Nonregular language I

We are interested in the limit of finite automata

- Some languages cannot be recognized

\[ \{0^n1^n \mid n \geq 0\} \]

- We might remember #0 first
  - But # of possible n’s is \( \infty \)
- Thus we cannot recognize it by finite automata
- However, this is not a formal proof
- It may be difficult to quickly tell if a language is regular or not
Consider two languages

\[ C = \{ w \mid \#0 = \#1 \} \]

\[ D = \{ w \mid \#01 = \#10 \} \]

It seems that both are not regular

Indeed, \( C \) is not regular but \( D \) is

This is an exercise in the book, so we don’t give details

To formally prove a language is not regular, we will introduce the pumping lemma
Pumping lemma I

- Strategy: by contradiction
- We prove
  
  \[ \text{regular} \Rightarrow \text{some properties} \]

- If “some properties” cannot hold, then the language is not regular
If $A$ regular $\Rightarrow \exists p$ (pumping length) such that 
$\forall s \in A, |s| \geq p, \exists x, y, z$ such that $s = xyz$ and

1. $\forall i \geq 0, xy^iz \in A$
2. $|y| > 0$
3. $|xy| \leq p$

Note that for $y^i$, we have $y^0 = \epsilon$
Proof of pumping lemma I

- Because $A$ is regular, $\exists$ a DFA to recognize $A$
  Let $p = \# \text{ states of this DFA}$
- If no string $s$ such that $|s| \geq p$, then the theorem statement is satisfied
- Now consider any $s$ with $|s| \geq p$

$$s = s_1 \cdots s_n$$
Proof of pumping lemma II

To process this string, assume the state sequence is

\[ q_1 \cdots q_{n+1} \]

Because \(|s| \geq p\), we have

\[ n + 1 > p \]

In \(1 \ldots p + 1\) two states must be the same (pigeonhole principle)

Fig 1.72
Proof of pumping lemma III

$\epsilon$, $q_j$, $q_l$, $q_a$,
Proof of pumping lemma IV

Assume

\[ q_j \text{ and } q_l \text{ with } j \leq p + 1, \ l \leq p + 1 \]

are two same states. Then let

\[ x = s_1 \cdots s_{j-1}, \ y = s_j, \cdots s_{l-1}, \ z = s_l \cdots s_n \]

Because \( j \neq l \),

\[ |y| > 0 \]

From \( l \leq p + 1 \), we have

\[ |xy| \leq p \]

Thus all conditions are satisfied