Consider \( \Sigma = \{0, 1\} \) and any language \( L \subseteq \Sigma^* \).

For all \( x \) in \( \Sigma^* \), define: \( \text{row}_L(x) \) to be the infinite vector

\[
\text{row}_L(x) = (L(x), L(x0), L(x1), L(x00), L(x01), \ldots, L(x^n), \ldots)
\]

Then \( x \equiv_L y \) if and only if \( \text{row}_L(x) = \text{row}_L(y) \).

**Full Myhill-Nerode Theorem:** \( L \) is regular \( \iff \) there are only finitely many distinct \( \text{row}_L(x) \) vectors \( \iff \) all PD sets for \( L \) are finite.

*Proof (\( \Rightarrow \)):* Suppose there only finitely many distinct rows.

Define \( M = (Q, \Sigma, \delta, s, F) \) as follows:

- \( Q = \{ \text{distinct rows} \} \)
- \( s = \text{row}_L(\varepsilon) \)
- \( F = \{ \text{row}_L(x) : x \in L \} \)

Then \( L(M) = L \), and since \( Q \) is finite, \( M \) is a DFA, so \( L \) is regular.

**Corollary:** The number \( K \) of distinct rows, which equals the max size of a PD set for \( L \), equals the min \# of states needed by a DFA for \( L \), and the \( M \) in the proof is the smallest DFA achieving this.

A PD set for \( L \) has all distinct rows.
More Non (?) regular Language Examples: Consider
\[ L_1 = \{ x \in \Sigma^* | \#_0(x) = \#_1(x) \} \]
\[ L_2 = \{ x \in \Sigma^* | \#_0(x) = \#_1(x) \} \].

"Proof" that \( L_1 \) is non-regular: Take \( S = 0^* \). Clearly
differentiate.
Let any \( x, y \in S \), \( x \neq y \), be given. Then we can write
\( x = 0^m \), \( y = 0^n \), where \( \text{wlog.} \, m \leq n \). Take \( z = 1^n \).
Then \( xz = 0^m 1^n \in L_1 \), since \( n \neq m \) but \( yz = 0^m 1^n \in L_1 \).
So \( L(xz) \neq L(yz) \), and since \( S \) is infinite, this proves the "there is"

"Wrong why?"

\[ x\bar{z} = 0001111 \in L_1 \]
\[ w = 0001111 \]

Re-define \( L_1 = \{ w : w \text{ can be broken as } w = u \cdot v \, \text{st.} \, \#_0(u) = \#_1(v) \} \).
Thus the "proof" is wrong. In fact, \( L_1 \) is regular!

\[ L_1 = \Sigma^* : \]
\[ \text{Given any } w \in \Sigma^n, \ \text{let } \ f(i) \]
\[ \text{For each } i, 0 \leq i \leq n, \text{define } m_i = \#_0(x_i \cdot x_{i+1}) - \#_1(x_i \cdot x_{i+1}) \]
\[ f(0) = -\#_1(x) \text{ which is } \leq 0 \]
\[ f(i) \text{ steps up by } +1 \text{ (or } 0 \text{ or } -1) \text{ from } f(i-1) \]
\[ f(n) = \#_0(x) \text{ which is } \geq 0 \].
Thus, if \( i \) st. \( f(i) = 0 \), Take \( u = x_i \cdot x_{i+1} \), \( v \in \Sigma^c \).
So \( \Sigma^c \) and since \( w \in \Sigma^* \) is arbitrary, \( L_1 = \Sigma^* \) which is regular: \( (0 \cup 1)^* \).

This proof does not work for \( L_2 \) because the "step up to 0" could be when \( x_i = 1 \). Then we don't get the needed 0 in \( X01 \).
Re-define $L_2 = \{ w \in \{0,1\}^* : w \text{ can be broken as } w = u0v \text{ st. } \#0(u) = \#1(v) \}$

Proof that $L_2$ is non-regular: Take $S = \emptyset$. Clearly: Infinite.

Let any $x, y \in S$, $x+y$ be given. Then we can write $x = 0^m$, $y = 0^n$ where \( \text{wlog } m < n \). Take $z = 1^n$.

Then $xz = 0^m 1^{n-1} \notin L_2$ since uses one of the 0s from the $0^m$ part, but then $\#0(u) \leq m-1$ while $\#1(v) = n-1$.

Whereas $yz = 0^m 1^{n-1} \notin L_2$, because it parses as $0^{n-1}01^{n-1}$ so $L(xz) \nsubseteq L(yz)$.

Since $x, y \in S$ are arbitrary and $S$ is infinite, $L_2 \notin REG$. 

---

### Ch 2 Preview

What Kind of formalism can describe languages like $L_2$?

```
G: S → 0 | 1S | 0S0 | 0S1 | 1S0 | 1S1 | ε | 0 | 1
G': S → 0S0 | 1S1 | ε | 0 | 1
```

Claim: $L(G') = L_2$. Abbreviate as $L(G) = L_2$.}

Note: $W = 10$ OK $W = 01$? $S → 0$

$ε \in L_2$, $W = 00$ OK? YES, since it acts as a base parses, as base 0 to $S → 1S$

$0 \in L_2$, OK? Must balance by sending the $S → SO$ case. $W = 0·0·0$ $W = 0·0·1$. $S → OS0$

$W = 0·1·0$ Not OK. $S → 1S$