

The Pumping Lemma Argument in the Court of Logic

Burt Rosenberg

March 2, 2017

Introduction

A language R has been accused of not being regular. Its guilt will be decided in the Court of Logic. The case is argued in the court between an Advocate, defending R against the charge of non-regularity, and a Prosecutor, trying to prove that charge. We will assume that both the Advocate and the Prosecutor are diligent and skillful. There is also a Judge in the Court of Logic. However, as the Judge must work completely by a fixed set of rules, the role is merely symbolic.

The court has two principles: that the innocent is never unjustly condemned, and that the guilty is not always be wrongly acquitted. That is: no regular language is ever classified as non-regular, and non-regular languages are often distinguished from regular languages. The principles are formalized by the definitions of *perfection* and *incomplete efficacy*.

Definition 0.1 (Perfection of Justice). There is an Advocate sufficiently skillful so that no matter what argument is made by the Prosecutor, no innocent language is condemned, i.e. no regular language is judged to be non-regular. Logic then demands that any language found guilty of being non-regular will certainly be non-regular.

Definition 0.2 (Incomplete Efficacy of Justice). No matter what the argument of the Advocate, the skillful Prosecutor can successfully prosecute some guilty languages, i.e. not all non-regular languages are judged to be regular. However it is possible that some guilty languages are acquitted.

Pumping Lemma Arguments

Perfect justice is possible by basing the court's judgement on a property possessed by all regular languages. Hopefully such a property is shared with few non-regular languages. One such property is captured by the *pumping lemma*: all sufficiently long strings in an infinite regular language can be *pumped*, i.e. each contains a substring that can be reinserted into the string multiple times at the place of its occurrence while remaining a string in the language.

The pumping lemma forms the basis for an argument made between the advocate and prosecutor to achieve a perfect and efficient adjudication of regular languages.

Theorem 0.1 (The Pumping Lemma Argument). The Pumping Lemma Argument is a protocol between an Advocate A and a Prosecutor P (in the presence of a fair Judge J) to condemn or acquit a language R accused of being non-regular. The protocol gives both perfection of justice and incomplete efficacy of justice. It consists of the following steps.

1. A makes its opening remarks, by declaring an integer p .
2. P either concedes the case or presents his argument: a string w .
3. J ascertains that w is of length at least p and is in the language R , else J acquits.
4. A responds to P 's argument by claiming $w = xyz$, with y a non-empty string and $|xy| < p$.
5. P then either concedes the case or presents an integer i , claiming xy^iz is not in the language.
6. J then checks that xy^iz is not in the language, and if so, judges R guilty as charged, else J acquits.

Proof. We first show *perfection*.

We can assume R is regular. If R is finite, A announces a p larger than the length of any string in R ; then P cannot announce a satisfactory w hence J acquits. Else A builds a machine M recognizing R and announces any p larger than the number of states in this machine. No matter what string w is provided by P in response to p , A runs M on w and identifies a repeated state. A divides the string $w = xyz$, where y is a substring of w which begins and ends on the repeated state; and A announces this xyz . Since $xy^iz \in R$ for any i , either P concedes, or J cannot verify the i provided, and J acquits.

We next show *incomplete efficacy*.

Not all languages are acquitted. The language,

$$R_{01} = \{0^i 1^i \mid i \geq 0\}$$

is not regular and it is not acquitted. P has only to present $0^p 1^p \in R_{01}$ to A and A has no response that can gain acquittal.

Some non-regular languages are acquitted. The language

$$R_{abc} = \{a^i b^j c^k \mid i, j, k \geq 0 \text{ and if } i = 1 \text{ then } j = k\}$$

can be shown to be non-regular by the Myhill-Nerode theorem. Any two strings in the infinite set,

$$\{ab^{2i}c^i \mid i = 1, 2, 3, \dots\}$$

can be distinguished: If $w_1 = ab^{2i_1}c^{i_1}$ and $w_2 = ab^{2i_2}c^{i_2}$ for $i_1 \neq i_2$, then $w_1c^{i_1} \in R_{abc}$ but $w_2c^{i_1} \notin R_{abc}$.

However A can gain acquittal against any P as follows: Let A announce $p = 1$. P responds with its best argument, some non-empty string w . A responds with $w = xyz$ where x the empty string, y is the length one string that is the first letter of w , and z being the rest of w . J will acquit, as P can produce no $i \geq 0$ that will cause $y^i z \notin abc$. \square

In the above proof, examples of incomplete efficacy were extreme examples. In the case of R_{01} , there is really only one sort of argument to make, one only needs to make it long enough. Other examples require the Prosecutor to pick its argument with care else the Advocate, by clever counter-argument, will gain acquittal of a non-regular language.

Example 0.1. The language,

$$R = \{w \in \{0,1\}^* \mid \text{the number of 0's and 1's in } w \text{ are equal}\}$$

is not regular. The Prosecutor should not make the argument $w = (01)^k$, because the Advocate will respond with $y = 01$. In which case the pumped strings are of the form $(01)^{k+i}$, and still in the language. Better the Prosecutor make the argument $w = 0^p 1^p$, to which the Advocate has no effective defense.

Example 0.2. The language

$$R_{ww} = \{ww \mid w \in \{0,1\}^*, \}$$

is not regular. The Prosecutor should not make the argument $w = (01)^k$, because the Advocate will respond with $y = 0101$. All pumped versions of w remain in R_{ww} . Instead the Prosecutor should argue $w = 0^p 10^p 1$, to which the Advocate has no effective defense.

Sufficiency of arguments

When arguing with the pumping lemma, both the Advocate and the Prosecutor must be skillful and diligent. This means that the arguments of the winning side are presented differently than the arguments of the losing side. The winning side need only present the one argument that wins the case; however the argument of the losing side must be presented in sufficient generality to convince the court that not just the particular argument given did not prevail, but that no argument given could have prevailed.

In the case of acquitting a regular language, this means that the Advocate is the “there exists” player, showing that it has at least one argument, while the Prosecutor is the “for all” player that shows that no argument is a sufficient counter. When the tables are turned, and it is a question of concluding guilt, then the Advocate is the “for all” player, showing that it could not have argued better, and the Prosecutor is the “there exists” player that only needs to present one winning argument.

Theorem 0.2 (Sufficient argument). To establish that a language R is regular, it must be shown that the Advocate can always win the case against an arbitrarily skillful Prosecutor. That is (omitting details for brevity),

$$\exists p \forall w \exists y \forall i xy^i z \in R.$$

Arguing the case of Regular languages, the Advocate is the “there exists” player and the Prosecutor is the “for all” player. To establish that a language R is non-regular, it must be shown that the Prosecutor can always win the case against an arbitrarily skillful Advocate. That is (omitting details for brevity),

$$\forall p \exists w \forall y \exists i xy^i z \notin R$$

Arguing the case of non-regular languages, the Prosecutor is the “there exists” player and the Advocate is the “for all” player.

Example 0.3. The language represented by the Regular Express $a(bc)^+$ is regular. The Advocate can argue in the particular, $p = 5$ and $y = bc$. The Prosecutor’s arguments must be made sufficiently general as $w = a(bc)^k$ for $k \geq 2$. Then,

$$\forall i \geq 0, xy^i z = a(bc)^{(k-1)+i} \in a(bc)^+,$$

winning the case for the Advocate.

Example 0.4. The language represented by $a^i b^j a^i$ with $i, j \geq 1$ is non-regular. The Advocate must argue in general as $p \geq 0$ and the Prosecutor can make the particular response $a^t b a^t$ where $t = \max(1, p)$. The Advocate will respond as $y = a^s$ for some $s \geq 1$, then Prosecutor can make the particular argument $i = 2$,

$$xy^2 z = a^{t+s} b a^t \notin a^i b^j a^i,$$

winning the case for the Prosecutor.