}$. You may have seen this same function written before as

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

Specifying the domain and codomain of a function is often beneficial, and so we will stick to the more verbose notation throughout this book.

Functions are *single-valued*, which means that if $f(x) = y$ and $f(x) = y'$, then $y = y'$. That is, each element of the domain maps to one element of the codomain.

The *image* of a function f with domain A and codomain B is the set of all elements of the codomain which are mapped to by some element of the domain, that is

$$im(f) = \{b \in B \mid \text{There is an } a \in A \text{ with } f(a) = b\}$$

Note that $im(f) \subseteq B$ and that it need not be the case that $im(f) = B$. For example, if we have

$$f : \mathbb{R} \to \mathbb{R}$$

$$x \mapsto x^2$$

then the codomain of f is $\mathbb{R}$ but $im(f) = \{r \in \mathbb{R} \mid r \geq 0\}$, the set of all nonnegative real numbers.

We say that a function is *injective*[3] iff no two members of the domain map to the same element of the codomain. We write $A \preceq B$ if there is an injection (an injective function) $f : A \to B$. It is simple to show that if $A \preceq B$ and A, B are finite, then $|A| \leq |B|$ (try this as an exercise). This is true for infinite sets as well, but the proof is not as trivial.

We say that a function is *surjective*[4] iff every element of the codomain is mapped to by some element of the domain (equivalently, if the image equals the codomain). As an exercise, show that if A, B are finite and there is a surjection (a surjective function) from A to B, then $|A| \geq |B|$. This is true for infinite sets as well, but the proof is not as trivial.

We say that a function is *bijective* (the function is a *bijection*[5]) iff the function is both injective and surjective. This means that bijections associate every element of the domain with a unique element of the codomain, and every element of the codomain with an element of the domain. In this way, the domain and codomain are in a one-to-one correspondence. The results we have for finite sets (and our interpretation of what a bijection is) imply that if A, B are finite and there is a bijection from A to B, then $|A| = |B|$. This is true for infinite sets as well, but the proof is not as trivial.

Technically, a set A is finite iff there is a bijection

$$f : A \to \{1, 2, ..., n\}$$

for some n. If there is no such bijection, then A is *infinite*.

If there is a bijection from A to $\mathbb{N}$, we say A is *countably infinite*. If A is neither finite nor countable, we say A is *uncountable*. If A is finite or countably infinite, we say that A is *countable*.

Relations

Relations are a generalization of functions which drop the requirement that the mapping be single-valued. A binary relation R between a domain A and a codomain B is a set of pairs (a, b) where $a \in A, b \in B$. In other words, a relation R between A and B is some subset of $A \times B$). We will normally consider relations whose domain equals their codomain. In this case, a relation R over a set S is a subset of $S^2 = S \times S$.

For example, with $S = \{1, 2\}$, $S \times S = \{(1, 1), (1, 2), (2, 1), (2, 2)\}$. We then might define some arbitrary relation R on S by $R = \{(1, 2), (1, 2)\}$.

[3]Some books call this "1-1" or "1-to-1"

[4]Some books call this "onto"

[5]Some books call this a "1-to-1 correspondence"

Common mathematical operators can be thought of as relations - for example, $<$ is a relation on $\mathbb{N}$, since it specifies the subset of $\mathbb{N} \times \mathbb{N}$ for which the proposition "a is less than b" holds.

We write Rab (or, equivalently, aRb) to say that "a holds the relation R to b". The former notation is called "prefix notation/Polish notation" and is often used in logic/cs as it is more easily machine-readable and in accordance with our normal notation for functions (which, as we saw, are relations). The latter notation is called infix notation, and has the benefit of seemeing more natural as it's more widespread (we normally write $a < b$, for example, rather than $<(a, b)$).

Notice that we identified relations with the *extension of the relation*. To define a binary relation is simply to define which pairs (a, b) for which the relation holds.

Relations need not be just binary. For example, our normal interpretation of the $+$ symbol (along with equality) specifies a relation - we might think of the ternary relation $+abc$ as expressing that $a + b = c$. The extension of this relation is then a set of ordered triples $\{(a, b, c) \mid a + b = c\}$.

As functions are relations, we can also think of a function $f : A \to B$ as being some subset of $A \times B$; the requirement that mappings be single valued ensures that for all a, there is only one b such that (a, b) is in the extension of the function.

3.3 Proof

A *proof* of a statement p is some sequence of assertions, beginning with simple assertions called *axioms* and ending with p such that each statement follows logically from the previous statement. We will have much more to say about what constitutes an acceptable axiom or inference later.

There are a few common proof techniques used when proving simple propositions. The first we consider is *proof by contradiction*.

Proof by Contradiction

Suppose we are asked to prove proposition p. Assume that p is not true and show that this leads to a contradiction. Conclude that p is true.

By way of example, suppose we are asked to prove that there are infinitely many prime numbers. We might prove this as follows

Proof. Suppose towards a contradiction that there are finitely many prime numbers. If this were the case, we could list out all the primes as $p_1, p_2, ...p_n$, where p_n is the greatest prime. Consider the number $p = p_1 \cdot p_2 \cdot ... \cdot p_n + 1$. Clearly we have $p_i \nmid p$ for all i (we write $a|b$ to

mean a divides b; similarly, $a \nmid b$ means a does not divide b). So p has no divisors except for itself and 1, and so is also prime. But $p > p_n$, contradicting that p_n was the greatest prime. So there are infinitely many prime numbers. □

Proof by Induction

Suppose that you are asked to prove that a proposition holds for all natural numbers. You might do this as follows: show that it holds for 0, and show that if it holds for n, it holds for $n+1$. This suffices to show that the proposition holds for all natural numbers (we have that it holds for 0, therefore it holds for 1, therefore it holds for 2, etc). This strategy is called *proof by induction*. Proving tthat the proposition holds for 0 (or, some arbitrary base number) is called the *base case* and proving the implication "holds for n implies holds for $n+1$" is called the *inductive step*.

By way of example, suppose we are asked to prove the well-known identity $\sum_1^n i = \frac{n(n+1)}{2}$. We might prove this as follows

Proof. BASE: let $n = 1$. Then $\sum_1^n i = 1 = (\frac{1(1+1)}{2})$ and so the base case holds.

INDUCT: Suppose that $\sum_1^n i = \frac{n(n+1)}{2}$. We want to show that this holds for $n+1$, that is, that $\sum_1^{n+1} i = \frac{(n+1)(n+2)}{2}$

We have that $\sum_1^n i = \frac{n(n+1)}{2}$. Then

$$(\sum_1^n i) + (n+1) = \frac{n(n+1)}{2} + (n+1)$$

Moving $n+1$ inside the sum on the left hand side and expanding the right hand side, we get

$$\begin{aligned}(\sum_1^{n+1} i) &= (n^2+n)/2 + (2n+2)/2 \\ &= (n^2+3n+2)/2 \\ &= ((n+1)(n+2))/2\end{aligned}$$

which is what we wanted to show, and so we are done. □

Direct Proof

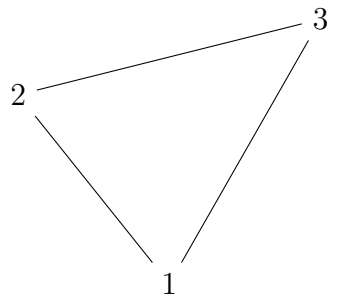
Sometimes, there is no need for a fancy-schmancy proof technique, and a direct proof is the easiest way to go about things. For example, suppose we were asked to prove that $(A \cap (B \cup C)) \subseteq ((A \cap B) \cup (A \cap C))$ for arbitrary sets A, B, C. We might prove this as follows.

Proof. Let $x \in (A \cap (B \cup C))$. Then $x \in A$ and $x \in (B \cup C)$. Then $x \in A$ and $x \in B$, or $x \in A$ and $x \in C$. If $x \in A$ and $x \in B$ then $x \in (A \cap B)$ so $x \in ((A \cap B) \cup (A \cap C))$. If $x \in A$ and $x \in C$ then $x \in (A \cap C)$ so $x \in ((A \cap B) \cup (A \cap C))$.

So each element of $(A \cap (B \cup C))$ is an element of $((A \cap B) \cup (A \cap C))$, and we are done. $\square$

3.4 Graphs

A *graph* is a tuple $G = (V, E)$ where V is a set of "vertices" or "nodes", and E is a binary relation (the edge relation) on V (ie, $E \subseteq V \times V$). For example, if $G = (V, E)$ with $v = \{1, 2, 3\}$ and $E = \{(1, 2), (2, 1), (1, 3), (3, 1), (2, 3), (3, 2)\}$, we might draw this as



where the straight lines are *undirected edges* due to the fact that whenever we have an edge from a to b (ie, a pair $(a, b) \in E$) we also have an edge from b to a (ie, a pair $(b, a) \in E$). If all edges are of this form the graph is said to be *undirected*. If this is not the case, the graph is said to be *directed* and edges are drawn with arrowheads.

Relations are often well represented as graphs; a relation R on a set S has a graph whose nodes are the elements of S and which has an edge (a, b) for each $(a, b) \in R$.

Neighbourhood: the neighbourhood of a node n of a graph G is the set of all nodes which are adjacent to n in G (ie, $\{n' \mid nEn'\}$).

Degree: the degree os a node n is the size of n's neighbourhood.

k-Regular: we say that a graph G is k-regular iff every node $n \in G$ has degree k. If we take E to be our edge relation, this can be schematized as

Note that finite 2-regular simple graphs are composed of a collection of disjoint cycles, and 1-regular simple graphs are composed of a collection of disjoint pairs of elements. The normal ordering relation on the integers (ie, the graph $(\mathbb{Z}, <)$) is an acyclic 2-regular graph (draw the graph to verify).

Part II

Propositional Logic

4 Language and Structure

The first logical system (alternatively called a *Language*) we consider is *Propositional Logic*, which we will abbreviate $\mathcal{LP}$. To specify a language, we need to describe the following

- The lexicon - ie what the valid "words" are in the language
- The syntax. This specifies how words can be connected together to form sentences.
- The semantics. This is the relation between sentences (which are *a priori* just meaningless sequences of symbols) and the meaning which a sentence prescribes.

It is the interplay between syntax and semantics - formalism and meaning - that gives logic its power.

Syntax

A special subset of a language's lexicon is its *signature*, denoted σ. For $\mathcal{LP}$, the signature can be any set of varaibles, which might look like

$$p, q, r, x, x', x_{1000}, \text{ etc}$$

which we will use to express propositions. If we have signature σ, we write $\mathcal{LP}(\sigma)$ to indicate that we are building sentences in $\mathcal{LP}$ with signature σ. $\mathcal{LP}$'s lexicon also contains the distinguished symbols

$$\wedge, \vee, \rightarrow, \leftrightarrow, \neg$$

called the *logical connectives* which are not allowed to be part of any signature, as well as parentheses

$$(,)$$

which are also not allowed to be part of the signature. Finally there are the *truth-values*

$$\top, \bot$$

which will be explained later.

An *expression* of $\mathcal{LP}(\sigma)$ is a string of symbols from the the lexicon of $\mathcal{LP}(\sigma)$. The length of this expression is the number of symbols it has.

As of yet, there is no notion of what a *meaningful* sentence of $\mathcal{LP}(\sigma)$ might look like. To achieve that, we define the *well-formed formulae (wffs)* of $\mathcal{LP}(\sigma)$. The following are wffs of $\mathcal{LP}(\sigma)$.

- $\bot$ is a wff as are all elements of σ
- If ϕ is a wff, so is $\neg\phi$
- If ϕ, ψ are wffs, so are $(\phi \wedge \psi), (\phi \vee \psi), (\phi \to \psi)$, and $(\phi \leftrightarrow \psi)$.

Nothing else is a wff of $\mathcal{LP}(\sigma)$.

For example, the following are wffs with $\sigma = \{p, q, r\}$

$$(p \to r) \to (p \to (q \to r))$$

$$\neg((p \vee \neg p) \wedge (q \wedge \neg q))$$

whereas the following are not wffs

$$p((\bot$$

$$q \to q \wedge r)$$

This defines the syntax, and all we are left with is defining the semantics.

Semantics

We can associate wffs with meaningful statements by giving each of the propositional connectives an interpretation. We will speak of truth and falsity as *truth values*, which we will write as $\top$ and $\bot$ respectively. The *truth value* of a statement will be $\top$ if the statement is true, and will be $\bot$ otherwise.

The truth-value of a wff depends on the truth-values assigned to the propositional symbols in the formula. We define how to propage truth-values up from propositions to wffs by way of *truth tables*.

Suppose we want to know the truth-value of $(p \wedge q)$ from the truth vales of p and q. The truth table below explains how:

This table says that if p is true and q is true, so is $p \wedge q$. Otherwise, $p \wedge q$ is false. That gives a natural interpretation of $p \wedge q$ as expressing "p and q".

This table says that $p \vee q$ is true iff at least one of p or q is true. That gives a natural interpretation of $p \vee q$ as expressing "p or q".

This says that $\neg p$ is true iff p is false. This gives a natural interpretation of $\neg p$ as expressing "not p".

Table 1: Conjunction (logical and)

p	q	$p \wedge q$
$\top$	$\top$	$\top$
$\top$	$\bot$	$\bot$
$\bot$	$\top$	$\bot$
$\bot$	$\bot$	$\bot$

Table 2: Disjunction (logical or)

p	q	$p \vee q$
$\top$	$\top$	$\top$
$\top$	$\bot$	$\top$
$\bot$	$\top$	$\top$
$\bot$	$\bot$	$\bot$

Table 3: Negation (not)

p	$\neg p$
$\top$	$\bot$
$\bot$	$\top$

Table 4: Conditional (if... then)

p	q	$p \to q$
$\top$	$\top$	$\top$
$\top$	$\bot$	$\bot$
$\bot$	$\top$	$\top$
$\bot$	$\bot$	$\top$

This says that $p \to q$ is false only when $p = \top$ and $q = \bot$, and should be interpreted as saying "if p, the q" or more colloquially "p implies q". In a sentence of the form $p \to q$, p is called the *antecedent* and q is called the *consequent*.

The motivation for the first two lines of the truth table should be intuitive given the suggested interpretation. For the third line, consider the sentence "if 4 is a multiple of 3 then 6 is a multiple of 3". The antecedent of that sentence is false and its consequent is true, and we intuitively think that the whole sentence is true. Similarly, we also agree that a sentence like "if 4 is a multiple of 3 then 5 is a multiple of 3" is true, even though both antecedent and consequent are false.

This says that $p \leftrightarrow q$ is true iff p, q have the same truth value. This gives a natural nterpretation of $p \leftrightarrow q$ as "p if and only if q".

Table 5: Biconditional (iff)

p	q	$p \leftrightarrow q$
$\top$	$\top$	$\top$
$\top$	$\bot$	$\bot$
$\bot$	$\top$	$\bot$
$\bot$	$\bot$	$\top$

Part III

First Order Logic

Part IV

Semantics

Part V

Syntax

Part VI

Computation

Part VII

Complexity

Part VIII

Philosophy

Part IX

Suggested Readings