

So far you've seen the Pumping Lemma for regular languages, which we used to show some languages are irregular, and an irregular language that nonetheless satisfies the Pumping Lemma. That's not great, so we need something better.

Recall the minimal-state DFA. We defined two states p, q to be indistinguishable if $\forall w \in \Sigma^*, \delta^*(p, w) \in F \Leftrightarrow \delta^*(q, w) \in F$. Well, we can do something similar for strings. We must have gotten to p, q via strings u, v (i.e. $\delta^*(q_0, u) = p$ & $\delta^*(q_0, v) = q$). We then say u, v are indistinguishable since uw is accepted iff vw is accepted. We can phrase this another way.

Def) Let L be a language. A string $z \in \Sigma^*$ is a distinguishing extension of strings $x, y \in \Sigma^*$ if exactly one of xz and yz is in L .

Def) Let L be a language. Strings $x, y \in \Sigma^*$ are indistinguishable, written $x =_L y$, if there is no distinguishing extension of x and y wrt L .

Notice that $=_L$ has a nice property.

Lemma) Let L be a language. Then $=_L$ is an equivalence relation.

PF) Clearly $x =_L x$ for all $x \in \Sigma^*$ since $xz = xz$ for all $z \in \Sigma^*$ and hence are always in L or not together. Hence $=_L$ is reflexive.

Similarly $x =_L y \Rightarrow y =_L x$. It makes no difference to say $xz \in L \Leftrightarrow yz \in L$ or $yz \in L \Leftrightarrow xz \in L$. Hence $=_L$ is symmetric.

Lastly, suppose $a =_L b$ and $b =_L c$. Then $\forall z \in \Sigma^*$, we have $az \in L \Leftrightarrow bz \in L$ and $bz \in L \Leftrightarrow cz \in L$,

but then

$az \in L \Rightarrow bz \in L \Rightarrow cz \in L$
 $cz \in L \Rightarrow bz \in L \Rightarrow az \in L$,

hence $az \in L \Leftrightarrow cz \in L$ and $=_L$ is transitive.

So $=_L$ divides Σ^* into a countable number of equivalence classes. With DFAs, these correspond a set of indistinguishable states. When this set has size 1, for every class, we have the minimal DFA. Moreover, this implies that DFAs have a finite number of these classes. Even better, this is only true for regular languages, so we can use this as a foolproof regularity test.

Thm (Myhill-Nerode) A language L is regular iff $=_L$ has a finite number of equivalence classes. Moreover, the number of equivalence classes corresponds to the number of states in the minimal DFA for L .

Pf) Let L be a regular language.

Let $D = (Q, \Sigma, \delta, q_0, L)$ be a DFA with $L(D) = L$.

For each $q \in Q$, define

$$W_q = \{w \in \Sigma^* \mid \delta^*(q_0, w) = q\}.$$

W_q is the set of inputs which leaves D in state q .

Notice that $\{W_q\}_{q \in Q}$ partitions Σ^* (no string is in more than one cell).

Moreover, $\forall q \in Q \forall w, v \in W_q, w =_L v$. Since for all $z \in \Sigma^*$,

$$\delta^*(q_0, wz) = \delta^*(q, z) = \delta^*(q_0, vz).$$

Thus $=_L$ has at most $|Q| < \infty$ equivalence classes.

Now suppose $=_L$ has a finite number n of equivalence classes.

Let W_i denote the i^{th} equivalence class of $=_L$.

Define the DFA $D = (Q, \Sigma, \delta, q_0, F)$, where

$$- Q = \{q_i \mid i \in \mathbb{Z}_n\} \quad (q_i \text{ corresponds to } W_i)$$

$$- \delta(q_i, a) = q_j \quad \text{where } \text{if } w \in W_i, wa \in W_j \text{ (note that which } w \text{ doesn't matter because any extension of } w \in W_i \text{ has the same behavior by definition of } =_L)$$

$$- q_0 = q_i \quad \text{where } \epsilon \in W_i \text{ (}\epsilon = \text{no input yet} = \text{start state)}$$

$$- F = \{q_i \mid W_i \subseteq L\}$$

hand wave for
the sake of time →

To conclude the proof, we constructed D so that D accepts all strings in L and none not in L , so $L(D) = L$, hence L is regular.

Lastly, note two facts. We showed that \equiv_L has at most $|Q|$ equivalence classes for any DFA for L . Moreover, we showed that $|Q|$ is at most required to be the number n of equivalence classes. Thus for the minimal DFA,

$$n \leq |Q| \leq n,$$

or $|Q| = n.$

□

Ex) Prove $L = \{w \in \mathbb{Z}_2^* \mid \#(w, 0) \text{ is even}\}.$

Define the sets

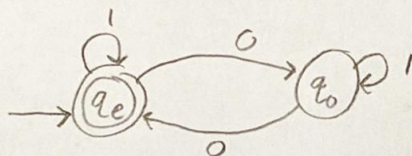
$$W_1 = \{w \in \mathbb{Z}_2^* \mid \#(w, 0) \text{ is even}\},$$

$$W_2 = \{w \in \mathbb{Z}_2^* \mid \#(w, 0) \text{ is odd}\}.$$

It's easy to verify that these are the equivalence classes of \equiv_L . Since there's only 2, L is regular.

□

In fact, the minimal DFA has 2 states.



Ex) Prove $L = \{0^n 1^n \mid n \geq 0\}$ is irregular.

For $n \geq 0$, consider 0^n . For each n , $\exists!$ extension $z(1^n)$ for which $0^n z \in L$ and $0^n z \notin L$ for all other z .

As such $x \equiv_L y \Rightarrow x = y$, so for each $n \geq 0$, there is an equivalence class. \equiv_L has an infinite number of equivalence classes then, so L is irregular.

□