- **Reading:** Sipser, §4.2 “The Diagonalization Method,” pages 174–178 (from just before Definition 4.12 until just before Corollary 4.18) and §1.4.
Countable Unions of Countable Sets

**Proposition:** The union of countably many countable sets is countable.

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

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

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

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

<table>
<thead>
<tr>
<th>( j = 0 )</th>
<th>( 1 )</th>
<th>( 2 )</th>
<th>( 3 )</th>
<th>( 4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S_i )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( S_0 )</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>( S_1 )</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>( S_2 )</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>( S_3 )</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( D )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

"Y" in row \( i \), column \( j \) means \( j \in S_i \)

- Let \( D = \{i \in \mathcal{N} : i \in S_i\} \) be the diagonal.
- \( D = YNNYY \ldots = \{0, 3, \ldots\} \)
- Let \( \overline{D} = \mathcal{N} - D \) be its complement.
- \( \overline{D} = NYYYN \ldots = \{1, 2, \ldots\} \)
- **Claim:** \( \overline{D} \) is omitted from the enumeration, contradicting the assumption that every set of natural numbers is one of the \( S_i \)s.
- **Pf:** \( \overline{D} \) is different from each row because they differ at the diagonal.
Cardinality of Languages

• An alphabet $\Sigma$ is finite by definition

• **Proposition:** $\Sigma^*$ is countably infinite.
  **Proof:**

• So every language is either finite or countably infinite

• $P(\Sigma^*)$ 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 $\Sigma$, there exists a non-regular language over $\Sigma$.

**Proof:**

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

In fact, “almost all” languages must be non-regular.
Existence of Non-regular Languages

**Theorem:** For every alphabet $\Sigma$, there exists a non-regular language over $\Sigma$.

**Q:** Could we do this proof using DFAs instead?

**Q:** Can we get our hands on an explicit non-regular language?
Goal: Explicit Non-Regular Languages

It appears that a language such as

\[ L = \{ x \in \Sigma^* : |x| = 2^n \text{ for some } n \geq 0 \} \]

\[ = \{ a, b, aa, ab, ba, bb, aaaa, \ldots , bbbb, aaaaaaaa, \ldots \} \]

can’t be regular because the “gaps” in the set of possible lengths become arbitrarily large, and no DFA could keep track of them.

But this isn’t a proof!

**Approach:**

1. Prove some general property \( P \) of all regular languages.
2. Show that \( L \) does not have \( P \).
Pumping Lemma (Basic Version)

If \( L \) is regular, then there is a number \( p \) (the pumping length) such that every string \( s \in L \) of length at least \( p \) can be divided into \( s = xyz \), where \( y \neq \varepsilon \) and for every \( n \geq 0 \), \( xy^nz \in L \).

\[
\begin{array}{c|c|c|c}
 n = 1 & x & y & z \\
 n = 0 & x & z \\
 n = 2 & x & y & y & z \\
 \ldots & \\
\end{array}
\]

• Why is the part about \( p \) needed?

• Why is the part about \( y \neq \varepsilon \) needed?
Pumping Lemma Example

• Consider

\[ L = \{ x : x \text{ has an even # of } a's \text{ and an odd # of } b's \} \]

• Since \( L \) is regular, pumping lemma holds.
  
  (i.e, every sufficiently long string \( s \) in \( L \) is “pumpable”)

• For example, if \( s = aab \), we can write \( x = \varepsilon, y = aa, \) and \( z = b \).
Pumping the even $a$’s, odd $b$’s language

• Claim: $L$ satisfies pumping lemma with pumping length $p = 4$.

• Proof:

• **Q**: Can the Pumping Lemma be used to prove that $L$ is regular?