HANDOUT FOR 1ST LECTURE

AXIOMATIC SET THEORY FALL 2024

1. Formulation of the Continuum Hypotheses

Defn. $A \preceq B$ iff $\exists f : A \to B$ that is 1-1.
Defn. $A \sim B$ iff $\exists f : A \to B$ that is 1-1 and onto.
Defn. $A \prec B$ iff $A \preceq B$ but not $B \preceq A$.
Defn. A is *finite* iff A has n members for some $n \in \mathbb{N}$.[1] Otherwise A is said to be *infinite*.[2]
Defn. A is *countable* iff $A \preceq \mathbb{N}$.

Schröder-Bernstein Theorem. If $A \preceq B$ and $B \preceq A$, then $A \sim B$.

Corollary. If $A \subset B$ and $B \preceq A$, then $A \sim B$.

N.B.: $\mathbb{N} \times \mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\{A \in \mathscr{P}(\mathbb{N}) \mid A \text{ is finite}\}$ are all countable. But $\mathscr{P}(\mathbb{N})$ is uncountable by:

Cantor's Theorem. $X \prec \mathscr{P}(X)$, for any set X.
Proof: Clearly, $X \preceq \mathscr{P}(X)$. For let $g : X \to \mathscr{P}(X)$ be s.t. for all $y \in X$, $g(y) = \{y\}$. Hence, it needs to be shown that $\mathscr{P}(X) \not\preceq X$. So, assume, contrary to what needs to be shown, that $\mathscr{P}(X) \preceq X$. By the Schröder-Bernstein theorem, it follows that $X \sim \mathscr{P}(X)$. Thus, there exists a bijection $f : X \to \mathscr{P}(X)$. However, we can show that no map $f : X \to \mathscr{P}(X)$ is onto. Let $C = \{x \in X \mid x \notin f(x)\}$. Suppose $C = f(y)$ for some $y \in X$. Then

$$y \in C \Rightarrow y \notin f(y) \Rightarrow y \notin C,$$

while

$$y \notin C \Rightarrow y \in f(y) \Rightarrow y \in C.$$

Thus, $y \in C$ iff $y \notin C$, which amounts to a contradiction. So, f is not onto. Therefore, $X \prec \mathscr{P}(X)$. ■

[1] If this does not sound especially rigorous or precise, that's because it's not. Here's a better definition if, as in ZF, we take a natural number to be the set of its predecessors. In other words, we use the following definition: $0 = \emptyset$; $n + 1 = n \cup \{n\}$. Then we say that a set A is finite just in case $A \sim n$ for some $n \in \mathbb{N}$.

[2] The first attempt at a definition of an infinite set did not use the natural numbers. According to this definition, a set is infinite iff $A \sim B$ for some proper subset B of A. A set that is infinite in this sense is now called *Dedekind infinite*.

Claim. $\mathbb{R} \sim \mathscr{P}(\mathbb{N})$.
Take this in steps:

(1) $\mathscr{P}(\mathbb{N}) \sim \mathcal{S}$, where $\mathcal{S}$ is the set of all infinite sequences of 0's and 1's, where an infinite sequence s of 0's and 1's is just a mapping $s : \mathbb{N} \to \{0, 1\}$. This is because of the following.

For each $A \subseteq \mathbb{N}$, the characteristic function χ_A of A is the mapping $\chi_A : \mathbb{N} \to \{0, 1\}$ such that for each $n \in \mathbb{N}$,

$$\chi_A(n) = \begin{cases} 1 & \text{if } n \in A \\ 0 & \text{otherwise} \end{cases}$$

Clearly for any $A \subseteq \mathbb{N}$, there is a unique $\chi_A \in \mathcal{S}$ and any $s \in \mathcal{S}$ is the characteristic function of some $A \subseteq \mathbb{N}$.

(2) By way of notation, let (a, b) be an open interval of the real line, i.e., $(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$. It's fairly immediate that $(0, 1) \sim (0, 2)$ since the function $f(x) = 2x$ is a bijection. Our strategy now is to inject $(0, 1)$ into $\mathcal{S}$, and then to inject $\mathcal{S}$ into $(0, 2)$. If we compose that injection with f^{-1}, then we have an injection of $\mathcal{S}$ into $(0, 1)$. This establishes in the first instance that $(0, 1) \preceq \mathcal{S}$ and in the second that $\mathcal{S} \preceq (0, 1)$. By the Schröder-Bernstein theorem, it follows that $(0, 1) \sim \mathcal{S}$.

The respective injections are defined as follows. For any given $r \in (0, 1)$, let $r(n)$ be the value of the $(n + 1)$-th place following the "decimal" point in the expansion of r in base 2. Thus, $r = .r(0)r(1)r(2) \cdots r(n) \cdots$ But in this form r just is a map from $\mathbb{N}$ to $\{0, 1\}$, and thus $r \in \mathcal{S}$. Thus $(0, 1) \preceq \mathcal{S}$.

The injection in the other direction is a little trickier. The reason is that some elements of $(0, 1)$ have more that one representation expressed in binary (or any) notation. For example, $0.0111111111 \cdots$ is the same real number as $0.100000 \cdots$. First we need some terminology. A subset X of $\mathbb{N}$ is said to be *cofinite* iff its relative complement in $\mathbb{N}$, i.e., $\mathbb{N} \backslash X$, is finite. If X is cofinite, then its characteristic function χ_X becomes all 1s after some $n \in \mathbb{N}$, and thus as a binary expression picks out a real number that is also represented by a different sequence of 0s and 1s. We must find some way around this duplicity in representation. The trick is to avoid the duplicity of representation by injecting $\mathscr{P}(\mathbb{N})$ into $(0, 2)$ as follows. For each $A \subseteq \mathbb{N}$, let

$$g(A) = \begin{cases} 1 + \sum_{n \in \mathbb{N}} (\chi_A(n)/2^{n+1}) & \text{if } A \text{ is cofinite} \\ \\ \sum_{n \in \mathbb{N}} (\chi_A(n)/2^{n+1}) & \text{otherwise.} \end{cases}$$

Thus the cofinite sets get mapped into the interval $(1, 2)$ while all the rest are mapped into $(0, 1)$. The composite function $f^{-1} \circ g$ then injects $\mathscr{P}(\mathbb{N})$ into $(0, 1)$. Thus, $\mathcal{S} \sim \mathscr{P}(\mathbb{N}) \preceq (0, 1)$.

(3) $(0, 1) \sim \mathbb{R}$. There are many familiar ways to show this.

Continuum Hypothesis (CH): There is no $X \subseteq \mathbb{R}$ such that $\mathbb{N} \prec X \prec \mathbb{R}$.

CH certainly seems to be a well-posed mathematical statement about the reals that is either true or false. Cantor thought it was true, but couldn't prove it. Thus, the terminology 'hypothesis'.

2. Bigger and Bigger Sets

Obviously, if we iterate the power set operator on $\mathbb{N}$, we get sets of larger and larger cardinality:

$$\mathbb{N} \prec \mathscr{P}(\mathbb{N}) \prec \mathscr{P}(\mathscr{P}(\mathbb{N})) \prec \cdots .$$

Notation. $\mathscr{P}^0(\mathbb{N}) = \mathbb{N}$; $\mathscr{P}^{n+1}(\mathbb{N}) = \mathscr{P}(\mathscr{P}^n(\mathbb{N}))$.

Thus: $\mathbb{N} \prec \mathscr{P}(\mathbb{N}) \prec \mathscr{P}^2(\mathbb{N}) \prec \mathscr{P}^3(\mathbb{N}) \prec \cdots \mathscr{P}^n(\mathbb{N}) \prec \cdots$.

Something bigger? Let $\mathscr{P}^\omega(\mathbb{N}) = \bigcup_{n\in\mathbb{N}} \mathscr{P}^n(\mathbb{N})$.

Then,

$$\mathbb{N} \prec \mathscr{P}(\mathbb{N}) \prec \cdots \prec \mathscr{P}^\omega(\mathbb{N}) \prec \mathscr{P}(\mathscr{P}^\omega(\mathbb{N})) = \mathscr{P}^{\omega+1}(\mathbb{N}) \prec \cdots \prec \mathscr{P}^{\omega+n}(\mathbb{N}) \prec \cdots \prec \mathscr{P}^{\omega+\omega}(\mathbb{N}) = \mathscr{P}^{\omega\cdot 2}(\mathbb{N})$$

Clearly we can iterate the power set operator n times on $\mathscr{P}^{\omega\cdot 2}(\mathbb{N})$ for arbitrary $n \in \mathbb{N}$, yielding $\mathscr{P}^{\omega\cdot 2+n}(\mathbb{N})$. Taking the union over all these gives us $\mathscr{P}^{\omega\cdot 3}(\mathbb{N})$. This combination of procedures can be iterated over and over, giving us $\mathscr{P}^{\omega\cdot n}(\mathbb{N})$ for arbitrary $n \in \mathbb{N}$. Taking the union over all of these in turn gives us

$$\mathscr{P}^{\omega\cdot\omega}(\mathbb{N}) =_{df} \mathscr{P}^{\omega^2}(\mathbb{N}).$$

Clearly, we can iterate everything done so far arbitrarily many finite times to get $\mathscr{P}^{\omega^n}(\mathbb{N})$ for arbitrary $n \in \mathbb{N}$. Taking the union of all these yields $\mathscr{P}^{\omega^\omega}(\mathbb{N})$. Obviously, we can then get:

$$\mathscr{P}^{\omega^\omega\cdot\omega}(\mathbb{N}) = \mathscr{P}^{\omega^{\omega+1}}(\mathbb{N}) \prec \mathscr{P}^{\omega^{\omega\cdot 2}}(\mathbb{N}) \prec \mathscr{P}^{\omega^{\omega\cdot\omega}}(\mathbb{N}) = \mathscr{P}^{\omega^{\omega^2}}(\mathbb{N}) \prec \mathscr{P}^{\omega^{\omega^\omega}}(\mathbb{N}).$$

There's no end to this iteration, only we run out of notation, and so we set

$$\mathscr{P}^{\epsilon_0}(\mathbb{N}) =_{df} \bigcup\{\mathscr{P}^\omega(\mathbb{N}), \mathscr{P}^{\omega^\omega}(\mathbb{N}), \mathscr{P}^{\omega^{\omega^\omega}}(\mathbb{N}), \ldots\}.$$

We're hardly done, though, since by iteration of these procedures we can get $\mathscr{P}^{\epsilon_n}(\mathbb{N})$ for arbitrary $n \in \mathbb{N}$, and so $\mathscr{P}^{\epsilon_\omega}(\mathbb{N})$, and, with enough work,

$$\mathscr{P}^{\epsilon_{\epsilon_0}}(\mathbb{N}) \prec \mathscr{P}^{\epsilon_{\epsilon_{\epsilon_0}}}(\mathbb{N}) \prec \mathscr{P}^{\epsilon_{\epsilon_{\epsilon_{\epsilon_0}}}}(\mathbb{N}) \prec \cdots$$

until we're forced to introduce some additional notion again. If we do that systematically we can union over the iterations from ω to ϵ_0 to the limit of subscripting ϵ, etc., etc., so we are limited only by our ingenuity of notation (and this is not mathematically limited).

However, you can verify that any given stage of construction, we have generated still only countably many sets of cardinality greater than ω. Can we do better? The answer is yes. We can construct an α such that $\mathscr{P}^\beta(\mathbb{N}) \prec \mathscr{P}^\alpha(\mathbb{N})$ for uncountably many β's. To

find such an α is to construct what's called an uncountable *ordinal*. So, we need to lay out the basics of the theory of ordinals.

3. Ordinals and Cardinals

Defn. We say that R *well-orders* A iff R linearly orders A and every non-empty subset of A has an R-least member.
Defn. Say that (A, R) is a well-ordering if R well-orders A.
Defn. A set is *transitive* just in case every element is also a subset.
Defn. A set α is an *ordinal* just in case α is transitive and the membership relation ϵ_α on α well-orders α.

You can verify that for any ordinal α, the set $\alpha \cup \{\alpha\}$ is also an ordinal, called the *successor* of α. Note that the empty set $\emptyset$ is an ordinal. Hence so is its successor $\{\emptyset\}$, and so on. We identify this sequence with the natural numbers:
$0 = \emptyset$
$1 = \{\emptyset\} = \{0\}$
$2 = \{\emptyset, \{\emptyset\}\} = \{0, 1\}$
$3 = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} = \{0, 1, 2\}$
$\vdots$
so that in general $n + 1 = n \cup \{n\} = \{0, \ldots, n\}$. Thus, the set of natural numbers $\mathbb{N}$, a.k.a, the set of finite ordinals, is then defined to be the smallest set S such that $\emptyset \in S$ and for any x, if $x \in S$, then $x \cup \{x\} \in S$. $\mathbb{N}$ is also an ordinal and is just ω from above by another name. Each of the "exponents" of the power set operator $\mathcal{P}$ constructed above is a countably infinite ordinal.

Defn. α is a *successor* ordinal iff $\alpha = \beta \cup \{\beta\}$ for some ordinal β. An ordinal other than 0 is said to be *limit* ordinal iff it is not a successor ordinal.

Examples of limit ordinals from above are: $\omega, \omega \cdot 2, \omega^2, \omega^\omega, \epsilon_0, \epsilon_1, \epsilon_\omega, \epsilon_{\epsilon_{\epsilon_0}}$.

Representation Theorem: If (A, R) is a well-ordering, then there is a unique ordinal α such that

$$(A, R) \simeq (\alpha, \epsilon_\alpha),$$

where $\simeq$ indicates order isomorphism.

Notation: Write Type(A, R) for α if $(A, R) \simeq (\alpha, \epsilon_\alpha)$.

Notation. For ordinals, write $\alpha < \beta$ for $\alpha \in \beta$.

Lemma: Any set S of ordinals is well-ordered by $<$ on S.

Notation. Let $| \alpha |$ be the least ordinal β such that $\alpha \sim \beta$. E.g., $| \epsilon_0 | = \omega$.

To get the first uncountable ordinal, consider the set

$$\mathcal{R} = \{R \in \mathscr{P}(\omega \times \omega) \mid R \text{ well-orders } \omega\}.$$

For each $R \in \mathcal{R}$, Type(ω, R) is a countable ordinal, and

$$S = \{\text{Type}(\omega, R) \mid R \in \mathcal{R}\}$$

is the set of all countable ordinals. Since $(S, <)$ is a well-ordering, there is an ordinal $\beta = \text{Type}\ (S, <)$. But β cannot be countable, else $\beta \in S$ which is contradictory. So β is an uncountable ordinal, indeed, the smallest uncountable ordinal. As such, it is called a *cardinal*.

Defn. An ordinal α is a *cardinal* iff $\mid \alpha \mid = \alpha$.

Note that ω as well as every finite ordinal is a cardinal, but that any other countable ordinal is not a cardinal.

The procedure we used above for generating from ω the smallest ordinal of greater cardinality can be applied to any ordinal α. In general, for any ordinal α, α^+ is the smallest ordinal of greater cardinality than α. We now have a technique for generating ordinals of larger and larger cardinality.

Notation: Let α be any ordinal and let γ be any limit ordinal. Then:

$$\begin{aligned} \aleph_0 &= \omega \\ \aleph_{\alpha+1} &= \aleph_\alpha^+ \\ \aleph_\gamma &= \bigcup\{\aleph_\eta \mid \eta < \gamma\} \end{aligned}$$

Note that the indices are not restricted to countable ordinals. For example, $\aleph_1$ is the first uncountable ordinal. Hence, $\aleph_{\aleph_1}$ is an ordinal, in fact the least ordinal with uncountably many smaller infinite cardinals. We can even construct a cardinal κ so large that $\kappa = \aleph_\kappa$. For example:

$$\kappa = \bigcup\{\aleph_0, \aleph_{\aleph_0}, \aleph_{\aleph_{\aleph_0}}, \ldots\}.$$

4. The Continuum Hypothesis Reformulated and Generalized

You'll note that we have developed two sequences of sets of larger and larger infinite cardinality, on the one hand the $\mathscr{P}^\beta(\mathbb{N})$ sequence and, on the other hand, the $\aleph_\alpha$ sequence. How do they compare? The one fixed point we have is that $\mathscr{P}^0(\mathbb{N}) = \omega = \aleph_0$.

Continuum Hypothesis (CH): $\mathscr{P}^1(\mathbb{N}) \sim \aleph_1$.

Generalized Continuum Hypothesis (GCH): $\mathscr{P}^\alpha(\mathbb{N}) \sim \aleph_\alpha$, for arbitrary ordinal

α.

Now, note that since $\aleph_1$ is well-ordered by the membership relation on itself, CH requires that $\mathbb{R}$ can be well ordered. For if $\aleph_1 \sim \mathbb{R}$, there is a bijection $f : \aleph_1 \to \mathbb{R}$ and we can define a well ordering on $\mathbb{R}$ by

$$r_1 \lhd r_2 \text{ iff } f^{-1}(r_1) < f^{-1}(r_2).$$

Can you think of a well-ordering of $\mathbb{R}$? Similarly, GCH requires that any $\mathscr{P}^\alpha(\mathbb{N})$ can be well-ordered. Consider the general proposition:

(AC): Any set can be well-ordered.

Can we prove AC? From what? Note that we'll have problems if we proceed naively, assuming that any collection whatsoever is a set. For example, consider the Russell "set" $T =_{df} \{x \mid x \notin x\}$. It quickly follows that $x \in T$ iff $x \notin T$. To attempt to avoid "paradoxes" such as this, we had best lay down the assumptions we are allowed to make concerning sets.

5. The Axioms of Zermelo-Fraenkel Set Theory (ZF)

The language $\mathcal{L}_{\text{ZF}}$ of ZF has, besides logical symbols, only a single, two-place predicate ϵ for set membership. In principle, all axioms (and theorems) can be expressed using just ϵ. Sometimes, though, abbreviations might be used. For example, $x \subseteq y$ is short for:

$$\forall z(z \in x \to z \in y).$$

However, I will state the axioms in English to the extent possible. You should be able to translate the English into formal notation.

Extensionality: Sets are equal if they have exactly the same members.

Comprehension Scheme: For any set A and any predicate $\varphi(x)$ that can be expressed in $\mathcal{L}_{\text{ZF}}$, there is a set $B = \{x \in A \mid \varphi(x)\}$. [This is called a scheme, since it gives us an infinite number of axioms, one for each predicate $\varphi(x)$.]

Pairing: Given any sets x and y, there is a set that contains both x and y as members.

Union: The union over any set is a subset of some set, i.e., for any set A there is a set B such that $x \in B$ if there exists a $C \in A$ such that $x \in C$.

Replacement Scheme: Suppose we have a set A and a predicate $\varphi(x, y)$ in the language $\mathcal{L}_{\text{ZF}}$ such that for any $x \in A$ there exists a unique y such that $\varphi(x, y)$. Then there is a set B containing each y such that $\varphi(x, y)$ for some $x \in A$. [Again, this is a scheme since there is an axiom for each predicate $\varphi(x, y)$ meeting the conditions.]

Infinity: There exists a set S such that $\emptyset \in S$ and, for any x, if $x \in S$, then $x \cup \{x\} \in S$.

Power Set: For any set X, there exists a set containing every subset of X.

The totality of axioms so far is known as ZF^- to indicate that the following is missing.

Foundation: Every set has an ϵ-least element. i.e., every non-empty set X has an element disjoint with X. [This rules out "pathological" cases where membership is cyclic or regresses infinitely, i.e., cases such as $x_0 \in x_1 \in \cdots \in x_n \in x_0$ or $\cdots \in x_3 \in x_2 \in x_1 \in x_0$.]

The full set of axioms is known as ZF. Very often, Foundation is irrelevant, motivating the consideration of ZF^-.

Notation: ZFC = ZF + AC.