

## Lecture 3

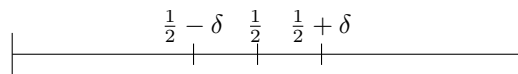
*Date:* Oct. 12, 2009. *Instructor:* C. Say. *Scribe:* Ergin Özkucur.

### Review

In the previous lectures, we have seen that one-way deterministic finite automata (1DFA) are equivalent to two-way deterministic finite automata (2DFA) and they both recognize regular languages. However, 2DFA have better state complexity than 1DFA.

We also have seen two cases for the one-way probabilistic finite automata (1PFA);

- In **unbounded error** case, for string  $x$ ,  $|\text{pr}(x \text{ is accepted}) - \text{pr}(x \text{ is rejected})|$  has no positive lower bound.
- In **bounded error** (isolated cut-point) case, for a string  $x$ , the probability that the machine makes an error is  $\epsilon < \frac{1}{2} - \delta$ . (Recall that a PFA with an arbitrary cut-point has an equivalent PFA with cutpoint  $\frac{1}{2}$ )



The bound on error of a PFA can be reduced to an arbitrary bound with a repetition procedure called probability amplification. The error can be reduced to  $2^{-t}$  with  $O(t)$  repetition.

One problem here may seem that, by the definition of a 1PFA, the input string can be read once, however we want to simulate a 1PFA several times on an input string. Such a 1PFA can be constructed with the idea similar to the Theorem 1.25 in [1], where two 1DFAs are merged and executed in parallel with cartesian product of the state sets. In 1PFA case, the transition matrix is the repeated Tensor product ( $\otimes$ ) of the original 1PFA, where

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots \\ a_{21}B & a_{22}B & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

State complexity is high for probability amplification with combinatorial PFA.

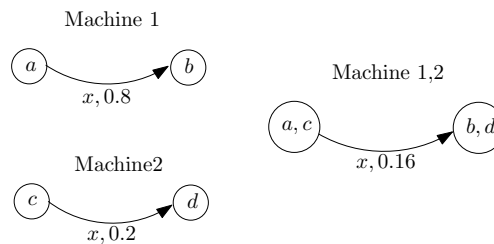
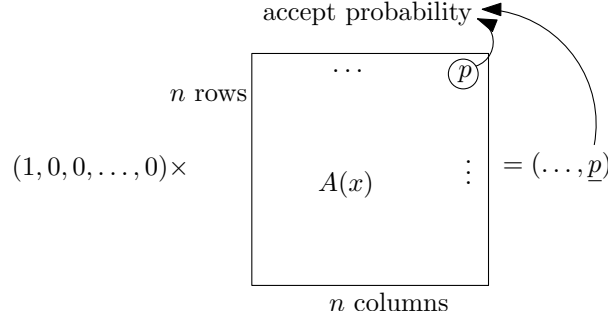


Figure 1: Merging PFAs

**Theorem 1.** *The class of languages recognized by IPFAs with bounded error equals the class of regular languages [2].*

*Proof.* The idea of the proof is that we can construct a DFA for a PFA. Let the state set of our PFA be  $S = \{s_0, \dots, s_{n-1}\}$ . Accept state is  $s_{n-1}$ . Transition probabilities for a string can be represented as  $n \times n$  matrices for each symbol.

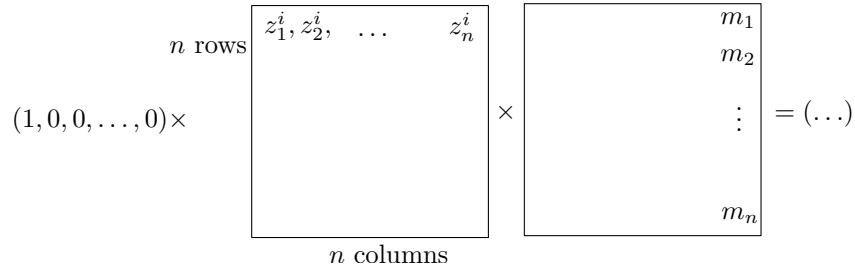
For any string  $x$ , the matrix  $A(x)$ : We will look at index and show that index is



finite. Let  $L$  be language recognised by given PFA. Recall  $\equiv_L$  for string indistinguishability.

Let  $x_1, x_2, \dots, x_k$  be strings which are pairwise inequivalent by  $\equiv_L$ . So for every  $i, j \leq k, i \neq j$  there exists a string  $y$  such that  $x_i y \in L$  or vice versa. Let the first row of  $A(x_i)$ , for all  $i$ , be  $(z_1^i, z_2^i, \dots, z_n^i)$ . Let  $p(x)$  denote the probability of acceptance of  $x$ . For string  $y$ , let the last column of  $A(y)$  be  $(m_1, m_2, \dots, m_n)$

$$A(x_i y) = A(x_i) A(y)$$



$$\begin{aligned}
 p(x_i y) &= z_1^i m_1 + z_2^i m_2 + \dots + z_n^i m_n \quad (x_i y \text{ accept}) \\
 p(x_j y) &= z_1^j m_1 + z_2^j m_2 + \dots + z_n^j m_n \quad (x_j y \text{ reject}) \\
 \lambda &< z_1^i m_1 + z_2^i m_2 + \dots + z_n^i m_n \\
 \lambda &\geq z_1^j m_1 + z_2^j m_2 + \dots + z_n^j m_n
 \end{aligned}$$

since  $\delta \leq |p(x) - \lambda|$  for all  $x$

$$2\delta \leq (z_1^i - z_1^j)m_1 + \dots + (z_n^i - z_n^j)m_n$$

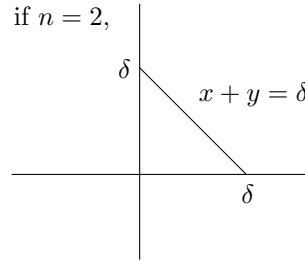
since the values of the  $m_i \leq 1$

$$2\delta \leq |z_1^i - z_1^j| + \dots + |z_n^i - z_n^j| \quad \text{for any } i \neq j \quad (1)$$

$$\sigma_i = \{(z_1, \dots, z_n) | z_j^i \leq z_j, 1 \leq j \leq n, \sum_j (z_j - z_j^i) = \delta\}$$

$$\sigma = \{(z_1, \dots, z_n) | 0 \leq z_i, 1 \leq j \leq n, \sum_j z_j = \delta\}$$

$\sigma$  is an  $(n - 1)$  dimensional simplex which is a subset of the hyperplane  $x_1 + x_2 +$



$\dots + x_n = \delta$   $\sigma$  has  $(n - 1)$  dimensional volume

$$V_{n-1}(\sigma) = c\delta^{n-1}$$

where  $c$  is some constant not dependent on  $\delta$ .

$(n - 1)$ -dimensional volume:

- $n = 2 \implies$  length of a line
- $n = 3 \implies$  area of a plane

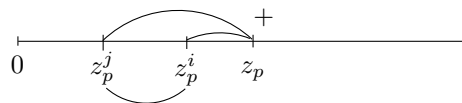
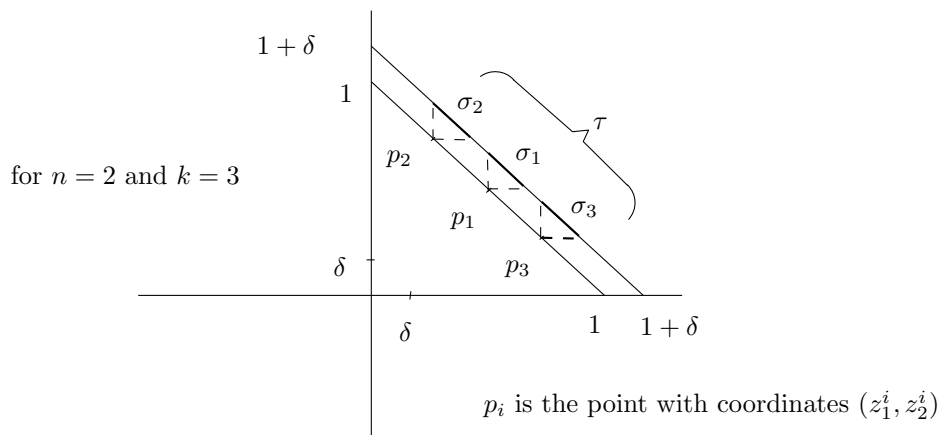
From  $\sum_j z_j^i = 1, (z_1, \dots, z_n) \in \sigma_i \implies \sum_j z_j = 1 + \delta$  and  $0 \leq z_j$ . Thus  $\sigma_i \subseteq \tau$  where  $\tau = \{(z_1, \dots, z_n) | \sum_j z_j = 1 + \delta, 0 \leq z_j, 1 \leq j \leq n\}$

A point  $(z_1, \dots, z_n) \in \sigma_i$  is an interior point of  $\sigma_i$  if and only if  $0 < z_p - z_p^i$  for all  $p, 1 \leq p \leq n$ .

Claim:  $\sigma_1$  and  $\sigma_j, i \neq j$  have no interior point in common.

Proof: otherwise,  $(z_1, \dots, z_n)$  is interior to both  $\sigma_i$  and  $\sigma_j$  then  $0 < z_p - z_p^i$ , and  $0 < z_p - z_p^j$  hence

$$|z_p^i - z_p^j| < |z_p - z_p^i| + |z_p - z_p^j|, 1 \leq p \leq n$$



$$\sum_p |z_p^i - z_p^j| < \underbrace{\sum_p |z_p - z_p^i|}_{\delta} + \underbrace{\sum_p |z_p - z_p^j|}_{\delta}, \quad 1 \leq p \leq n$$

From Equation 1,  $\sum_p |z_p^i - z_p^j| < 2\delta$  so they do not intersect.

Sum of volumes:

$$kc\delta^{n-1} \leq c(1 + \delta)^{n-1}$$

$$k \leq \frac{c(1 + \delta)^{n-1}}{c\delta^{n-1}}$$

so  $k$  is finite. □

### State Complexity

There exists PFA  $M$  with just two states and a sequence  $\lambda_n$ ;  $1 \leq n \leq \infty$ , of isolated cutpoints such that for each  $n$ , the DFA  $B_n$  with the least number of states satisfying  $L(M, \delta_n) = L(B_n)$  has at least  $n$  states. The PFA:

$$S = \{s_0, s_1\}, F = \{s_1\}, \Sigma = \{0, 2\}$$

$$A(0) = \begin{pmatrix} 1 & 0 \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}, A(2) = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ 0 & 1 \end{pmatrix}$$

If  $\mathbf{x} = x_1, x_2, \dots, x_n \in \Sigma^*$ ,

$$p(\mathbf{x}) = \frac{x_n}{3} + \frac{x_{n-1}}{3^2} + \dots + \frac{x_1}{3^{n-1}}$$

$$p(\mathbf{x}) = 0.x_n x_{n-1}, \dots x_1 \text{ (in base 3)}$$

$$\lambda_n = 0.\underbrace{22 \dots 211}_{n+1}$$

The cutpoint is isolated. For  $\lambda_6 = 0.2222211$ , a string with highest  $p(\mathbf{x})$  below  $\lambda_6$  is  $0.22222022 \dots$ . And a string with lowest  $p(\mathbf{x})$  above  $\lambda_6$  is  $0.222222$ .

Members of the language have at least  $n$  2's at the end. With Myhill-Nerode theorem, the DFA has  $n + 1$  states.

## References

- [1] Michael Sipser. *Introduction to the Theory of Computation*. Course Technology, 2006.
- [2] M. Rabin. Probabilistic automata. *Information and Control* 6, 3:220–245, 1963.