MSJ Math Club

Godel's Incompleteness Theorem

4 December 2014

1 Introduction

Before we start proving this theorem, we need some basic background information.

1.1 Definitions

A *theory* is defined as all the statements that can be proven from a set of axioms (assumed truths) and inference rules (basically proof methods, so we don't get a scenario like *What the Tortoise Said to Achilles*). A theory is said to be *complete* if all true and valid statements (ones that follow the theory's syntactical structure) are provable. A theory is said to be *consistent* if one cannot prove both a statement P and $\neg P$.

We also want to introduce symbols that may be unfamiliar to those less experienced. As seen above, $\neg P$ means the logcial negation ("not") of P. $\forall$ means "for all", for example $(\forall x)(x = x)$ means that all x's are equal to themselves. $\exists$ means "there exists", for example $\neg(\exists x)(x+2 = x+1)$ means there does not exist an x such that $x + 2$ equals $x + 1$ (clearly). $A \vee B$ means "A or B". $A \supset B$ means "A implies B". $A \wedge B$ means "A and B". $f(x) := x+2$ means "$f(x)$ is defined as $x + 2$".

1.2 The Liar's and Richard's Paradoxes

"This statement is false" is a classic paradox. If it is true, then it is false; If it is false, then it is true. To resolve this paradox, however, we must notice that it is written in a language that allows for sentences to refer to other sentences as objects. In most mathematical theories, and in all consistent theories, there is no way to assess the truth of another sentence within the theory. (However there is something that comes close to that, and that is what Godel's theorem is about.)

We have a similar situation with Richard's paradox, which in effect is a more elaborate version of the liar paradox. Suppose we take all boolean functions of integers $f_0, f_1, f_2, f_3, \ldots$. Each function basically describes a property of the integer; for example $f_0(x)$ can mean "x is prime", and $f_1(x)$ can mean "x is even". At some point, there is a n such that $f_n(x)$ means "x is such that $\neg f_x(x)$ (is true)". Now we want to find $f_n(n)$, and we arrive at the same paradox. (this time with a bit of diagonalization!)

2 Foundations

This section provides an overview of the first two sections of Godel's proof, which sets the necessary background for the actual proof part.

2.1 The Axiomatic System

It is not possible to prove anything without assuming something first. And as such, everything in a theory is ultimately derived from the axioms of a theory.

2.1.1 Propositional Logic

Propositional logic deals with propositions or *sentences*. Basically, it says stuff such as $(A \wedge (A \supset B)) \supset B$ (If A and A implies B, then B), more commonly known as "tautologies". There are a few basic axioms of propositional logic (10, classically). These are $P \supset (Q \supset P)$, $(P \supset Q) \supset ((P \supset (Q \supset H)) \supset (P \supset H)$, $P \supset (Q \supset P \wedge Q)$, $P \supset P \vee Q$, $P \supset Q \vee P$, $P \wedge Q \supset P$, $P \wedge Q \supset Q$, $(P \supset Q) \supset ((H \supset Q) \supset (F \vee H \supset Q))$, $(P \supset Q) \supset ((P \supset \neg Q) \supset \neg P)$, and $\neg\neg P \supset P$. Two inference rules exist, one may substitute P, Q, and H or any sentential variable for any valid sentence (referred to as substitution), and one may prove Q from $P \supset Q$ and P (referred to as modus ponens).

2.1.2 Peano Axioms

The following are the fundamental axioms of number theory, which was the main target of Godel's incompleteness theorems. (Godel's results can actually be expanded to all axiomatic systems that are complex enough) (Full axioms is this plus first-order logic)
A. $0 \in \mathbb{N}$

B. $\forall x, x = x$
C. $\forall x, y,\ x = y \supset y = x$
D. $\forall x, y, z,\ x = y \land y = z \supset x = z$
E. $(a \in \mathbb{N}) \land (a = b) \supset b \in \mathbb{N}$
s is known as the *successor function.*
F. $\forall x \in \mathbb{N}, sx \in \mathbb{N}$
G. $\neg \exists x (sx = 0)$
H. $s(m) = s(n) \supset m = n$
I. $(0 \in K) \land (\forall n (n \in K \supset sn \in K)) \supset (K = \mathbb{N})$
The last is called the axiom of induction and K is a set.

2.2 Proofs

To understand Godel's incompleteness theorem, one must understand how proofs work, and how is it possible to check proofs. Many pieces of computer software are actually able to check proofs (search Formal Verification).

A proof is essentially a sequence of statements such that every statement follows logically from previous statements. To check a proof, one must take into account the axioms of a theory, and the inference rules. We may easily generate an algorithm for checking any proof. For each statement, we must check if the statement is an axiom, or the statement can be obtained by applying the inference rules on previous statements. Because there are a finite number of statements in a proof, this checking algorithm always terminates.

3 The Proof

Godel used a combination of the above ideas to formulate his awesome proof that not all statements in (Peano-based) number theory are provable.

3.1 Godel Numbering

The first part of the proof involves assigning a number to every symbol used in number theory. This isn't too difficult, but there are an infinite number of potential variables, and there are many types of variables (statements, natural numbers, sets/functions, sets of sets, etc). We can do this, for example, by letting each non-variable symbol be a small prime number. Since there are infinite primes, we can let a large prime (define as larger than all small primes) represent each sentential variable. Then we can let the squares of the large primes be the natural numbers, cubes of the large primes be sets, fourth powers be sets of sets and so on.

We can then use numbers to represent statements. Let $n(a)$ be the symbol number of the symbol a (see above). Let P_i denote the ith prime ($P_1 = 2$). Let the Godel number $N([S])$ of a statement $S = a_1a_2a_3a_4...$ ($a_1, a_2, a_3...$ are symbols used in the statement S) be equal to the value obtained by the product of $P_i^n(a_i)$ for all natural numbers i less than the length of the statement (number of symbols in it).

Similarly, we can do the same with proofs. Let the Godel number of a proof $Q([P])$ be equal to the product of 2 to the power of the Godel number of the first statement in the proof, 3 to the power of the Godel number of the second statement in the proof, 5 to the power of the Godel number of the third statement in the proof and so on. Obviously, these numbers are really big, but they are finite.

The nice thing about these numbers is that we can enumerate and check proofs and manipulate statements. We can first create a function $Sub(x, y, z)$ that converts the Godel number of a statement x into the Godel number of a statement with a variable y substituted for z. We can also create a function $B(x, y)$ where x is the Godel number of a proof, and y is the Godel number of a statement to mean that the proof is a valid proof of the statement. We can check it by, as mentioned above, checking if each statement in the sequence of x is either an axiom or follows by the inference rules (modus ponens and substitution), and then making sure the last statement in the sequence is actually y. Godel actually constructed the entire function in his proof, be we shall skip it for lack of time.

3.2 The Godel Sentence/Statement

Let $Bew(y) := (\exists x)(B(x, y))$. This basically states that the statement with Godel number y is *provable.* Let q be the Godel number of the variable y. Let $u := N(\text{"}\neg Bew(Sub(y, q, y))\text{"})$ (Godel number of the statement in quotes, substitute q for its numeral $sss...ss0$). The we construct the following statement:

$$G = \neg Bew(Sub(u, q, u))$$

Which basically means: "$\neg Bew(Sub(y, q, y))$ is unprovable when you substitute y for u". Notice, however, that when you make the substitution, the result is $\neg Bew(Sub(u, q, u))$, which is the equation itself! Therefore, the equation is basically saying that itself is unprovable.

G="I am unprovable. muahaha."

If G is false, then it is provable, which means that our theory is *inconsistent*; you can prove both that G is both provable and unprovable at the same time (since provable means its true so its unprovable). Therefore, if our theory is consistent, then G must be true and unprovable, and therefore, not all truths are provable.