A special class of equivalence relation for strings in $\Sigma^*$

$x \sim y$, $x, y \in \Sigma^*$ are "equivalent"
under a right invariant property if

$\forall z \in \Sigma^*$ \quad x \sim y \implies x.z \sim y.z$

**Intuition**: Given a DFA, $M$, all string $w \in \Sigma^*$ such that

$\hat{\delta}(q_0, w) = q'$

$x \sim_M y$ if $\hat{\delta}(q_0, x) = \hat{\delta}(q_0, y)$

This equivalence relation $\sim_M$ is right invariant since

$\hat{\delta}(q_0, x.z) = \hat{\delta}(q_0, y.z) \forall z$

Why is $\sim_M$ transitive?

If $x \sim_M y$ and $y \sim_M z$ then $x \sim_M z$
Any equivalence relation on a set partitions its elements into (disjoint) equivalence classes.

Can we achieve a reduction in the number of states (equivalence classes) for a given regular language?

1. What is the min no. of equivalence classes - related to minimum state DFA for a language $L$?

2. Is this machine "unique"?

We want to "tighten" our definition of the equivalence relation in the following way:

$$x \sim_L y \iff \forall z \in \Sigma^* \left[ x.z \sim_L y.z \right]$$
The relation that we had defined on the basis of the machine does not force
\[ x \sim y \quad \text{even if} \quad xz \sim yz \quad z \in \Sigma^* \]

\[ R_L: \text{relation for a language } L \]

Myhill-Nerode Relation
\[ x \sim_L y \quad \text{iff} \quad \exists z \in \Sigma^* \text{ either} \]
\[ x \cdot z \text{ and } y \cdot z \text{ are both in } L \text{ or both are not in } L. \]

Claim: \( R_L \) is right invarient equivalence relation

why is it equivalence?
reflexive, symmetric, transitive
\[ x \sim_L y \quad \text{iff} \quad x \sim_L x \quad \forall z \in \Sigma^* \]

RI. if \( x \sim_L y \quad \text{then} \quad z \in \Sigma^* \quad x \cdot z \sim_L y \cdot z \]
We know that for all \( z' \in \Sigma^* \), \( x \cdot z' \) and \( y \cdot z' \) are both in \( L \) or not in \( L \) for some \( x \mathrel{R_L} y \).

We want to show that \( \forall u \in \Sigma^* x \cdot u \mathrel{R_L} y \cdot u \) or in other words, \( \forall z \in \Sigma^* x \cdot u \cdot z \) and \( y \cdot u \cdot z \) are both in \( L \) or not in \( L \).

Choose \( z' = u \cdot z \).

---

The equivalence classes \( [z] \) for the reln \( R_L \) for a regular language correspond to the states of min state DFA.

---

Myhill Nerode Theorem

A language \( L \) is regular iff the no. of equivalence classes \( [R_L] \) is finite.
For proof we will go through an indirect construction using $R_m$.

The following statements are equivalent:

1. $L$ is a regular language.
2. $L$ is the union of some number of equivalence classes of a right invariant equivalence relation of finite index ($\#\text{equivalence classes}$ is finite).
3. $R_L$ has finite index.

1. $\Rightarrow$ 2 $\Rightarrow$ 3 $\Rightarrow$ 1

Moreover, $x R_m y \equiv x R_L y$.

$\#\text{equivalence classes of } R_L \leq \#\text{equivalence classes of } R_m$.

$\forall x R_m y \Rightarrow \exists z \in z$.

$x \cdot z R_m y \cdot z$ (property?)

So either both $x z$ are in $L$ or not in $L$.

$\Rightarrow x R_L y$. 

Given $R_L$ with finite index, we will construct a DFA for $L$.

Let $[x]$ denote the equivalence class of any string $x \in \Sigma^*$.

$$M = (Q, \Sigma, \delta, \epsilon, F)$$

$$\delta([x], a) = [x.a] \quad a \in \Sigma$$

Is it consistent? i.e. if $y \in [x]$ is

$$[y.a] = [x.a] \quad \text{(by R.1)}$$

$F$ is a state, if $[x] \in F$.

We must justify that $M$ accepts exactly $L$.

$$\delta([\varepsilon], x) \in F \quad \text{iff} \quad x \in L$$

"$[x]$ by our previous definition $\delta$ so $[x] \in F$ if $x \in L$ q.e.d"