Countability

Reading: Sipser, “The Diagonalization Method,”
from just before Definition 4.12 up to Corollary 4.18,
pages 202–207 (174–178 2^{nd} ed.).

Examples of Regular Languages

- $\{w \in \{a, b\}^* : |w|$ even & every 3rd symbol is an $a\}$
- $\{w \in \{a, b\}^* :$ There are not 7 a's or 7 b's in a row$\}$
- $\{w \in \{a, b\}^* : w$ has both an even number of a's and an even number of b's$\}$
- $\{w : w$ is written using the ASCII character set and every substring delimited by spaces, punctuation marks, or the beginning or end of the string is in the American Heritage Dictionary$\}$

Questions about regular languages

Give $X =$ a regular expression, DFA, or NFA, how could you tell if:

- $x \in L(X)$, where x is some string?
- $L(X) = \emptyset$?
- $x \in L(X)$ but $x \notin L(Y)$?
- $L(X) = L(Y)$, where Y is another RE/FA?
- $L(X)$ is infinite?
- There are infinitely many strings that belong to both $L(X)$ and $L(Y)$?

Goal: Existence of Non-Regular Languages

Intuition:

- Every regular language can be described by a finite string (namely a regular expression).
- To specify an arbitrary language requires an infinite amount of information.
  - For example, an infinite sequence of bits would suffice.
  - Σ^* has a lexicographic ordering, and the i'th bit of an infinite sequence specifying a language would say whether or not the i'th string is in the language.

$\Rightarrow$ Some languages must not be regular.

How to formalize?

Countability

- A set S is **finite** if there is a bijection $\{1, \ldots, n\} \leftrightarrow S$ for some $n \geq 0$.
- **Countably infinite** if there is a bijection $f : \mathcal{N} \leftrightarrow S$

  This means that S can be "enumerated," i.e. listed as $\{s_0, s_1, s_2, \ldots\}$ where $s_i = f(i)$ for $i = 0, 1, 2, 3, \ldots$

  So $\mathcal{N}$ itself is countably infinite

  So is $\mathcal{Z}$ (integers) since $\mathcal{Z} = \{0, -1, 1, -2, 2, \ldots\}$

  Q: What is f?
- **Countable** if S is finite or countably infinite
- **Uncountable** if it is not countable

Facts about Infinite Sets

- **Proposition:** The union of 2 countably infinite sets is countably infinite.

  If $A = \{a_0, a_1, \ldots\}$, $B = \{b_0, b_1, \ldots\}$

  The $A \cup B = C = \{c_0, c_1, \ldots\}$

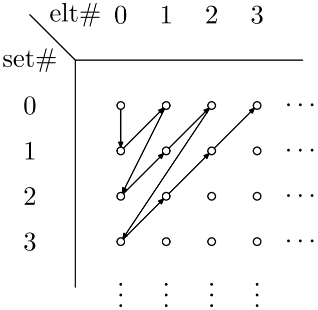
  where $c_i = \begin{cases} a_{i/2} & \text{if } i \text{ is even} \\ b_{(i-1)/2} & \text{if } i \text{ is odd} \end{cases}$

  Q: If we are being fussy, there is a small problem with this argument. What is it?

- **Proposition:** If there is a function $f : \mathcal{N} \to S$ that is onto S then S is countable.

Countable Unions of Countable Sets

- **Proposition:** The union of countably many countably infinite sets is countably infinite



$(0,0)$,
$(0,1), (1,0)$,
$(0,2), (1,1), (2,0)$,
$\ldots$

Each element is “reached” eventually in this ordering

Q: What is the bijection $\mathcal{N} \leftrightarrow \mathcal{N} \times \mathcal{N}$?

Are there uncountable sets? (Infinite but not countably infinite)

Theorem: $\mathcal{P}(\mathcal{N})$ is uncountable
(The set of all sets of natural numbers)

Proof by contradiction: (i.e. assume that $\mathcal{P}(\mathcal{N})$ is countable and show that this results in a contradiction)

- Suppose that $\mathcal{P}(\mathcal{N})$ were countable.
- There there is an enumeration of all subsets of $\mathcal{N}$ say $\mathcal{P}(\mathcal{N}) = \{S_0, S_1, \ldots\}$

Diagonalization

S_i \ $j =$	0	1	2	3	4	$\cdots$
S_0	Y	N	N	Y	N	$\cdots$
S_1	N	N	N	N	N	$\cdots$
S_2	Y	Y	N	Y	Y	$\cdots$
S_3	N	N	N	Y	N	$\cdots$
$\vdots$				D		

"Y" in row i, column j means $j \in S_i$

- Let $D = \{i \in \mathcal{N} : i \in S_i\}$ be the diagonal
- $D = YNNY \ldots = \{0, 3, \ldots\}$
- Let $\overline{D} = \mathcal{N} - D$ be its complement
- $\overline{D} = NYYN \ldots = \{1, 2, \ldots\}$
- **Claim:** $\overline{D}$ is omitted from the enumeration, contradicting the assumption that every set of natural numbers is one of the S_is.

  Pf: $\overline{D}$ is different from each row; they differ at the diagonal.

Cardinality of Languages

- An alphabet Σ is finite by definition
- **Proposition:** Σ^* is countably infinite
- So every language is either finite or countably infinite
- $\mathcal{P}(\Sigma^*)$ is uncountable, being the set of subsets of a countable infinite set.

  i.e. There are uncountably many languages over any alphabet

  Q: Even if $|\Sigma| = 1$?

Existence of Non-regular Languages

Theorem: For every alphabet Σ, there exists a non-regular language over Σ.

Proof:

- There are only countably many regular expressions over Σ.
  $\Rightarrow$ There are only countably many regular languages over Σ.
- There are uncountably many languages over Σ.
- Thus at least one language must be non-regular.

$\Rightarrow$ In fact, "almost all" languages must be non-regular.

Q: Could we do this proof using DFAs instead?

Q: Can we get our hands on an *explicit* non-regular language?