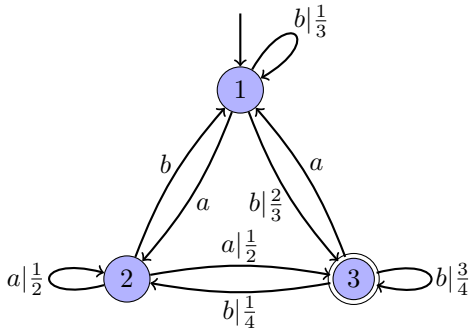
Probabilistic automata and Markov chains

Amaury Pouly
IRIF/CNRS – Université Paris Diderot



Lectures notes of the Master Parisien de Recherche en Informatique

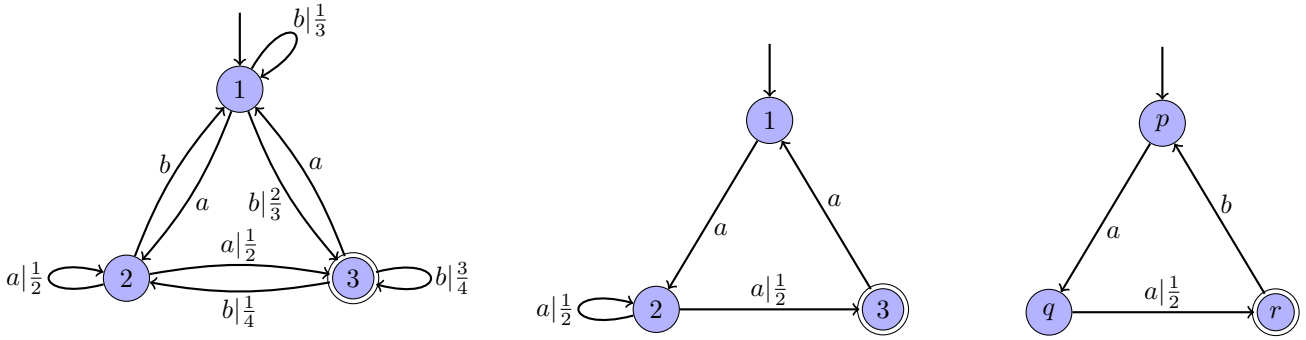
Course 2.16 – Finite automata based computation models

Academic year 2018 – 2019

Contents

These lectures notes are are intended to be mostly self-contained. As much as possible, I try to use similar notations to *Jacques Sakarovitch*'s part of the course on weighted automata and transducers[Sak19].

*Version 0.4—89a4292, compiled on February 13, 2019



(a) Automaton $\mathcal{A}$ (b) Automaton $\mathcal{B}$ (c) Automaton $\mathcal{C}$

Figure 1: Examples of probabilistic automata

1 Probabilistic automata

Probabilistic automata are a generalization of finite automata introduced by [Rab63]. They are also a particular case of weighted automata where the weights are rational (or real) and the transition matrices are stochastic (probabilities sums to 1). The interpretation of this model is that the automaton associates to each word a probability of acceptance.

A matrix $M \in \mathbb{R}^{P \times Q}$ is said to be *stochastic* if all entries are between 0 and 1, and the sum of all entries on each row is equal to 1, *i.e.* $\sum_{q \in Q} M_{pq} = 1$ for all $p \in P$. A *probabilistic automaton* is a tuple $\mathcal{A} = \langle A, Q, S, \mu, T \rangle$ where

- A is a finite *alphabet*,
- Q is a finite set of *states*,
- $S \in [0,1]^{1 \times Q}$ is stochastic (row) vector of *initial probabilities*,
- $T \in \{0,1\}^{Q \times 1}$ is a $0-1$ (column) vector of *accepting states*,
- $\mu(a) \in [0,1]^{Q \times Q}$ is a stochastic matrix of *transition probabilities*, for every $a \in A$.

Unless otherwise stated, we always requires the probabilities to be rational numbers. We naturally extend μ to define a *morphism* from the set of words to the set of $Q \times Q$ matrices, using the usual matrix product. To every word $w \in A^*$, we can now associate the *probability of acceptance* $\mathbb{P}_{\mathcal{A}}(w) = S\mu(w)T$.

It will occasionally be useful to more fined-grained probabilities. Given two states $q, q' \in Q$ and a word $w \in A^*$, we define the probability of going from state q to state q' by reading w to be $\mathbb{P}_{\mathcal{A}}\left(q \xrightarrow{w} q'\right) = \mu(w)_{q,q'}$.

Example 1 (Automaton of Figure 1a)**.** We use the notation $a|p$ on an edge of q to q' to signify that the transition labelled by a has probability p; formally $\mu(a)_{qq'} = p$. If the probability is 1 then we sometimes write just a. In this example, there in a unique initial state, labelled by an incoming arrow, which therefore has probability 1. We identified the accepting states by an extra circle.

This probabilistic automaton is represented by the tuple $\mathcal{A} = \langle A, Q, S, \mu, T \rangle$ where $A = \{a, b\}$, $Q = \{1, 2, 3\}$ and

$$S = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}, \qquad \mu(a) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \end{bmatrix}, \qquad \mu(b) = \begin{bmatrix} \frac{1}{3} & 0 & \frac{2}{3} \\ 1 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} \end{bmatrix} \quad \text{and} \quad T = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Consider the word bb, it can be accepted through two paths:

- $1 \to 1 \to 3$ with probability $\frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}$
- $1 \to 3 \to 3$ with probability $\frac{2}{3} \cdot \frac{3}{4} = \frac{1}{2}$

Note that the path $1 \to 3 \to 2$ has probability $\frac{1}{3} \cdot \frac{1}{4}$ but ends at 2 which is *not a accepting state*. Thus the probability of acceptance of bb is $\frac{2}{9} + \frac{1}{2} = \frac{13}{18}$.

Exercise 2. Recall the definition of the morphism μ.

Exercise 3. In Example 1, check that I, $\mu(a)$ and $\mu(b)$ are stochastic. Check that the acceptance probability of bb matches the formal definition, *i.e.* $S\mu(bb)T = \frac{13}{18}$. What is the acceptance probability of $aabb$?

Exercise 4. Show that the product of two stochastic matrices is stochastic.

Exercise 5. Given $\langle A, Q, S, \mu, T\rangle$, two states $q, q' \in Q$ and a word w, what is the interpretation of $\mu(w)_{q,q'}$? Prove it. Therefore what is the meaning of $S\mu(w)$? *Hint: you have almost done this in the part on weighted automata already.*

It is often convenient to create probabilistic automata where the transition matrix is not stochastic because the probabilities sum to *less* than 1. This is the case in Figure 1c: the probability to leave state 2 is only $\frac{1}{2}$. This can handled in two ways: either by allowing *substochastic* matrices, where the sum in each row less or equal to 1. Or, by adding a *sink state* which is not accepting and collects all the missing probabilities. The two approaches are equivalent: any path that reaches the sink state will never leave it and thus has probability of acceptance 0.

Exercise 6. Illustrate the substochastic and sink state approaches on $\mathcal{C}$ from Figure 1c. Show that indeed every word has the same probability in each approach.

In the context of weighted automata, it is natural to consider the weighted language of words recognized by an automaton, where the weight is the probability of acceptance. In the context of probabilistic automata, a new interesting notion of language emerges. Let $\mathcal{A}$ be a probabilistic automaton and $0 \leqslant \lambda \leqslant 1$, define the language recognized by $\mathcal{A}$ as

$$\mathcal{L}_{\mathcal{A}}(\lambda) = \{w \in A^* : \mathbb{P}_{\mathcal{A}}(w) > \lambda\}.$$

In other words, $\mathcal{L}_{\mathcal{A}}(\lambda)$ is the set of words accepted by $\mathcal{A}$ with probability at least λ. Any such $\mathcal{L}_{\mathcal{A}}(\lambda)$ is called a *stochastic language* and λ is called a *cut-point*. Note however that $\mathcal{L}_{\mathcal{A}}(\lambda)$ is not a weighted language, there is one language for each λ.

Example 7 (Automaton of Figure 1a)**.** We have seen in Example 1 that $\mathbb{P}_{\mathcal{A}}(bb) = \frac{13}{18}$ thus $bb \in \mathcal{L}_{\mathcal{A}}(\lambda)$ for every $\lambda < \frac{13}{18}$, but $bb \notin \mathcal{L}_{\mathcal{A}}(\lambda)$ for every $\lambda \geqslant \frac{13}{18}$.

Exercise 8 (Automaton of Figure 1a)**.** Find a word that is not in $\mathcal{L}_{\mathcal{A}}(\frac{1}{2})$ and one that is in $\mathcal{L}_{\mathcal{A}}(\lambda)$ for all $\lambda < \frac{2}{3}$. Can you find a word in $\mathcal{L}_{\mathcal{A}}(\frac{2}{3})$? Find an infinite regular language that is included in $\mathcal{L}_{\mathcal{A}}(\lambda)$ for all $\lambda < \frac{2}{3}$.

Exercise 9 (Automata of Figure 1)**.** What is the relationship between $\mathcal{B}$ of Figure 1b and $\mathcal{A}$ of Figure 1a, in particular can you relate $\mathcal{L}_{\mathcal{A}}(\lambda)$ and $\mathcal{L}_{\mathcal{B}}(\lambda)$?

1.1 Relation to regular languages

It is natural to try to understand how stochastic languages compares to other classes of language, and in particular if they are decidable language. A first simple step toward this goal is to compare them to regular languages.

Exercise 10. Prove that every regular language is stochastic. *Hint: take a finite automata and consider its transition matrix: $\mu(a)_{q,q'} = 1$ if there is an edge from q to q' labelled by a, 0 otherwise.*

Exercise 11. Let A be a finite alphanet. prove that the collection of regular languages over A^* is countable.

1.1.1 Non-regular stochastic languages

A first observation is that there exist some stochastic languages that are not regular, this was proven in [Rab63] using a counting argument.

Theorem 12. *Stochastic languages strictly contains regular languages.*

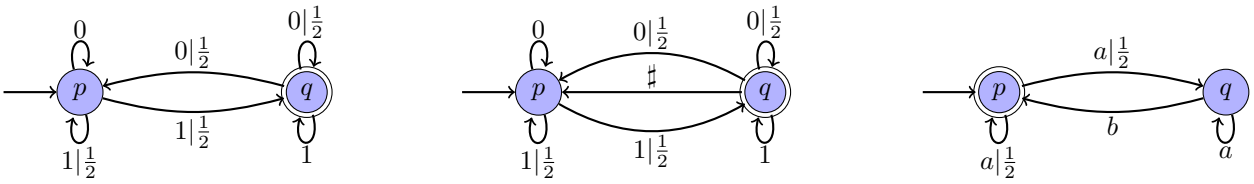
Proof. Every regular language is stochastic, see Exercise 10. Conversely, we will construct a nonregular stochastic language. Consider $\mathcal{A} = \langle A, Q, S, \mu, T\rangle$ where $A = \{0, 1\}$, $Q = \{p, q\}$ and

$$S = \begin{bmatrix}1 & 0\end{bmatrix}, \qquad \mu(0) = \begin{bmatrix}1 & 0\\ \frac{1}{2} & \frac{1}{2}\end{bmatrix}, \qquad \mu(1) = \begin{bmatrix}\frac{1}{2} & \frac{1}{2}\\ 0 & 1\end{bmatrix}, \qquad T = \begin{bmatrix}0\\ 1\end{bmatrix}.$$

This automaton is illustrated in Figure 2a. Given a word $w \in A^*$, define $[w] = \sum_{i=1}^{|w|} w_i 2^{i-|w|-1}$. We now claim that $S\mu(w) = \begin{bmatrix}1 - [w] & [w]\end{bmatrix}$. Indeed check that $[\varepsilon] = 0$, $[w0] = \frac{[w]}{2}$ and $[w1] = \frac{1+[w]}{2}$. Then we check by induction that

$$S = \begin{bmatrix}1 - [\varepsilon] & [\varepsilon]\end{bmatrix}, \qquad \begin{bmatrix}1 - [w] & [w]\end{bmatrix}\mu(0) = \begin{bmatrix}1 - \frac{1}{2}[w] & \frac{1}{2}[w]\end{bmatrix}, \qquad \begin{bmatrix}1 - [w] & [w]\end{bmatrix}\mu(1) = \begin{bmatrix}\frac{1-[w]}{2} & \frac{1+[w]}{2}\end{bmatrix}.$$

It follows that the probability of acceptance of w is $\mathbb{P}_{\mathcal{A}}(w) = [w]$. But now note that $[w]$ is dense in $[0, 1]$ for $w \in A^*$. It follows that if $\lambda < \mu$ then $\mathcal{L}_{\mathcal{A}}(\lambda) \supsetneq \mathcal{L}_{\mathcal{A}}(\mu)$. Indeed, by density we can find w such that $\lambda < \mathbb{P}_{\mathcal{A}}(w) \leqslant \mu$ since $\lambda < \mu$. Therefore the collection $\{\mathcal{L}_{\mathcal{A}}(\lambda) : \lambda \in [0, 1]\}$ is uncountable. But the collection of regular languages is countable, thus there exists a λ such that $\mathcal{L}_{\mathcal{A}}(\lambda)$ is not regular. □



(a) Automaton $\mathcal{A}$ (b) Automaton $\mathcal{B}$ (c) Automaton $\mathcal{C}$

Figure 2: Examples of stochastic automata whose language is not regular.

We will need the following classical theorem on the characterization of regular languages.

Theorem 13 (*Myhill-Nerode*)**.** *Let L be a language, we say that two words u and v are* right equivalent *for L, and write $u \equiv_L v$, if for every $w \in A^*$, we have $uw \in L$ if and only if $vw \in L$. Prove that $\equiv_L$ is an equivalence relation. Show that a language L is regular if and only if the number of equivalence classes of A^* with respect to $\equiv_L$ is finite. Furthermore, the number of equivalence classes corresponds to the number of states of the smallest deterministic finite automaton that recognizes L.*

Exercise 14. Prove Theorem 13. For the last statement, you can show that the number of equivalence classes is a *bound* of the number of states (and not necessarily that it is optimal). *Hint:* the equivalence classes correspond to states of an automaton that recognizes L.

1.1.2 Universally non-regular probabilistic automata

The original construction by Rabin showed that there exists an automaton $\mathcal{A}$ such that $\mathcal{L}_{\mathcal{A}}(\lambda)$ is not regular for at least one λ. Surprisingly, a small modification of this automaton given by [FS15] allows us to strengthen this statement.

Theorem 15. *There exists a* universally non-regular *probabilistic automaton,* i.e. *an automaton $\mathcal{B}$ such that $\mathcal{L}_{\mathcal{B}}(\lambda)$ is non-regular for all $\lambda \in (0,1)$.*

Proof. Consider automaton $\mathcal{B}$ illustrated in Figure 2b, it is defined over the alphabet $A' = A \cup \{\sharp\} = \{0, 1, \sharp\}$. It is the same as automaton $\mathcal{A}$ from the proof of Theorem 12 with an extra transition from q to p labelled by $\sharp$. It is clear that any word accepted with positive probability must be of the word $u\sharp v$ where $u, v \in A^*$. Furthermore, it is almost immediate that $\mathbb{P}_{\mathcal{B}}(u\sharp v) = \mathbb{P}_{\mathcal{A}}(u)\mathbb{P}_{\mathcal{A}}(v) = [u][v]$. Recall that the set $[A^*] = \{[w] : w \in A^*\}$ is dense in $[0,1]$.

Now fix $\lambda \in (0,1)$ and take $u, v \in A^*$ such that $\lambda < [u] < [v]$. Then by density of $[A^*]$ we can find $w \in A^*$ such that $\frac{\lambda}{[u]} < [w] < \frac{\lambda}{[v]}$. But then $\mathbb{P}_{\mathcal{B}}(u\sharp w) = [u][w] > \lambda$ whereas $\mathbb{P}_{\mathcal{B}}(v\sharp w) = [v][w] < \lambda$. This shows that $u \not\equiv_{\mathcal{L}_{\mathcal{B}}(\lambda)} v$. Again by density of $[A^*]$, we can find infinitely many such pairs u, v and thus $\mathcal{L}_{\mathcal{B}}(\lambda)$ cannot be regular by Theorem 13. □

Exercise 16. Let $\mathcal{C} = \langle A, Q, S, \mu, T\rangle$ be the automaton illustrated in Figure 2c. Give A, Q, S, μ and T. Show for every word $x(n_1, \ldots, n_k) := a^{n_1}ba^{n_2}\cdots a^{n_k}b$ we have $\mathbb{P}_{\mathcal{C}}(x(n_1, \ldots, n_k; m)) = 2^{-m}\prod_{i=1}^{k}(1-2^{-n_i})$. Show that if $u = x(n_1, \ldots, n_k)$ and $w = x(n_{k+1}, \ldots, n_\ell)$ then $\mathbb{P}_{\mathcal{C}}(uw) = \mathbb{P}_{\mathcal{C}}(u)\mathbb{P}_{\mathcal{C}}(w)$. Show that $\{\mathbb{P}_{\mathcal{C}}(x(n_1, \ldots, n_k)) : n_1, \ldots, n_k \in \mathbb{N}, k \in \mathbb{N}\}$ is dense in $[0,1]$. Conclude that $\mathcal{C}$ is universally non-regular. *Hint:* use the same proof idea as Theorem 15.

1.1.3 Isolated cut-points

An interesting observation in the examples above is that stochastic languages that are non-regular tend to verify that $\mathcal{L}_{\mathcal{A}}(\lambda) \neq \mathcal{L}_{\mathcal{A}}(\lambda + \varepsilon)$ for small ε. For example, this was essential in the proof of existence of such languages. On the other hand, simple examples that only recognize regular language tend to satisfy the opposite property that the language is unchanged by small perturbation in the threshold. The latter are called isolated cut-points.

Formally, a cut-point λ is called *isolated* with respect to some probabilistic automaton $\mathcal{A}$ if there exists $\delta > 0$ such that

$$|\mathbb{P}_{\mathcal{A}}(w) - \lambda| \geqslant \delta,\ \forall w \in A^*.$$

We will call δ the *isolation threshold* (for λ), although there is no standard name for it.

Theorem 17. *If λ is isolated with respect to $\mathcal{A}$ then $\mathcal{L}_{\mathcal{A}}(\lambda)$ is regular. Furthermore, if $\mathcal{A}$ has n states and r final states, then $\mathcal{L}_{\mathcal{A}}(\lambda)$ can be recognized by a finite deterministic automaton with at most $(1 + \frac{r}{\delta})^{n-1}$ states where δ is the isolation threshold.*

Proof. First assume there is a unique final state. Write $\mathcal{A} = \langle A, Q, S, \mu, T\rangle$ where $Q = \{s_1, \dots, s_n\}$ and s_n is the only final state. Let $L = \mathcal{L}_{\mathcal{A}}(\lambda)$ and assume that λ is isolated with threshold $\delta > 0$. Let $x_1, \dots, x_k \in A^*$ be pairwise $\equiv_L$ inequality words (see Theorem 13). Then by definition, for every $i \neq j$, there exist $y \in A^*$ such that $x_i y \in L$ but $x_j y \notin L$ (or the other way around). Since λ is isolated we must have that

$$\mathbb{P}_{\mathcal{A}}(x_i y) - \mathbb{P}_{\mathcal{A}}(x_j y) \geqslant 2\delta.$$

Let $(\xi_1^i, \dots, \xi_n^i)$ be the first row of $\mu(x_i)$. Let $(\eta_1, \dots, \eta_n)$ be the last column of $\mu(y)$, for this particular y. Check that $\mathbb{P}_{\mathcal{A}}(x_i y) = S\mu(x_i y)T = S\mu(x_i)\mu(y)T = \xi_1^i\eta_1 + \cdots \xi_n^i\eta_n$ and thus

$$\mathbb{P}_{\mathcal{A}}(x_i y) - \mathbb{P}_{\mathcal{A}}(x_j y) = (\xi_1^i - \xi_1^j)\eta_1 + \cdots (\xi_n^i - \xi_n^j)\eta_n \geqslant 2\delta.$$

But since $\mu(y)$ is stochastic, we must have $0 \leqslant \eta_\ell \leqslant 1$ for all ℓ. This implies that

$$|\xi_1^i - \xi_1^j| + \cdots |\xi_n^i - \xi_n^j| \geqslant 2\delta, \qquad \text{for } i \neq j. \tag{1}$$

In other words, the points ξ_i and ξ_j cannot be too close to each other for the L^1 norm. Coupled with the fact that they are stochastic vectors (and thus live in $[0,1]^n$), this will put a bound on k.

Let $\|x\| = |x_1| + \cdots + |x_n|$ denote the L^1 norm and $B_R(p) = \{x \in \mathbb{R}^n : \|x - p\| < R\}$ denote the L^1 open ball of radius R and center $p \in \mathbb{R}^n$. We will use the fact that $B_R(p)$ has volume cR^n where c only depends on n (in fact $c = \frac{2^n}{n!}$). Now we can rephrase (1) as $\|\xi^i - \xi^j\| \geqslant 2\delta$ which implies that $B_\delta(\xi^i) \cap B_\delta(\xi^j) = \varnothing$ for $i \neq j$. On the other hand, by stochasticity, we have that $\|\xi^i\| = 1$ thus if $x \in B_R(\xi^i)$ then $\|x\| \leqslant \|x - \xi^i\| + \|\xi^i\| < 1 + \delta$ thus $B_\delta(\xi^i) \subseteq B_{1+\delta}(0)$. Therefore,

$$B_\delta(\xi^1) \uplus \cdots \uplus B_\delta(\xi^k) \subseteq B_{1+\delta}(0)$$

where $\uplus$ denotes the disjoint union. By taking the volume, we get that $kc\delta^n \leqslant c(1+\delta)^n$ and thus

$$k \leqslant (1 + \tfrac{1}{\delta})^n.$$

This show that the number of equivalence classes with respect to $\equiv_L$ is finite and therefore L is regular. Furthermore this gives us a bound on the number of states by Theorem 13. It is possible to improve the bound further by noting that the ξ_i are stochastic vectors, therefore they belong to the hyperplane H defined by $x_1 + \cdots x_n = 1$, which is a $n-1$ dimensional subspace. Therefore we get that

$$(B_\delta(\xi^1) \cap H) \uplus \cdots \uplus (B_\delta(\xi^k) \cap H) \subseteq B_{1+\delta}(0) \cap H$$

where all intersections with H are nonempty. We conclude by noticing that if $B_R(p) \cap H$ is nonempty, its volume in H is $c'R^{n-1}$ (where in fact $c' = \frac{\sqrt{n}}{(n-1)!}$). □

The previous theorem suggests that finding an equivalent deterministic finite automaton might increase the number of states. Furthermore, the increase fundamentally depends on the isolation threshold, which we do not know if it lower bounded by a function of n. The next theorem shows that, in fact, it is not.

Theorem 18. *There exists a probabilistic automaton $\mathcal{A}$ with only two states and a sequence $(\lambda_n)_{n\in\mathbb{N}}$ of isolated cut points such that $\mathcal{L}_{\mathcal{A}}(\lambda_n)$ cannot be recognized by a deterministic finite automaton with less than n states.*

Proof. Let $\mathcal{A} = \langle A, Q, S, \mu, T\rangle$ where $A = \{0, 2\}$, $Q = \{s_0, s_1\}$ and

$$S = \begin{bmatrix}1 & 0\end{bmatrix}, \qquad \mu(0) = \begin{bmatrix}1 & 0\\ \frac{2}{3} & \frac{1}{3}\end{bmatrix}, \qquad \mu(2) = \begin{bmatrix}\frac{1}{3} & \frac{2}{3}\\ 0 & 1\end{bmatrix}, \quad \text{and} \quad T = \begin{bmatrix}0\\1\end{bmatrix}.$$

This automaton is very similar to the dyadic automaton of Figure 2a but in base 3. Similarly, we get that if $w \in A^*$ then

$$\mathbb{P}_{\mathcal{A}}(w) = \sum_{i=1}^{|w|} w_i 3^{i-|w|-1}.$$

Contrary to the previous automaton, the set $P = \{\mathbb{P}_{\mathcal{A}}(w) : w \in A^*\}$ is not dense anymore. Indeed, P is included in the *Cantor set* C, that is exactly the set of real numbers of $[0,1]$ that do not require the digit 1. It is well-known that the Cantor set satisfies the following property. For every $p \in C$, there exists $(a, b) \subset [0, 1]$ such that $C \cap (a, b) = \{p\}$, therefore p is isolated. Now fix $n \in \mathbb{N}$ and consider the cut point

$$\lambda_n = 0.2222\cdots2211 = \sum_{i=1}^{n-1} 2 \cdot 3^{-i} + 3^{-n} + 3^{-n-1}.$$

Then the word $2^n \in \mathcal{L}_{\mathcal{A}}(\lambda_n)$ since it has probability of acceptance

$$\mathbb{P}_{\mathcal{A}}(2^n) = \sum_{i=1}^{n} 2 \cdot 3^{i-k-1} > \lambda_n.$$

Conversely, if $w \in A^*$ has length $|w| \leqslant n-1$ then

$$\mathbb{P}_{\mathcal{A}}(w) \leqslant \sum_{i=1}^{|w|} 2 \cdot 3^{-i-|w|-1} \leqslant \sum_{i=1}^{n-1} 2 \cdot 3^{-i} \leqslant \lambda_n.$$

It follows that $\mathbb{P}_{\mathcal{A}}(\lambda_n)$ is nonempty and must reject all words of length less than n. Therefore any deterministic finite automaton that recognizes this language must have at least n states (see Exercise 19). □

Exercise 19. Let L be a nonempty regular language that contains no words of length less than n. Show that any deterministic finite automaton that recognizes L must have at least n states. *Hint: use Theorem 13.*

1.2 Decision problems

We now look at various classical problems on probabilistic automata. Contrary to regular languages, basic questions like emptiness are very hard.

1.2.1 The emptiness problem and its variants

Given a stochastic language, the first question that comes to mind is whether this language is empty or not. Surprisingly, even this simple problem turns out to be undecidable. Note that Rabin defines the language $\mathcal{L}_{\mathcal{A}}(\lambda)$ as words with probability of acceptance *strictly greater* than λ, but some authors prefer to use another convention where the probability is *greater or equal* to λ. To avoid any ambiguity, we distinguish the two problems and follow a recent proof strategy [GO10].

Problem 20 (*Strict Emptiness*)**.** *Given a probabilistic automaton $\mathcal{A}$ and a cut-point λ, decide whether there exists a word w such that $\mathbb{P}_{\mathcal{A}}(w) > \lambda$.*

Problem 21 (*Emptiness*)**.** *Given a probabilistic automaton $\mathcal{A}$ and a cut-point λ, decide whether there exists a wordw such that $\mathbb{P}_{\mathcal{A}}(w) \geqslant \lambda$.*

In fact, both problems are essentially equivalent and reduce to the following variant of the problem where we look for words with a specific probability of acceptance.

Problem 22 (*Equality*)**.** *Given a probabilistic automaton $\mathcal{A}$ and a cut-point λ, decide whether there exists a word w such that $\mathbb{P}_{\mathcal{A}}(w) = \lambda$.*

We will start with the equality problem and reduction from the *Post Correspondence Problem* (PCP) given by Bertoni. Recall that PCP is a classical example of undecidable problem.

Problem 23 (PCP)**.** *Given A a finite alphabet and $\phi_1, \phi_2 : A \to \{0,1\}^*$ two functions that we naturally extend to morphisms over A^*, decide whether there exists $w \in A^*$ such that $\phi_1(w) = \phi_2(w)$.*

It will be important, for a later reduction, to consider a particular sub-class of probabilistic automaton where only certain probabilities appear. An automaton is called *simple* if every transition probability is in $\{0, \frac{1}{2}, 1\}$.

Theorem 24. *The Equality Problem is undecidable, even for simple automata.*

Proof. We will reduce from the PCP: let $\phi_1, \phi_2 : A \to \{0,1\}^*$ be an instance. We modify this instance into φ_1, φ_2 by inserting 10 after every letter of $\phi_i(a)$ so that $\varphi_i(a) \in \{010, 110\}^*$ and therefore $\varphi_i(a) \in \{0,1\}^*10$. Clearly, $\varphi_1(w) = \varphi_2(w)$ if and only if $\phi_1(w) = \phi_2(w)$ so this modification preserves the undecidability.

We will build a probabilistic automaton $\mathcal{A}$ such that $\mathcal{A}$ accepts a word with probability $\frac{1}{2}$ if and only if this PCP instance has a solution. We do so by encoding $\{0,1\}^*$ into probabilities. Similarly to the proof of Theorem 12, define

$$[w] = \sum_{i=1}^{|w|} w_i 2^{-i-1} \qquad \text{for every } w \in \{0,1\}^*.$$

Check that $[\cdot]$ is injective over $\{0,1\}^*10$ and therefore for every words $w \in A^*$,

$$[\phi_1(w)] = [\phi_2(w)] \text{ if and only if } \phi_1(w) = \phi_2(w). \tag{2}$$

We can now consider the following two automata for $i \in \{1,2\}$: $\mathcal{A}_i = \langle A, Q, S, \mu_i, T\rangle$ where $Q = \{p,q\}$ and

$$S = \begin{bmatrix}1 & 0\end{bmatrix}, \qquad \mu(a) = \begin{bmatrix}2^{-|\phi_i(w)|} & [\phi_i(w)] \\ 0 & 1\end{bmatrix}, \quad \text{and} \quad T = \begin{bmatrix}0\\1\end{bmatrix}.$$

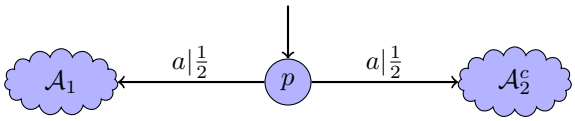
One checks that $\mu(a)$ is substochastic* by checking that $[w] \leqslant 1 - 2^{-|w|}$ for every $w \in \{1,0\}^*10$. For every $u, v \in \{0,1\}^*$, check that $[uv] = [u] + 2^{-|u|}[v]$. Then check that if $a, b \in A$ we have that

$$\mu_i(a)\mu_i(b) = \begin{bmatrix}2^{-|\phi_i(a)|-|\phi_i(b)} & [\phi_i(a)] + 2^{-|\phi_i(a)|}[\phi_i(b)] \\ 0 & 1\end{bmatrix} = \begin{bmatrix}2^{-|\phi_i(ab)|} & [\phi_i(ab)] \\ 0 & 1\end{bmatrix}$$

and therefore for every word $w \in A^*$,

$$\mathbb{P}_{\mathcal{A}_i}(w) = S\begin{bmatrix}2^{-|\phi_i(w)|} & [\phi_i(w)] \\ 0 & 1\end{bmatrix}T = [\phi_i(w)].$$

From $\mathcal{A}_2$, we can create another automaton $\mathcal{A}_2^c$ such that $\mathbb{P}_{\mathcal{A}_2^c}(w) = 1 - \mathbb{P}_{\mathcal{A}_i}(w)$ for every $w \in A^*$ (see Exercise 25). We can create the automaton $\mathcal{B}$ below that computes the "average" of $\mathcal{A}_1$ and $\mathcal{A}_2^c$ (where the initial transitions from p to automata are labelled by all possible letters $a \in A$).



We then obtain that for every letter $a \in A$ and every word $w \in A^*$,

$$\mathbb{P}_{\mathcal{B}}(aw) = \tfrac{1}{2} \iff \tfrac{1}{2}\mathbb{P}_{\mathcal{A}_1}(w) + \tfrac{1}{2}\mathbb{P}_{\mathcal{A}_2^c}(w) = \tfrac{1}{2} \iff [\varphi_1(w)] = [\varphi_2(w)] \iff \varphi_1(w) = \varphi_2(w)$$

using (2). Therefore $\mathcal{B}$ accepts a word with probability $\frac{1}{2}$ if and only if the PCP instance has a solution.

In this proof, note that automata $\mathcal{B}$ only uses dyadic transition probabilities, and therefore we can make it simple by introduing more states (see Exercise 26). □

Exercise 25. Given a probabilistic automaton $\mathcal{A}$, build another automaton $\mathcal{A}^c$ such that $\mathbb{P}_{\mathcal{A}^c}(w) = 1 - \mathbb{P}_{\mathcal{A}}(w)$ for every word w.

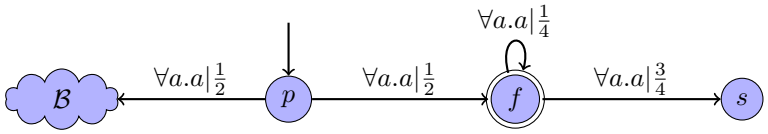
Exercise 26. Recall that a *dyadic number* (or *dyadic rational*) is a number of the form $a2^{-p}$ for some $a \in \mathbb{Z}$ and $p \in \mathbb{N}$. Given a probabilistic automaton $\mathcal{A}$ where all transition probabilities are dyadic, build a simple automaton $\mathcal{B}$ such that $\mathbb{P}_{\mathcal{B}}(w) = \mathbb{P}_{\mathcal{A}}(w)$ for every word w.

Proposition 27. *Given a simple probabilistic automaton $\mathcal{A}$, one can compute (simple) probabilistic automata $\mathcal{B}$ and $\mathcal{C}$ such that the following propositions are equivalent:*

- *there exists a word w such that $\mathbb{P}_{\mathcal{A}}(w) = \frac{1}{2}$,*
- *there exists a word w such that $\mathbb{P}_{\mathcal{B}}(w) \geqslant \frac{1}{4}$,*
- *there exists a word w such that $\mathbb{P}_{\mathcal{C}}(w) > \frac{1}{8}$.*

Proof. The idea of the proof is that $x = \frac{1}{2}$ if and only if $x(1-x) \geqslant \frac{1}{4}$. Therefore from $\mathcal{A}$, build $\mathcal{A}^c$ such that $\mathbb{P}_{\mathcal{A}^c}(w) = 1 - \mathbb{P}_{\mathcal{A}}(w)$ (see Exercise 25). Then build $\mathcal{B}$ the product automaton of $\mathcal{A}$ and $\mathcal{A}^c$, which satisfies $\mathbb{P}_{\mathcal{B}}(w) = \mathbb{P}_{\mathcal{A}}(w)\mathbb{P}_{\mathcal{A}^c}(w)$ (see Exercise 28). By construction, all transition probabilities of $\mathcal{B}$ are already multiple of $\frac{1}{4}$.

To build $\mathcal{C}$, we start by noticing that since all transition probabilities of $\mathcal{B}$ are multiple of $\frac{1}{4}$, $\mathbb{P}_{\mathcal{B}}(w)$ is a multiple of $4^{-|w|}$ for every word w. Thus $\mathbb{P}_{\mathcal{B}}(w) \geqslant \frac{1}{4}$ if and only if $\mathbb{P}_{\mathcal{B}}(w) > \frac{1}{4} - 4^{-|w|}$. Consider the automaton $\mathcal{C}$ below, where $\forall a.a|x$ means that we add transitions for all possible letters a.



*It is possible to build stochastic matrices directly by taking $S = \begin{bmatrix}1 & 0\end{bmatrix}$, $\mu_i(a) = \begin{bmatrix}1-[\phi_i(a)] & [\phi_i(a)] \\ 1-[1\phi_i(a)] & [1\phi_i(a)]\end{bmatrix}$ and $T_1 = \begin{bmatrix}0\\1\end{bmatrix}, T_2 = \begin{bmatrix}1\\0\end{bmatrix}$. This requires to tweak $[\cdot]$ into $[w] = \sum_{i=1}^{|w|} w_i 2^{i-|w|-1}$ and ϕ_i to insert a 1 before every letter so that $\varphi_i(a) \in 1\{0,1\}^*$.

One checks that for every letter a and every word w, we have that

$$\mathbb{P}_{\mathcal{C}}(aw) = \tfrac{1}{2}\mathbb{P}_{\mathcal{B}}(w) + \tfrac{1}{2}4^{-|w|}$$

and therefore

$$\mathbb{P}_{\mathcal{C}}(aw) > \tfrac{1}{8} \iff \mathbb{P}_{\mathcal{B}}(w) > \tfrac{1}{4} - 4^{-|w|} \iff \mathbb{P}_{\mathcal{B}}(w) \geqslant \tfrac{1}{4}.$$

□

Exercise 28. Given two probabilistic automata $\mathcal{A}$ and $\mathcal{B}$ over the same alphabet, build a a product automaton $\mathcal{C}$ that satisfies $\mathbb{P}_{\mathcal{C}}(w) = \mathbb{P}_{\mathcal{A}}(w)\mathbb{P}_{\mathcal{B}}(w)$ for every word w. Show that if $\mathcal{A}$ and $\mathcal{B}$ are simple then all transition probabilities of $\mathcal{C}$ are multiple of $\frac{1}{4}$. *Hint:* consider the cartesian product of the states.

As a consequence of Theorem 24 and Proposition 27, we get:

Theorem 29. *The emptiness and strict emptiness problems are undecidable, even for simple automata and cut-point* $\frac{1}{2}$.

Exercise 30. Given $\mathcal{A}$ a probabilistic automaton and a rational cut-point $\lambda \in (0,1)$, build automata $\mathcal{B}$ and $\mathcal{C}$ such that $\mathbb{P}_{\mathcal{A}}(w) \geqslant \lambda$ (resp. $\mathbb{P}_{\mathcal{A}}(w) > \lambda$) if and only if $\mathbb{P}_{\mathcal{B}}(w) \geqslant \frac{1}{2}$ (resp. $\mathbb{P}_{\mathcal{C}}(w) \geqslant \frac{1}{2}$). Show that if $\mathcal{A}$ is simple and λ is dyadic then you can make $\mathcal{B}$ and $\mathcal{C}$ simple.

1.3 The isolation problem

We saw in Section 1.1.3 that isolated cut-points are very special since they define regular languages. On the other hand, the emptiness language is undecidable in general for probabilistic automata but decidable for finite automata. Therefore, if we can detect that a cut-point is isolated, it would give us a way to decide emptiness in certain cases.

Problem 31 (*Isolation*)**.** *Given a probabilistic automaton $\mathcal{A}$ and a cut-point λ, decide whether λ is isolated with respect to $\mathcal{A}$.*

Unfortunately, this problem is undecidable in general and even when the threshold is fixed. An elegant way to prove this is to reduce to a variant of the PCP problem for infinite words, which is also undecidable.

Problem 32 ($\omega-PCP$)**.** *Given A a finite alphabet and $\phi_1, \phi_2 : A \to \{0,1\}^*$ two functions that we naturally extend to morphisms over A^*, decide whether there exists $w \in A^{\mathbb{N}}$ such that $\phi_1(w) = \phi_2(w)$.*

Exercise 33. Show that $\omega-$PCP is undecidable.

In particular, we will use a classical feature of the $\omega-$PCP problem: if an instance is not solvable, then there is uniform bound on the how far the first different letter can be.

Lemma 34. *Let $\phi_1, \phi_2 : A \to \{0,1\}^*$ be an instance of the $\omega-PCP$ that has no solution. Then there exists $n_0 \in \mathbb{N}$ such that for every infinite (or non-empty finite) word w, there exists $i \leqslant n_0$ such that $\phi_1(w)_i \neq \phi_2(w)_i$. In other words, $\phi_1(w)$ and $\phi_2(w)$ differ already in their first n_0 letters, and n_0 is independent of w.*

Proof. Consider the tree where the root is labelled $(\varepsilon, \varepsilon)$ and given a node (u, v) of the tree, if $u_i = v_i$ for all $i \leqslant \min(|u|, |v|)$, then this node has children $(u\phi_1(a), v\phi_2(a))$ for all $a \in A$. In other words, we write on the nodes the result of finite labelling of the $\omega-$PCP and we continue only if we haven't found a differing letter (but labels are allowed to differ in length, in which case we only compare up to the shortest one). This tree is finitely branching since each node has 0 or $|A|$ children and A is finite. This tree has no infinite path for it would imply that this instance of $\omega-$PCP has a solution. Therefore by König's lemma, the tree is finite. Let n_0 be the longest length of a word that appears in a label of the tree. Since the labels of the nodes are of the form $(\phi_1(w), \phi_2(w))$, this shows the result. □

Theorem 35. *The isolation problem is undecidable, even for a fixed rational cut-point* $0 < \lambda < 1$.

Proof. We will show the result for $\lambda = \frac{1}{2}$, this can be extended to any rational $\lambda \in (0,1)$ by Exercise 30.

The problem will essentially be the same as for Theorem 24 with a twist. Let $\phi_1, \phi_2 : A \to \{0,1\}^*$ be an instance of the $\omega-$PCP. We modify this instance so that $\phi_i(w) \in \{0,1\}^*1$ for every non-empty word w. This can be done by adding a "1" after each letter of $\phi_i(a)$ for every $a \in A$. Clearly, this does not change the undecidability of $\omega-$PCP.

Like in the proof of Theorem 24, we define $[w] = \sum_{i=1}^{|w|} w_i 2^{-i-1}$ for every $w \in \{0,1\}^*$ and build a probabilistic automaton $\mathcal{B}$ such that $\mathbb{P}_{\mathcal{B}}(aw) = \frac{1}{2} + \frac{1}{2}([\phi_1(w)] - [\phi_2(w)])$ for every $a \in A$ and $w \in A^*$. Recall that $[wx] = [w] + 2^{-|w|}[x]$ for all words w, x. We will now show that $\frac{1}{2}$ is isolated if and only if this instance of $\omega - PCP$ is not solvable.

Assume that this instance has a solution $w \in A^{\mathbb{N}}$. Let $n \in \mathbb{N}$, then there exists a finite prefix u of w such that $|\phi_1(u)| \geqslant n$ and $|\phi_2(u)| \geqslant n$ (since $\phi_1(w)$ and $\phi_2(w)$ are infinite words). Since the instance is solvable, $\phi_1(w) = \phi_2(w)$ and

thus the first n letters of $\phi_1(u)$ and $\phi_2(u)$ are the same, *i.e.* $\phi_1(u) = px$ and $\phi_2(u) = py$ for some $p \in A^n$ and $x, y \in A^*$. But then

$$\begin{aligned}
|[\phi_1(u)] - [\phi_2(u)]| &= |[px] - [py]| \\
&= |[p] + 2^{-|p|}[x] - [p] - 2^{-|p}[y]| \\
&= 2^{-n}|[x] - [y]| \\
&\leqslant 2^{1-n} && \text{since } [x], [y] \in [0,1].
\end{aligned}$$

Therefore,

$$|\mathbb{P}_{\mathcal{B}}(aw) - \tfrac{1}{2}| = \tfrac{1}{2}\,|[\phi_1(w)] - [\phi_2(w)]| \leqslant 2^{-n}.$$

This shows that $\frac{1}{2}$ is not isolated, since there are words accepted with probabilities arbitrarily close to the cut-point.

Conversely, assume that this instance has no solution. Then by Lemma 34, there exists $n_0 \in \mathbb{N}$ such that for every infinite (or non-empty finite) word $w \in A^{\mathbb{N}}$, there exists $i \leqslant n_0$ such that $\phi_1(w)_i \neq \phi_2(w)_i$. Recall that we modified the instances so that $\phi_i(w) \in \{0,1\}^*1$ for every word w. Let $w \in A^*$, then we can write $\phi_1(w) = ua1x$ and $\phi_2(w) = ub1y$ where $|u| \leqslant n_0$, $a, b \in \{0,1\}$ are distincts and $x, y \in \{0,1\}^*$. It follows that

$$\begin{aligned}
|[\phi_1(w)] - [\phi_2(w)]| &= |[ua1x] - [ub1y]| \\
&= |(a-b)2^{-|u|} + ([x] - [y])2^{-|u|-2}| \\
&\geqslant |a-b|2^{-|u|} - |[x] - [y]|2^{-|u|-2} \\
&\geqslant 2^{-|u|} - 2 \cdot 2^{-|u|-2} && \text{since } [x], [y] \in [0,1] \\
&\geqslant 2^{1-n_0} && \text{since } |u| \leqslant n_0.
\end{aligned}$$

It follows that for all $w \in A^*$,

$$|\mathbb{P}_{\mathcal{B}}(aw) - \tfrac{1}{2}| = \tfrac{1}{2}\,|[\phi_1(w)] - [\phi_2(w)]| \geqslant 2^{-n_0}.$$

Since n_0 is independent of w, this shows that $\frac{1}{2}$ is isolated. □

There is a slight discrepancy in Theorem 35 for the case $\lambda = 0$ and $\lambda = 1$. It is clear that those two cases are symmetric, by taking the complement of the automaton. If we fix the cut-point to 1, the isolation problem is the same asking if there are words accepted with probabilities arbitrarily close to 1. This is related to asking what is the value of an automaton. The *value* of probabilistic automaton $\mathcal{A}$ over alphabet A is

$$\mathrm{val}(\mathcal{A}) = \sup\{\mathbb{P}_{\mathcal{A}}(w) : w \in A^*\}.$$

In other words, is it the supremum of the probability of acceptance over all possible input words.

Problem 36 (*Value* 1)**.** *Given a probabilistic automaton $\mathcal{A}$, decide whether $\mathcal{A}$ has value* 1, i.e. *for every $\varepsilon > 0$, there exists w such that $\mathbb{P}_{\mathcal{A}}(w) > 1 - \varepsilon$.*

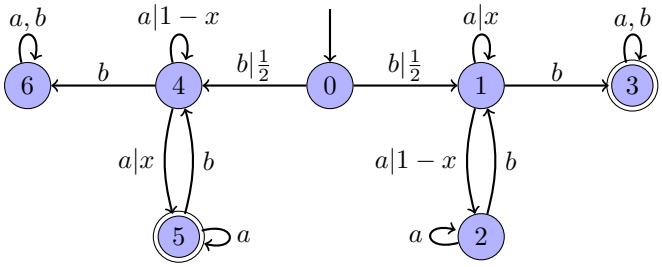


Figure 3: Auxiliary automaton for the value 1 problem.

Proposition 37. *Let $x \in (0,1)$, then the automaton $\mathcal{A}_x$ from Figure 3 has value* 1 *if and only if $x > \frac{1}{2}$.*

Proof. The automaton $\mathcal{A}_x$ is in fact built similarly to $\mathcal{C}$ in Figure 2c but with probabilities x and $1-x$ instead of $\frac{1}{2}$, and by putting two copies together. First notice that observe that

$$\mathbb{P}_{\mathcal{A}_x}\left(1 \xrightarrow{a^n b} 1\right) = 1 - x^n, \qquad \mathbb{P}_{\mathcal{A}_x}\left(1 \xrightarrow{a^n b} 3\right) = x^n, \qquad \mathbb{P}_{\mathcal{A}_x}\left(4 \xrightarrow{a^n b} 4\right) = 1 - (1-x)^n, \qquad \mathbb{P}_{\mathcal{A}_x}\left(4 \xrightarrow{a^n b} 6\right) = (1-x)^n.$$

Let $n_1, \dots, n_k \geqslant 1$ and $w = a^{n_1} b \cdots a^{n_k} b$, then (see Exercise 38)

$$\mathbb{P}_{\mathcal{A}_x}\left(1 \xrightarrow{w} 3\right) = 1 - \prod_{i=1}^{k}(1 - x^{n_i}) \qquad \text{and} \qquad \mathbb{P}_{\mathcal{A}_x}\left(4 \xrightarrow{w} 6\right) = 1 - \prod_{i=1}^{k}(1 - (1-x)^{n_i}) \leqslant \sum_{i=1}^{k}(1-x)^{n_i}. \tag{3}$$

If $x \leqslant \frac{1}{2}$ then $\mathbb{P}_{\mathcal{A}_x}\left(1 \xrightarrow{w} 3\right) \leqslant \mathbb{P}_{\mathcal{A}_x}\left(4 \xrightarrow{w} 6\right)$ and since 6 is not accepting, w cannot be accepted with arbitrarily high probability. Conversely, if $x > \frac{1}{2}$ then for every $\varepsilon > 0$ we will build a sequence $(n_k)_k$,

$$\sum_{k=1}^{\infty}(1-x)^{n_k} \leqslant \varepsilon \qquad \text{and} \qquad \sum_{k=1}^{\infty} x^{n_k} = \infty.$$

Let $c \in \mathbb{R}$ to be fixed later, and $n_k = c + \ln_x \frac{1}{k}$. Check that $\sum_{k=1}^{\infty} x^{n_k} = x^c \sum_{k=1}^{\infty} \frac{1}{k} = \infty$. On the other hand, since $x \in (0, \frac{1}{2})$ and by continuity, there exists $\beta > 1$ such that $1-x = x^\beta$. Then $\sum_{k=1}^{\infty}(1-x)^{n_k} = \sum_{k=1}^{\infty} x^{\beta n_k} = x^{\beta c} \sum_{k=1}^{\infty} \frac{1}{k^\beta} \leqslant \varepsilon$ if we choose c sufficiently small. If we now consider the word $w = a^{n_1} b \cdots a^{n_k} b$ for some $k \in \mathbb{N}$ then

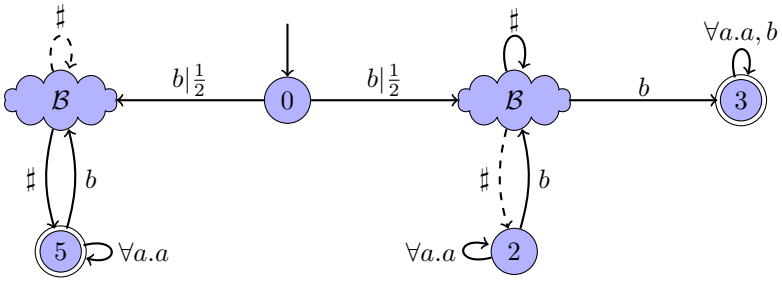
$$\begin{aligned}
\mathbb{P}_{\mathcal{A}_x}(bw) &= \tfrac{1}{2}\left(1 - \mathbb{P}_{\mathcal{A}_x}\left(4 \xrightarrow{w} 6\right)\right) + \tfrac{1}{2}\mathbb{P}_{\mathcal{A}_x}\left(1 \xrightarrow{w} 3\right) \\
&\geqslant 1 - \tfrac{1}{2}\sum_{i=1}^{k}(1-x)^{n_i} - \tfrac{1}{2}\prod_{i=1}^{k}(1-x^{n_i}) \\
&\geqslant 1 - \tfrac{1}{2}\sum_{i=1}^{\infty}(1-x)^{n_i} - \tfrac{1}{2}\exp\left(\sum_{i=1}^{k}\ln(1-x^{n_i})\right) \\
&\geqslant 1 - \tfrac{1}{2}\varepsilon - \tfrac{1}{2}\exp\left(-\sum_{i=1}^{k} x^{n_i}\right)
\end{aligned}$$

but since $\sum_{i=1}^{\infty} x^{n_i} = \infty$, we can find a $k \in \mathbb{N}$ such that $\exp\left(-\sum_{i=1}^{k} x^{n_i}\right) \leqslant \varepsilon$ and then $\mathbb{P}_{\mathcal{A}_x}(bw) \geqslant 1 - \varepsilon$. □

Exercise 38. In the proof of Proposition 37, prove (3).

Theorem 39. *The value* 1 *problem is undecidable.*

Proof. We will reduce from the strict emptiness problem with fixed cut-point $\frac{1}{2}$. Let $\mathcal{B}$ be a probabilistic automaton over alphabet A, which we assume does not contain a and b. We will now combine $\mathcal{A}_x$ from Figure 3 and $\mathcal{B}$. The idea is to replace the transitions in $\mathcal{A}_x$ that involve x by copies of $\mathcal{B}$. Consider the automaton $\mathcal{C}$ below, over alphabet $A \cup \{b, \sharp\}$, where the transitions coming out of $\mathcal{B}$ are from the accepting states of $\mathcal{B}$, and the *dashed* transitions coming out of $\mathcal{B}$ are from the *non-accepting* the states. Furthermore, the only accepting states of $\mathcal{C}$ are 5 and 3. The notation $\forall a.a$ means that we add all transitions labelled by $a \in A$. For convenience, we haven't added the sink state 6 and all the transitions to it for missing/invalid letters.

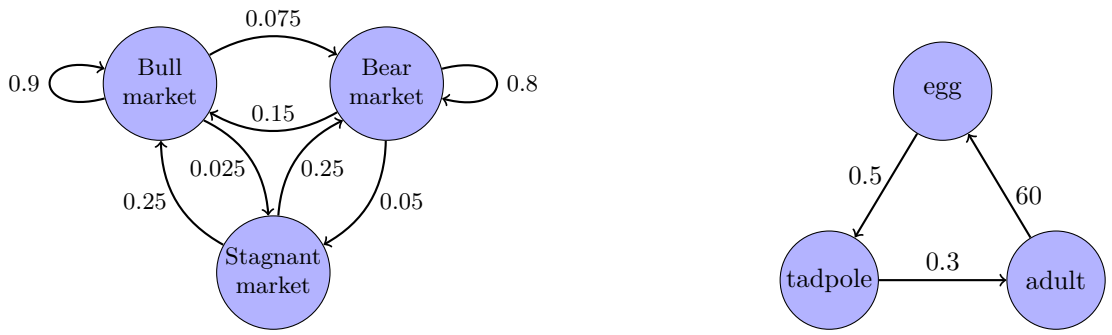


We now claim that $\mathcal{C}$ has value 1 if and only if $\exists w \in A^*$ such that $\mathbb{P}_{\mathcal{B}}(w) > \frac{1}{2}$.

Assume there exists such a word w and let $x = \mathbb{P}_{\mathcal{B}}(w)$. Let $\varepsilon > 0$, then $\mathcal{A}_x$ has value 1 by Proposition 37, so there exists $n_1, \dots, n_k$ such that $\mathbb{P}_{\mathcal{A}_x}(ba^{n_1} b \cdots a^{n_k} b) \geqslant 1 - \varepsilon$. But now observe that

$$\mathbb{P}_{\mathcal{C}}(b(w\sharp)^{n_1} b \cdots (w\sharp)^{n_k} b) = \mathbb{P}_{\mathcal{A}_x}(ba^{n_1} b \cdots a^{n_k} b) = 1 - \varepsilon.$$

Indeed, for the left copy of $\mathcal{B}$, after reading a word $w\sharp$, we get to state 5 with probability $\mathbb{P}_{\mathcal{B}}(w) = x$ and to the initial state of $\mathcal{B}$ with the remaining probabilistic $1 - x$, thus perfectly emulating the a transition of $\mathcal{A}_x$. The right copy of $\mathcal{B}$ is similar except that we inverted the accepting states so the transition probabilities are swapped.



(a) A Markov chain representing a hypothetical stock market (b) A linear dynamical system modelling a frog population

Figure 4: Examples of Markov chain and linear dynamical systems

Conversely, assume that $\mathbb{P}_{\mathcal{B}}(w) \leqslant \frac{1}{2}$ for every word $w \in A^*$. It is clear that any word accepted by $\mathcal{C}$ with positive probability must be of the form $w' = u_0 v_0 \sharp u_1 v_1 \sharp \cdots u_k v_k$ where $u_i \in b^*$ and $v_i \in A^*$. Let $x = \max_{i=1}^{k} \mathbb{P}_{\mathcal{B}}(v_i)$, then $x \leqslant \frac{1}{2}$ by assumption and by construction, any transition $v_i\sharp$ perfectly emulates a a transition of $\mathcal{A}_x$ (but with probability $\leqslant x$). Thus $\mathbb{P}_{\mathcal{C}}(w') \leqslant \mathbb{P}_{\mathcal{A}_x}(u_0 a u_1 \cdots u_k a) \leqslant \operatorname{val} \mathcal{A}_x$. But $\operatorname{val} \mathcal{A}_x < 1$ by Proposition 37 since $x \leqslant \frac{1}{2}$, this shows that $\operatorname{val} \mathcal{C} \leqslant \operatorname{val} \mathcal{A}_x < 1$. □

2 Markov chains and linear dynamical systems

A *Markov chain* is a particular case of probabilistic automata where the alphabet is unary. In this case, we can simplify the presentation and describe a Markov chain in *dimension* d by a tuple $\mathcal{M} = \langle S, A, T \rangle$ where

- $S \in [0,1]^{1 \times d}$ is stochastic (row) vector of *initial probabilities*,
- $T \in \{0,1\}^{d \times 1}$ is a $0-1$ (column) vector of *accepting states*,
- $A \in [0,1]^{d \times d}$ is a stochastic matrix of *transition probabilities*.

Similarly to probabilistic automata, we usually assume that initial probabilities and transition probabilities are rational numbers. In the case of Markov chains, there is a unique probability of acceptance for every length. It is given for every $n \in \mathbb{N}$ by

$$\mathbb{P}_{\mathcal{M}}(n) = SA^nT.$$

More generally, we will consider *linear dynamical systems* (*LDS*) $\langle S, A, T \rangle$ where we lift the restriction that I and A be stochastic. In particular, the values of a LDS do not need to be within $[0,1]$.

Example 40. Figure 4a illustrates an hypothetical stock market that can exhibit three behaviors during a week: bull, bear or stagnant. For example, following a bull week, the market has 90% chances of being bull the next week but it will become bear with a 7.5% probability. If we start from an initial distribution over the three states and put it in a vector I, and let A be the transition matrix, then SA^n gives the probability distribution over the three states after n weeks. We can thus analyse the long-term behavior of the system. For example if we take $T = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^t$ then SA^nT gives the probability of being in a certain state (say bull) after n weeks. The emptiness problem now becomes: does there exists n such that $SA^nT \geqslant \lambda$, in other words, is there is any week where the probability of the market being bull is higher than λ.

Example 41. Figure 4b illustrates a simplified model for the dynamics of a frog population. Frogs have three life stages: egg, tadpole and adult. Every year, 50% of the eggs survive to become tadpoles, 30% of the tadpoles become adults and every pair of adults produces 120 eggs and dies. The corresponding transition matrix, also known as the *Leslie* matrix, is

$$A = \begin{bmatrix} 0 & 0 & 60 \\ 0.5 & 0 & 0 \\ 0 & 0.3 & 0 \end{bmatrix}.$$

Starting from an initial state of the pond, for example 50 eggs, 20 tadpoles and 2 adults, we can get the state of the population after n years by computing SA^n where $S = \begin{bmatrix} 50 & 20 & 2 \end{bmatrix}$. We can study the long-term behavior of this system, for example the total population size is given by SA^nT where $T = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^t$.

2.1 Linear recurrent sequences

An alternative point of view is to consider the sequence $(u_n)_{n\in\mathbb{N}}$ given by $u_n = \mathbb{P}_{\mathcal{M}}(n)$. A useful property of this sequence is that it is linear. Formally, a *linear recurrent sequence* (LRS) of *order* k is any sequence $(u_n)_{n\in\mathbb{N}}$ that satisfies the recurrence relation

$$u_{n+k} = a_{k-1}u_{n+k-1} + \cdots + a_0 u_n$$

for all $n \in \mathbb{N}$, for some numbers $a_0, \ldots, a_{k-1} \in \mathbb{R}$. When all numbers u_n and a_i are rational, we say that it is a *rational* LRS, and if all numbers u_n and a_i are integers, then it is an *integer* LRS. There is a strong connection between LRS and matrix powers that comes from linear algebra.

Theorem 42 (*Cayley–Hamilton*)**.** *Let $A \in \mathbb{R}^{d\times d}$ be a matrix and let $p(\lambda) = \det(\lambda I_d - A)$ be its* characteristic polynomial*, then $p(A) = 0$. In particular, A^d is a linear combination of $I_d, A, \ldots, A^{d-1}$.*

Proof. We admit the proof and simply show how the last statement follows from $p(A) = 0$. Indeed, $p(\lambda)$ is a determinant of $d \times d$ matrix, thus it is a polynomial of degree d in λ. Furthermore, it is not hard to see that p is *monic*, *i.e.* $p(\lambda) = \lambda^d + q(\lambda)$ where q has degree at most $d-1$. Therefore, $p(A) = 0$ implies that $A^d = -q(A) = \sum_{i=0}^{d-1} a_i A^i$ where the a_i are the coefficients of q. □

Proposition 43. *Let $d \in \mathbb{N}$, let $S \in \mathbb{Q}^{1\times d}$, $A \in \mathbb{Q}^{d\times d}$ and $T \in \{0,1\}^{d\times 1}$. Then the sequence $(SA^nT)_{n\in\mathbb{N}}$ is a rational LRS of order d. Furthermore if all entries of I and A are integers, then it is an integer LRS. In particular, if $\mathcal{M}$ is a Markov chain, then $(\mathbb{P}_{\mathcal{M}}(n))_{n\in\mathbb{N}}$ is a rational LRS. Conversely, if $(u_n)_{n\in\mathbb{N}}$ is rational LRS of order d, then there exists a LDS $\langle S, A, T\rangle$ of dimension d such that $u_n = SA^nT$ for all $n \in \mathbb{N}$. Furthermore, if $(u_n)_n$ is an integer LRS then S, A and T have integer coefficients.*

Proof. By Cayley–Hamilton theorem, A^d is a linear combination of $I_d, A, \ldots, A^{d-1}$ so we can find $a_0, \ldots, a_{d-1} \in \mathbb{Q}$ such that

$$A^d = \sum_{i=0}^{d-1} a_i A^i.$$

For every $n \in \mathbb{N}$, let $u_n = SA^{n+d}T$, then we have that

$$u_{n+d} = SA^{n+d}T = SA^nA^dT = SA^n\left(\sum_{i=0}^{d-1} a_i A^i\right)T = \sum_{i=0}^{d-1} a_i SA^{n+i}T = \sum_{i=0}^{d-1} a_i u_{n+i}.$$

Thus $(u_n)_n$ is a LRS. If all entries of I and A are rational, then the characteristic polynomial p of A has rational entries thus the coefficients a_i are rationals. Similarly if I and A are rational, then the coefficients of p are integers.

Conversely, if $(u_n)_n$ is a LRS of order d, let $a_0, \ldots, a_{d-1}$ be the coefficients of the recurrence relation. Consider the matrices

$$S = \begin{bmatrix}1 & 0 & \cdots & 0\end{bmatrix}, \qquad A = \begin{bmatrix}0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 1 & \ddots & 0\\ \vdots & \vdots & \ddots & \ddots & 0\\ 0 & \cdots & \cdots & 0 & 1\\ a_0 & a_1 & \cdots & \cdots & a_{d-1}\end{bmatrix}, \qquad T = \begin{bmatrix}u_0\\ u_1\\ \vdots\\ u_{d-1}\end{bmatrix}.$$

Then we check that for every $n \in \mathbb{N}$,

$$A\begin{bmatrix}u_n\\ u_{n+1}\\ \vdots\\ u_{n+d-1}\end{bmatrix} = \begin{bmatrix}u_{n+1}\\ \vdots\\ u_{n+d-1}\\ a_0u_n + \cdots + a_{d-1}u_{n+d-1}\end{bmatrix} = \begin{bmatrix}u_{n+1}\\ \vdots\\ u_{n+d-1}\\ u_{n+d}\end{bmatrix} \quad \text{and thus} \quad A^nT = \begin{bmatrix}u_n\\ u_{n+1}\\ \vdots\\ u_{n+d-1}\end{bmatrix} \tag{4}$$

follows by induction. This implies that $SA^nT = u_n$ and concludes. □

Remark 44. The proof of Proposition 43 could give the impression that any Markov chain or LDS $\langle S, A, T\rangle$ verifies equation (4), *i.e.* it shifts consecutive terms by one. *This is not the case in general*, see Exercise 45.

Exercise 45. Consider the following two LDS $\langle S, A_1, T\rangle$ and $\langle S, A_2, T\rangle$:

$$S = \begin{bmatrix}1 & 0\end{bmatrix}, \qquad T = \begin{bmatrix}0\\1\end{bmatrix}, \qquad A = \begin{bmatrix}0 & 1\\ -1 & 2\end{bmatrix}, \qquad \text{and} \qquad B = \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}.$$

By following the proof of Proposition 43, let $u_n = SA^nT$, find the recurrence relation (of order 2) satisfied by u_n, find u_0 and u_1 and give an explicit expression for u_n. Find an explicit expression for B^n and show that $u_n = SB^nT$. Then prove that

$$A^n \begin{bmatrix} u_n \\ u_{n+1} \end{bmatrix} = \begin{bmatrix} u_{n+1} \\ u_{n+2} \end{bmatrix} \qquad \text{but} \qquad B^n \begin{bmatrix} u_n \\ u_{n+1} \end{bmatrix} \neq \begin{bmatrix} u_{n+1} \\ u_{n+2} \end{bmatrix}.$$

Proposition 46. *Let $\lambda \in \mathbb{Q}$, $(u_n)_n$ and (v_n) be two rational LRS. Then $(\lambda)_n$, $(\lambda u_n)_n$, $(u_n + v_n)_n$, $(u_n v_n)_n$.*

Proof. The first item is trivial since it satisfies $u_{n+1} = u_n$. Let $(u_n)_n$ and $(v_n)_n$ be two LRS of order d (we can always increase the order artificially) and let $a_0, \dots, a_{d-1}$ and $b_0, \dots, b_{d-1}$ be the coefficients of the recurrence relation. Let $w_n = \lambda u_n$, then

$$w_{n+d-1} = \lambda u_{n+d-1} = \lambda \sum_{i=0}^{d-1} a_i u_{n+i} = \sum_{i=0}^{d-1} a_i \lambda u_{n+i} = \sum_{i=0}^{d-1} a_i w_{n+i}$$

thus $(w_n)_n$ is a LRS. By Proposition 43, there exists S_1, S_2, A_1, A_2, T_1 and T_2 such that $u_n = S_1 A_1^n T_1$ and $v_n = S_2 A_2^n T_2$. Consider

$$\hat{S} = \begin{bmatrix} S_1 & S_2 \end{bmatrix}, \qquad \hat{A} = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}, \qquad \hat{T} = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix}.$$

Then we have that

$$\hat{S}\hat{A}^n\hat{T} = \begin{bmatrix} S_1 & S_2 \end{bmatrix} \begin{bmatrix} A_1^n & 0 \\ 0 & A_2^n \end{bmatrix} \begin{bmatrix} T_1 \\ T_2 \end{bmatrix} = S_1 A_1^n T_1 + S_2 A_2^n T_2 = u_n + v_n.$$

Thus by Proposition 43, $(u_n + v_n)_n$ is a LRS. Similarly, consider

$$\hat{I} = S_1 \otimes S_2, \qquad \hat{A} = A_1 \otimes A_2, \qquad \hat{T} = T1 \otimes T_2$$

where $\otimes$ denotes the Kronecker product (see solution of Exercise 28). Then by the mixed-product property,

$$\hat{S}\hat{A}^n\hat{T} = (S_1 \otimes S_2)(A_1 \otimes A_2)^n (T_1 \otimes T_2) = (S_1 A_1^n T_1) \otimes (S_2 A_2^n T_2) = u_n v_n.$$

Thus by Proposition 43, $(u_n v_n)_n$ is a LRS. □

Another interesting feature of LRS is that we can provide an explicit expression for its general term. Let $(u_n)_n$ be a LRS and let $a_0, \dots, a_{d-1}$ be the coefficients of its recurrence relation. We define the *characteristic polynomial* of the sequence to be

$$p(x) = x^d - a_1 x^{d-1} \cdots a_{d-1} x - a_d.$$

Proposition 47. *Let $(u_n)_n$ be a LRS and p its characteristic polynomial. Let $\lambda_1, \dots, \lambda_d$ be the (possibly repeated) (complex) roots of p. Then there are univariate polynomials $A_1, \dots, A_d$ of degree at most d such that*

$$u_n = A_1(n)\lambda_1^n + \cdots + A_d(n)\lambda_d^n. \tag{5}$$

In particular, $(u_n)_n$ is linear combination of the sequences $n^k \lambda_i^n$ for $i \in \{1, \dots, d\}$ and $0 \leqslant k < d$. Furthermore, all the coefficients that appear in the A_i are algebraic numbers[†]. Conversely, any sequence of this form is a LRS.

Proof. Put A in Jordan Normal Form (see Proposition 48) below, then $A = PJP^{-1}$ where $J = \operatorname{diag}(J_1, \dots, J_k)$. It follows that $A^n = PJ^nP^{-1}$ and $J^n = \operatorname{diag}(J_1^n, \dots, J_k^n)$. It is easy to check by induction that a block J_i of dimension k satisfies

$$J_i^n = \begin{bmatrix} \lambda_i^n & \binom{n}{1}\lambda_i^{n-1} & \cdots & \binom{n}{k}\lambda_i^{n-k} \\ & \ddots & \vdots & \\ & & \ddots & \binom{n}{1}\lambda_i^{n-1} \\ & & & \lambda_i^n \end{bmatrix}$$

and therefore the entries of J_i^n are a linear combination of $n^k \lambda_i^n$. Putting everything together, we get the result. □

Proposition 48 (*Jordan Normal Form* (JNF))**.** *Let $A \in \mathbb{R}^{d \times d}$ be a matrix, then there exists an invertible matrix P and block diagonal matrix $J = \operatorname{diag}(J_1, \dots, J_k)$ such that $A = PJP^{-1}$ where J_i is a* Jordan *block of the form*

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_i & 1 \\ & & & \lambda_i \end{bmatrix}$$

where $\lambda_1, \dots, \lambda_k$ are the (possibly repeated) eigenvalues of A.

[†]Algebraic numbers are roots of polynomials with integer (or rational coefficients). For example $x = \sqrt{2}$ is algebraic because $x^2 - 2 = 0$.

To summarize, we started with Markov chains that we generalized to linear dynamical systems. We then showed that the following objects are equivalent:

- linear dynamical systems,
- linear recurrent sequence,
- *exponential polynomials*: expressions of the form (5).

This equivalence is important because it shows that LRS are a universal object in some sense, they appear naturally in many contexts and it gives more tools to solve problems.

2.2 Decisions problems

Recall that in Section 1.2, we looked at the emptiness problem for probabilistic automata and showed that it is undecidable. It is clear that the proof does not apply anymore because we used binary expansion to encode words, something which is impossible with a unary alphabet. In fact, the problem becomes *a priori* much simpler. Indeed, fix $\lambda \in (0,1)$, then the emptiness problem for Markov chain becomes: decide whether there exists $n \in \mathbb{N}$ such that $SA^nT > \lambda$. Note in particular that this is a "deterministic" problem: there are no words to choose, we *just* need to check if a LRS contains an element bigger than λ. For this purpose, we introduce the following two problems (the names are not universally):

Problem 49 (*Markov Reachability/Equality*)**.** *Given a Markov chain $\langle S, A, T\rangle$ and a threshold $\lambda \in \mathbb{Q}$, decide whether $SA^nT = \lambda$ for some n.*

Problem 50 (*Markov inequality*)**.** *Given a Markov chain $\langle S, A, T\rangle$ and a threshold $\lambda \in \mathbb{Q}$, decide whether $SA^nT \geqslant \lambda$ for all n. version).*

Note that the Markov inequality problem naturally comes in two flavors, depending on whether the inequality is strict or not. It is clear that the Markov inequality problem is equivalent (in terms of decidability) to the emptiness problem since $\exists n. SA^nT > \lambda$ if and only if $\neg\forall n. SA^nT \leqslant \lambda$ if and only if $\neg\forall n.(1 - SA^nT) \geqslant 1 - \lambda$ and $1 - SA^nT$ is also a Markov chain.

While the Markov reachability problem hasn't necessarily received a lot of attention, the following well-known problems for integer LRS have been studied extensively.

Problem 51 (*Skolem*)**.** *Given a LRS $(u_n)_n$, decide whether it has a zero,* i.e. *whether $u_n = 0$ for some $n \in \mathbb{N}$.*

Problem 52 (*Positivity*)**.** *Given a LRS $(u_n)_n$, decide whether it is positive,* i.e. *whether $u_n > 0$ for all $n \in \mathbb{N}$.*

Note that the positivity problem also naturally comes in two flavors, depending on whether the inequality is strict or not. It is clear that the positivity problem is harder than the Skolem problem since we can reduce the latter to the former. On the other hand, the Skolem problem has now been open for *more than 70 years* [OW12] ! In particular, the Skolem problem is not known to be either decidable or undecidable.

Remark 53. The Skolem and positivity are classically defined with a threshold of 0. This is without loss of generally since for any $\lambda \in \mathbb{Q}$, $u_n = \lambda$ if and only if $u_n - \lambda = 0$ and $(u_n - \lambda)_n$ is a LRS.

Exercise 54. Some authors define the Skolem or Markov reachability problem as follows: given a matrix $A \in \mathbb{Q}^{d\times d}$, decide whether $(M^n)_{1,2} = 0$ for some n. Show that the two formulations are equivalent.

It is clear that the Markov reachability and inequality problems are particular cases of the Skolem and positivity problems for rational LRS. Nevertheless, one could hope that the stochastic aspect could make the problem easier. We will show that this is unfortunately not the case. The reduction follows [AAOW15] and uses the following intermediate problem.

Problem A. *Given a stochastic matrix $A \in \mathbb{Q}^{d\times d}$ and a vector $y \in \{0,1,2\}^d$, decide whether $eA^ny = 1$ where $e = \begin{bmatrix}1 & 0 & \dots & 0\end{bmatrix}$.*

Proposition 55. *The Skolem problem for rational LRS reduces to Problem A.*

Proof. Let $A \in \mathbb{Z}^{d\times d}$ be an instance of the Skolem problem. Without loss of generality (see Exercise 54), we are trying to decide whether $(A^n)_{1,2} = 0$ for some n. We will construct a stochastic matrix $\tilde{P}$ and vector $\tilde{v} \in \{0,1,2\}^{2k+1}$ such that for all n, $A^n_{1,2} = 0$ if and only if $e\tilde{P}^n\tilde{v} = 0$ where $e = \begin{bmatrix}1 & 0 & \dots & 0\end{bmatrix}$.

The first step consists in separating the positive and negative values in A, since a stochastic matrix can only have nonnegative entries. Let A^+ and A^- be nonnegative matrices defined by $A^+_{ij} = \max(0, A_{ij})$ and $A^+_{ij} = \max(0, -A_{ij})$, then $A = A^+ - A^-$. Now define

$$e = \begin{bmatrix}1 & 0 & \dots & 0\end{bmatrix}, \qquad P = \begin{bmatrix}A^+ & A^-\\ A^- & A^+\end{bmatrix}, \qquad v = \begin{bmatrix}x\\ -x\end{bmatrix}, \qquad x = \begin{bmatrix}0 & 1 & 0 & \dots & 0\end{bmatrix}^t.$$

One checks that $eP^nv = e(A^+ - A^-)^n x = A^n_{1,2}$ by using that $\varphi : \begin{bmatrix} X & Y \\ Y & X \end{bmatrix} \mapsto X - Y$ is a homomorphism from the ring of $2k \times 2k$ symmetric matrices to $k \times k$ matrices.

The second step is to rescale the matrix to make it stochastic[‡], now that it only has nonnegative entries. Let $s \in \mathbb{Q}$ such that sP is substochastic and define

$$\tilde{e} = \begin{bmatrix} e & 0 \end{bmatrix}, \qquad \tilde{P} = \begin{bmatrix} sP & \mathbf{1} - sP\mathbf{1} \\ 0 & 1 \end{bmatrix}, \qquad \tilde{v} = \begin{bmatrix} \mathbf{1} + v \\ 1 \end{bmatrix}, \quad \text{where} \quad \mathbf{1} = \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^t.$$

First it is clear that $\tilde{e}$ is stochastic since it contains a single 1, and the entries of $\tilde{v}$ are in $\{0,1,2\}$ since the entries of v are in $\{-1,0,1\}$. Moreover, $\tilde{P}$ is stochastic since on row i, the last entry is $1 - (sP\mathbf{1})_i = 1 - \sum_j sP_{ij}$, *i.e.* the remainder to make it stochastic. Then we have that

$$\tilde{e}\tilde{P}^n\tilde{v} = \begin{bmatrix} e & 0 \end{bmatrix} \begin{bmatrix} (sP)^n & \mathbf{1} - (sP)^n\mathbf{1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{1} + v \\ 1 \end{bmatrix} = \begin{bmatrix} e & 0 \end{bmatrix} \begin{bmatrix} (sP)^n v + \mathbf{1} \\ 1 \end{bmatrix} = e(sP)^n v + e\mathbf{1} = s^n A^n_{1,2} + 1.$$

Therefore, $\tilde{e}\tilde{P}^n\tilde{v} = 1$ if and only if $A^n_{1,2} = 0$. □

Proposition 56. *Problem A reduces to the Markov reachability problem with threshold* $\frac{1}{2}$.

Proof. Let $e = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}$, $A \in \mathbb{Q}^{d \times d}$ stochastic and $y \in \{0,1,2\}^k$ be an instance of Problem A. We will build a $(k+5) \times (k+5)$ stochastic matrix P such that $eA^ny = 1$ if and only if $SP^{n+2}T = \frac{1}{2}$ for some S and T, which is an instance of the Markov reachability problem.

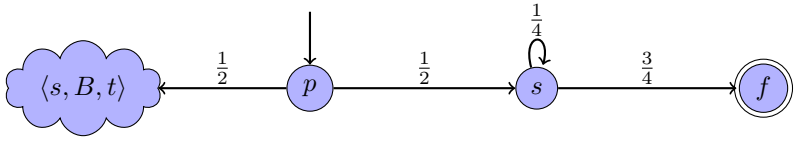
First, we need to put y in the matrix itself since we cannot have a value of 2 in the vector T and ensure that the resulting matrix is stochastic. Define

$$s = \begin{bmatrix} e & 0 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} \frac{1}{4}A & \frac{1}{4}y & \mathbf{1} - \frac{1}{4}(A\mathbf{1} - y) \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \qquad t = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad \mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}.$$

Note that s is stochastic and B is stochastic since A is stochastic and $y_i \leqslant 2$ thus each line has sum 1 and $\mathbf{1} - \frac{1}{4}A\mathbf{1} - \frac{1}{4}y \geqslant 0$. Then check (using that $A\mathbf{1} = \mathbf{1}$ since A is stochastic) that

$$sB^nt = \begin{bmatrix} e & 0 & 0 \end{bmatrix} \begin{bmatrix} (\frac{1}{4}A)^n & \frac{1}{4^n}A^{n-1}y & \mathbf{1} - \frac{1}{4^n}(A^n\mathbf{1} - y) \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = 4^{-n}eA^{n-1}y.$$

Next, we will use the following automaton, which is essentially the same as in the proof of Proposition 27, to "compensate" for the 4^{-n} factor.



Formally, define

$$S = \begin{bmatrix} 0 & \dots & 0 & 1 & 0 & 0 \end{bmatrix}, \qquad P = \begin{bmatrix} B & 0 & 0 & 0 \\ \frac{1}{2}s & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad T = \begin{bmatrix} t \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

It is immediate that S and P are stochastic since s are B are stochastic, and T only has $\{0,1\}$ entries. It can be checked by induction, or by reasoning on the automaton that

$$SP^nT = S \begin{bmatrix} B^n & 0 & 0 & 0 \\ \frac{1}{2}sB^{n-1} & 0 & \frac{1}{2}4^{1-n} & \frac{1}{2}(1 - 4^{1-n}) \\ 0 & 0 & 4^{-n} & 1 - 4^{-n} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} t \\ 0 \\ 0 \\ 1 \end{bmatrix} = S \begin{bmatrix} B^nt \\ \frac{1}{2}(sB^{n-1}t + 1 - 4^{1-n}) \\ 1 - 4^{-n} \\ 1 \end{bmatrix} = \frac{1}{2}(sB^{n-1}t + 1 - 4^{1-n}).$$

It follows that

$$SP^nT = \tfrac{1}{2} \;\Leftrightarrow\; sB^{n-1}t + 1 - 4^{1-n} = 1 \;\Leftrightarrow\; 4^{1-n}eA^{n-2}y + 1 - 4^{1-n} = 1 \;\Leftrightarrow\; eA^{n-2}y = 1.$$

□

[‡]We have already seen this trick for probabilistic automata, but in the case at hand, we have more constraints to satisfy on the vectors.

We can now show the main result of this section.

Theorem 57. *The following problem are interreducible*[§]*:*

- *the Skolem problem for integer LRS,*
- *the Skolem problem for rational LRS,*
- *the Markov reachability problem, even for fixed threshold.*

Proof. The Skolem problem for integer LRS and the Markov reachability problem are particular case of the Skolem problem for rational LRS. By Proposition 55 and Proposition 56, we have that the Skolem problem for rational LRS is reducible to the Markov reachability problem. Finally the Skolem problem for rational LRS is easily reducible to the Skolem problem for integer LRS: let $(u_n)_n$ be a rational LRS, by Proposition 43, write $u_n = SA^nT$ for some rational S, A, T. Then there exists $m \in \mathbb{N}$ such that mS, mA and mT have integer coefficients. But clearly, $v_n = (mS)(mA)^n(mT) = m^{n+2}u_n$ thus the Skolem problem for $(u_n)_n$ is equivalent to the Skolem problem for $(v_n)_n$, but the latter is an integer LRS by Proposition 43. □

Theorem 58. *The following problem are interreducible:*

- *the positivity problem for integer LRS,*
- *the strict positivity problem for integer LRS,*
- *the positivity problem for rational LRS,*
- *the strict positivity problem for rational LRS,*
- *the Markov reachability problem, even for fixed threshold,*
- *the strict Markov reachability problem, even for fixed threshold.*

Proof. It is straightforward to check that the proof of Theorem 57 also shows that all the non-strict problems are interreducible, and that all the strict problem are interreducible. It remains to see that a strict problem is interreducible with a non-strict one. This is the case for the integer LRS.

Let $(u_n)_n$ be an integer LRS, then $u_n \geqslant 0$ if and only if $u_n + 1 > 0$ and $u_n + 1$ is an integer LRS. Thus the non-strict positivity problem reduces to the strict one. Conversely, $u_n > 0$ if and only $u_n \geqslant 1$ thus the strict positivity problems reduces to the non-strict one. □

2.3 Skolem–Mahler–Lech theorem

We will now see one of the most famous results on the Skolem problem, that gives the structure of the set of zeroes of a LRS. We will follow a particularly simple proof that does not require too much number theory [Han86]. A set $A \subseteq \mathbb{N}$ is called:

- *periodic* if there exists r such that $q \in A$ if and only if $q + r \in A$ for all $q \in \mathbb{N}$.
- *ultimately periodic* if there exists q_0 and r such that $q \in A$ if and only if $q + r \in A$ for all $q \geqslant q_0$,
- *quasi-periodic* if it the union of a finite set and a periodic set.

Exercise 59. Show that A is periodic of period r if and only if there exists a finite set $P \subseteq \{0, \dots, r-1\}$ such that $A = \bigcup_{p \in P}(p + r\mathbb{N})$. Show that A is ultimately periodic if and only if then there exists $r \in \mathbb{N}$ and two finite sets F, P such that $Z = F \cup \bigcup_{p \in P}(p + r\mathbb{N})$.

Lemma 60. *Let $(A_i)_{i \in I}$ be a family of quasi-periodic sets with the same period r, then $A = \bigcap_{i \in I} A_i$ is quasi-periodic of period r.*

Let p be a fixed prime number, then for every rational number $q \neq 0$, there exists a unique integer $u \in \mathbb{Z}$ such that $q = p^u \frac{a}{b}$ where $a, b \in \mathbb{Z}$ and p does not divide a or b. We write this number $v_p(q) = i$, and by convention $v_p(0) = \infty$. This is called a *p-adic valuation* and it satisfies the following properties:

- for all $q, q' \in \mathbb{Q}$, $v_p(qq') = v_p(q) + v_p(q')$,
- for all $q, q' \in \mathbb{Q}$, $v_p(q + q') \geqslant \min(v_p(q), v_p(q'))$,

[§]This means there are both reducible to each other. In particular their (non-)decidability is equivalent.

- for all $n \in \mathbb{N}$, $v_p(n!) \geqslant \frac{n}{p-1}$.

Given a polynomial $P(x) = a_0 + a_1 x + \cdots + a_n x^n$ with rational coefficients, we define its valuation to be $v_p^j(P) = \min\{v_p(a_j), \ldots, v_p(a_n)\}$ for $j \leqslant n$, and $v_p^j(P) = \infty$ if $j > n$. It then follows that

- for all $n \in \mathbb{N}$, $v_p(P(n)) \geqslant v_p^0(P)$.

Lemma 61. *Let P be a polynomial with rational coefficients and $n \in \mathbb{Z}$. Let $R(x) = (x-m)P(x)$, then for every i, $v_p^i(P) \geqslant v_p^{i+1}(R)$.*

Proof. □

Proposition 62. *Let $(d_n)_n$ be a sequence of integers and let $b_n = \sum_{i=0}^n \binom{n}{i} p^i d_i$. Then either b_n is identically 0, or $\{n : b_n = 0\}$ is finite.*

Proof. □

Proposition 63. *Let $\langle S, A, T\rangle$ be LDS with integer coefficients and A invertible. If $p > 2$ does not divide $\det A$, then $\{n : SA^nT = 0\}$ is quasi-periodic of period $r < p^{k^2}$.*

Proof. For any $n \in \mathbb{N}$, let $\tilde{n}$ denote the class of n modulo p and extend it to $\tilde{A}$ coefficient-wise. Then $\tilde{A}$ is a matrix with coefficients in the field $\mathbb{F} = \mathbb{Z}/p\mathbb{Z}$, but since p does not divide $\det A$, then $\tilde{A}$ is invertible over $\mathbb{F}$. It follows (by Lagrange's theorem to $\mathrm{GL}_k(\mathbb{F})$) that there exists $r < p^{k^2}$ such that $\tilde{A}^r = I$ and thus $A^r = I + pM$ where M is an integer coefficient matrix.

Let $j \in \{0, \ldots, r-1\}$ and for all $n \in \mathbb{N}$, let $d_n = (SA^j)M^nT$, then

$$u_{j+rn} = SA^{j+rn}T = SA^j(A^r)^nT = SA^j(I+pM)^nT = \sum_{i=0}^n \binom{n}{i} p^i d_i.$$

It follows by Proposition 62 that $\{n : u_{j+rn} = 0\}$ is either finite or everything. Since there are finitely many j, then $\{n : u_n = 0\}$ is quasi-periodic. □

Theorem 64 (Skolem–Mahler–Lech)**.** *Let $(u_n)_n$ be a LRS, then the set $Z = \{n : u_n = 0\}$ is a quasi-periodic.*

Proof. We will show this result in the case of rational LRS only, and admit the general case. By Proposition 43, there exists a LDS $\langle S, A, T\rangle$ such that $u_n = SA^nT$. Since it is rational, there exists $m \in \mathbb{Z}$ such that $\langle mS, mA, mT\rangle$ is an integer LDS and clearly, $SA^nT = 0$ if and only if $(mS)(mA^n)(mT) = m^{n+2}SA^nT = 0$. Thus we can assume that $\langle S, A, T\rangle$ has integer coefficients. Let d be the dimension of A and $V = A^d(\mathbb{R}^d)$, then observe that A is invertible over V. Furthermore,

$$\{n : u_n = 0\} = \{n : SA^nT = 0\} = \{n \leqslant d : SA^nT = 0\} \cup \{n : SA^n(A^dT) = 0\}.$$

The first part is finite and the second part corresponds to the LDS $\langle S, A, A^dT\rangle$. Since A is invertible over V and $AV \subseteq V$, we can find another LDS $\langle S', B, T'\rangle$ such that $SA^nT = S'B^nT'$ and B is invertible (see Exercise 65). Then apply Proposition 63 to $\langle S', B', T'\rangle$ to conclude. □

Exercise 65. Let $\langle S, A, T\rangle$ be a LDS and V a linear subspace. Assume that $T \in V$, $AV \subseteq V$ and A is invertible over V. Show that there exists a LDS $\langle S', B, T'\rangle$ such that $SA^nT = S'B^nT'$ and B is invertible.

References

[AAOW15] S. Akshay, Timos Antonopoulos, Joël Ouaknine, and James Worrell. Reachability problems for markov chains. *Information Processing Letters*, 115(2):155 – 158, 2015.

[FS15] Nathanaël Fijalkow and Michał Skrzypczak. Irregular behaviours for probabilistic automata. In Mikolai Bojanczyk, Slawomir Lasota, and Igor Potapov, editors, *Reachability Problems*, pages 33–36, Cham, 2015. Springer International Publishing.

[GO10] Hugo Gimbert and Youssouf Oualhadj. Probabilistic automata on finite words: Decidable and undecidable problems. In Samson Abramsky, Cyril Gavoille, Claude Kirchner, Friedhelm Meyer auf der Heide, and Paul G. Spirakis, editors, *Automata, Languages and Programming*, pages 527–538, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[Han86] G. Hansel. Une démonstration simple du théorème de skolem-mahler-lech. *Theoretical Computer Science*, 43:91 – 98, 1986.

[OW12] Joël Ouaknine and James Worrell. Decision problems for linear recurrence sequences. In Alain Finkel, Jérôme Leroux, and Igor Potapov, editors, *Reachability Problems*, pages 21–28, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[Rab63] Michael O. Rabin. Probabilistic automata. *Information and Control*, 6(3):230 – 245, 1963.

[Sak19] Jacques Sakarovitch. Five lectures in the theory of weighted automata and transducers, 2018–2019.

Answer to exercises

Exercise 2. $\mu(w_1 \cdots w_n) = \mu(w_1) \cdots \mu(w_n)$.

Exercise 3. We check that $1+0+0=1$ for I. Then each line of $\mu(a)$ and $\mu(b)$, for example $0+\frac{1}{2}+\frac{1}{2}=1$ and $0+\frac{1}{4}+\frac{3}{4}$.

Exercise 4. Let $M \in [0,1]^{P\times Q}$ and $N \in [0,1]^{Q\times R}$ then $(MN)_{p,r} = \sum_{q\in Q} M_{p,q}N_{q,r}$. It follows that on line p we have

$$\sum_{r\in R}(MN)_{p,r} = \sum_{r\in R}\sum_{q\in Q} M_{p,q}N_{q,r} = \sum_{q\in Q} M_{p,q}\sum_{r\in R} N_{q,r} = \sum_{q\in Q} M_{p,q} = 1.$$

Exercise 5. Intuitively, $\mu(w)_{q,q'}$ is the probability that we end up in state q' by reading word w from state q. Formally, $\mu(w)_{q,q'}$ is the sum of the weights (probabilities) of all paths from q to q' that are labelled by w. To prove it, introduce $E = \{(q,a,q',\mu(a)_{q,q'}) : q,q' \in Q, a \in A\}$ the set of edges of the automaton, and use the results on weighted automata (see Lemma 7 and Corollary 8 of [Sak19]). Then $S\mu(w)$ is the probability distribution of the states starting from the initial distribution I. This is indeed a distribution because it is a stochastic vector.

Exercise 6. In the first approach, we simply write $\mathcal{B}$ using a substochastic matrix: $\mathcal{B} = \langle A, Q, S, \mu, T\rangle$ where $A = \{a,b\}$, $Q = \{p,q,r\}$ and

$$S = \begin{bmatrix}1 & 0 & 0\end{bmatrix}, \qquad \mu(a) = \begin{bmatrix}0 & 1 & 0\\ 0 & 0 & \frac{1}{2}\\ 0 & 0 & 0\end{bmatrix}, \qquad \mu(b) = \begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 1 & 0 & 0\end{bmatrix}, \quad \text{and} \quad T = \begin{bmatrix}0\\0\\1\end{bmatrix}.$$

In the second approach, we create a sink state $\perp$ to account for the missing probability: $\mathcal{B}' = \langle A, Q', S', \mu', T'\rangle$ where $Q' = \{p,q,r,\perp\}$ and

$$S' = \begin{bmatrix}1 & 0 & 0 & 0\end{bmatrix}, \qquad \mu'(a) = \begin{bmatrix}0 & 1 & 0 & 0\\ 0 & 0 & \frac{1}{2} & \frac{1}{2}\\ 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 1\end{bmatrix}, \qquad \mu'(b) = \begin{bmatrix}0 & 0 & 0 & 1\\ 0 & 0 & 0 & 1\\ 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}, \quad \text{and} \quad T' = \begin{bmatrix}0\\0\\1\\0\end{bmatrix}.$$

Notice that indeed $\mu'(a)$ and $\mu'(b)$ are stochastic. Furthermore, we have the following relationship between the two automata, for every letter $x \in A$, vector $v \in \mathbb{Q}^3$ and "sink probability" $\varepsilon \in \mathbb{Q}$:

$$\mu'(x)\begin{bmatrix}v\\ \varepsilon\end{bmatrix} = \begin{bmatrix}\mu(x)v\\ \varepsilon'\end{bmatrix}$$

for some ε'. Thus for every word w,

$$S'\mu'(w)T = \begin{bmatrix}I & 0\end{bmatrix}\mu'(w)\begin{bmatrix}T\\0\end{bmatrix} = \begin{bmatrix}I & 0\end{bmatrix}\begin{bmatrix}\mu(w)T\\ \varepsilon\end{bmatrix} = S\mu(w)T.$$

Exercise 8. Check that $\mathbb{P}_{\mathcal{A}}(bba) = \frac{1}{12}$ and $\mathbb{P}_{\mathcal{A}}(abb) = \frac{2}{3}$. Thus $bba \notin \mathcal{L}_{\mathcal{A}}(\frac{1}{2})$ and $abb \in \mathcal{L}_{\mathcal{A}}(\lambda)$ for all $\lambda < \frac{2}{3}$. In fact $\mathcal{L}_{\mathcal{A}}(\frac{2}{3}) = \varnothing$ as we will see. Check that $\mathbb{P}_{\mathcal{A}}((ab)^n b) = \frac{2}{3}$ for every $n \in \mathbb{N}$, thus $(ab)^*b \subseteq \mathcal{L}_{\mathcal{A}}(\lambda)$ for all $\lambda < \frac{2}{3}$.

Exercise 9. The edges of $\mathcal{B}$ are exactly the edges of $\mathcal{A}$ labelled by a, thus $\mathcal{L}_{\mathcal{B}}(\lambda) = \mathcal{L}_{\mathcal{A}}(\lambda) \cap a^*$.

Exercise 10. TODO

Exercise 11. Each regular language can be described by a regular expression, that is a finite word over the finite alphabet $A \cup \{(,),+,*,\varepsilon\}$. The set of words over a finite alphabet is countable.

Exercise 14. One immediately checks that $\equiv_L$ is reflexive, symmetric and transitive. Let L be a regular language and let $\mathcal{A} = \langle A, Q, q_0, \delta, q_f\rangle$ be a deterministic finite automaton, where q_0, q_f are the initial and final states and $\delta : Q \times A \to Q$ is the transition function (which we can assume is total), which we naturally extend to words in the obvious way. For each state q, define $L_q = \{w \in A^* : \delta(q_0, w) = q\}$ to be the set of words w such that the automaton is in state q after reading w from q_0. We claim that for all $u, v \in L_q$, $u \equiv_L v$. Indeed, if $u \in L_q$ and $w \in A^*$, then $\delta(q_0, uw) = \delta(\delta(q_0,u),w) = \delta(q,w)$ thus $uw \in L$ if and only if $\delta(q,w) = q_f$. Note that this condition is independent of $u \in L_q$ and thus $uw \in L$ if and only if $vw \in L$.

Conversely, assume that the number of equivalence classes is finite. We denote by $[u]$ the equivalence class of every $u \in W$. Now consider the deterministic finite automaton $\mathcal{A} = \langle A, Q, q_0, \delta, F\rangle$ where $Q = \{[u] : u \in A^*\}$ which is finite by asssumption, $q_0 = [\varepsilon]$, $F = \{[u] : u \in L\}$ and $\delta([u], a) = [ua]$. Note that F is well-defined because the condition $u \in L$ is independent of the particular u we choose since if $[u] = [v]$ then $u = u\varepsilon \in L$ if and only if $v = v\varepsilon \in L$. Similarly, δ is well-defined because if $[u] = [v]$ then $[ua] = [va]$. Indeed, $(ua)w \in L$ if and only if $u(aw) \in L$ if and only if $v(aw) \in L$ (by $[u] = [v]$) if and only if $(va)w \in L$. We now prove that $\mathcal{A}$ recognizes L: $\mathcal{A}$ recognizes u if and only if $\delta(q_0, u) \in F$ if and only if $\delta([\varepsilon], u) \in F$ if and only if $[u] \in F$ if and only if $u \in L$.

Exercise 16. This probabilistic automaton is represented by the tuple $\mathcal{C} = \langle A, Q, S, \mu, T\rangle$ where $A = \{a, b\}$, $Q = \{p, q\}$ and

$$S = \begin{bmatrix}1 & 0\end{bmatrix}, \qquad \mu(a) = \begin{bmatrix}\frac{1}{2} & \frac{1}{2}\\ 0 & 1\end{bmatrix}, \qquad \mu(b) = \begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix} \quad \text{and} \quad T = \begin{bmatrix}1\\ 0\end{bmatrix}.$$

One checks that

$$\mu(a)^n \begin{bmatrix}x\\ y\end{bmatrix} = \begin{bmatrix}2^{-n}x + (1-2^{-n})y\\ y\end{bmatrix}, \qquad \mu(b)\begin{bmatrix}x\\ y\end{bmatrix} = \begin{bmatrix}0\\ x\end{bmatrix}, \qquad \mu(a)^n\mu(b)\begin{bmatrix}x\\ y\end{bmatrix} = \begin{bmatrix}(1-2^{-n})x\\ x\end{bmatrix}.$$

Therefore,

$$\begin{aligned} S\mu(x(n_1,\dots,n_k))T &= S\mu(a)^{n_1}\mu(b)\cdots\mu(a)^{n_k}\mu(b)T\\ &= S\mu(a)^{n_1}\mu(b)\cdots\mu(a)^{n_{k-1}}\mu(b)\begin{bmatrix}(1-2^{-n_k})\\ 1\end{bmatrix}\\ &= S\begin{bmatrix}(1-2^{-n_1})\cdots(1-2^{-n_k})\\ (1-2^{-n_2})\cdots(1-2^{-n_k})\end{bmatrix}\\ &= \prod_{i=1}^{k}(1-2^{-n_i}). \end{aligned}$$

Let $u = x(n_1,\dots,n_k)$ and $w = x(n_{k+1},\dots,n_\ell)$ then $uw = x(n_1,\dots,n_\ell)$ thus $\mathbb{P}_\mathcal{C}(uw) = \mathbb{P}_\mathcal{C}(u)\mathbb{P}_\mathcal{C}(w)$ by a straightforward calculation. To see the density, fix $\lambda \in (0,1)$ and let $\mu_\infty = \log\lambda < 0$. Now consider the sequence defined by $\mu_0 = 0$ and $\mu_{i+1} = \mu_i + \log(1-2^{-n_i})$ where $n_{i+1} = \min\{k \geqslant 1 : \nu_i + \log(1-2^{-k}) > \mu_\infty\}$. Such a k exists because $\mu_i > \nu_\infty$ and $\log(1-2^{-k}) \to 0$ as $k \to \infty$. Then $\mu_k \to \mu_\infty$ as $k \to \infty$ thus $e^{\mu_k} \to \lambda$ as $k \to \infty$. But $e^{\mu_k} = \prod_{i=1}^{k}(1-2^{-n_i}) = \mathbb{P}_\mathcal{C}(x(n_1,\dots,n_k))$. The proof that $\mathcal{C}$ is universally non-regular is then the same as for Theorem 15.

Exercise 19. Since L is nonempty, there exists $x \in L$, which must therefore have length $|x| \geqslant n$. For every $1 \leqslant i \leqslant n$, let $u_i = x_1\cdots x_i$. Then $u_i \not\equiv_L u_j$ for $i < j$. Indeed, if we let $w = x_{j+1}\cdots x_n$ then $u_j w = x \in L$ but $|u_i w| = i + n - j < n$ thus $u_i w \notin L$. It follows that $\equiv_L$ has at least n equivalence classes. By Theorem 13, any deterministic finite automaton that recognizes L must therefore have at least n states.

Exercise 28. Let $\mathcal{A} = \langle A, Q, S, \mu, T\rangle$ and $\mathcal{B} = \langle A, Q', S', \mu', T'\rangle$. Define $\mathcal{C} = \langle A, Q\times Q', S\otimes S', \mu\otimes\mu', T\otimes T'\rangle$ where $\otimes$ is the *Kronecker* product: given $M \in \mathbb{R}^{Q\times Q}$ and $M' \in \mathbb{R}^{Q'\times Q'}$ then $M\otimes M' \in \mathbb{R}^{(Q\times Q')\times(Q\times Q')}$ is defined by

$$(M\otimes M')_{(p,p'),(q,q')} = M_{p,q}M'_{p',q'}.$$

We check that $\otimes$ preserves stochasticity:

$$\sum_{q\in Q}\sum_{q'\in Q'}(M\otimes M')_{(p,p'),(q,q')} = \sum_{q\in Q}M_{p,q}\sum_{q'\in Q'}M'_{p',q'} = \sum_{q\in Q}M'_{p,q} = 1$$

if M and M' are stochastic. And furthermore, it satisfies the *mixed-product* property:

$$(A\otimes B)(C\otimes D) = (AB)\otimes(CD).$$

Indeed,

$$\begin{aligned}(A\otimes B)(C\otimes D)_{(p,p'),(q,q')} &= \sum_{r\in Q}\sum_{r'\in Q'}(A\otimes B)_{(p,p'),(r,r')}(C\otimes D)_{(r,r'),(q,q')}\\ &= \sum_{r\in Q}\sum_{r'\in Q'}A_{p,r}B_{p',r'}C_{r,q}D_{r',q'}\\ &= \sum_{r\in Q}A_{p,r}C_{r,q}\sum_{r'\in Q'}B_{p',r'}D_{r',q'}\\ &= (AC)_{p,q}(BD)_{p'q'}\\ &= ((AC)\otimes(BD))_{(p,p'),(q,q')}.\end{aligned}$$

Therefore for every word $w \in A^*$ we have that

$$\begin{aligned}\mathbb{P}_\mathcal{C}(w) &= (S\otimes S')\mu''(a)(T\otimes T')\\ &= (S\otimes S')(\mu(a_1)\otimes\mu'(a_1))\cdots(\mu(a_{|a|})\otimes\mu'(a_{|a|}))(T\otimes T')\end{aligned}$$

$$= (S\mu(a_1)\cdots\mu(a_n)T) \otimes (S\mu(a_1)\cdots\mu(a_n)T) \qquad \text{by the mixed-product property}$$
$$= \mathbb{P}_{\mathcal{A}}(w)\mathbb{P}_{\mathcal{B}}(w).$$

If $\mathcal{A}$ and $\mathcal{B}$ are simple then it is clear that all probabilities that appear in $\mathcal{C}$ are product of the form xy where x and y are multiple of $\frac{1}{2}$, therefore they are multiple of $\frac{1}{4}$

Exercise 25. Let $\mathcal{A} = \langle A, Q, S, \mu, T\rangle$ be a probabilistic automaton and define $\mathcal{A}^c = \langle A, Q, S, \mu, S\rangle$ where $S_i = 1 - T_i$, *i.e.* S is the "complement" over T. Then for every word $w \in A^*$, we have that

$$\mathbb{P}_{\mathcal{A}^c}(w) = S\mu(w)S = \sum_{i=1}^{|Q|}(S\mu(w))_i S_i = \sum_{i=1}^{|Q|}(S\mu(w))_i(1-T_i) = \sum_{i=1}^{|Q|}(S\mu(w))_i - \sum_{i=1}^{|Q|}(S\mu(w))_i T_i = 1 - \mathbb{P}_{\mathcal{A}}(w)$$

by stochacity of $S\mu(w)$.

Exercise 26. Let $\mathcal{A} = \langle A, Q, S, \mu, T\rangle$ be a probabilistic automaton and let p be the smallest integer such that for all $a \in A$, $2^p\mu(a)$ has integer entries. In other words, p is the highest power of 2 appearing in the denominators of the transition probabilities. If $p = 0$ or $p = 1$ then $\mathcal{A}$ is simple. Otherwise, define $\mathcal{A}' = \langle A, Q', S', \mu', T'\rangle$ where $Q = Q \cup \{\hat{q} : q \in Q\}$, $S'_q = S_q$ and $S_{\hat{q}} = 0$, $T'_q = T_q$ and $T'_{\hat{q}} = 0$ for all $q \in Q$, and for all letters $a \in A$:

Exercise 30. TODO

Exercise 33. TODO

Exercise 38. It is clear that any such path from 1 to 3 must go through the cycle $1 \to 2 \to 1$ at most $k-1$ times, then follow $1 \to 3$ once and then $3 \to 3$ for the remaining time. Thus

$$\begin{aligned}
\mathbb{P}_{\mathcal{A}_x}\left(1 \xrightarrow{w} 3\right) &= \sum_{i=0}^{k-1} \mathbb{P}_{\mathcal{A}_x}\left(1 \xrightarrow{a^{n_1}b\cdots a^{n_i}b} 1\right)\mathbb{P}_{\mathcal{A}_x}\left(1 \xrightarrow{a^{n_{i+1}}b} 3\right)\mathbb{P}_{\mathcal{A}_x}\left(3 \xrightarrow{a^{n_{i+2}}b\cdots a^{n_k}b} 3\right) \\
&= \sum_{i=0}^{k-1}\prod_{j=1}^{i}(1-x^{n_j})\mathbb{P}_{\mathcal{A}_x}\left(1 \xrightarrow{a^{n_{i+1}}} 3\right) = \sum_{i=0}^{k-1} x^{n_{i+1}}\prod_{j=1}^{i}(1-x^{n_j}) \\
&= 1 - \prod_{j=1}^{k}(1-x^{n_j}) \qquad \text{by induction on } k.
\end{aligned}$$

and similarly

$$\begin{aligned}
\mathbb{P}_{\mathcal{A}_x}\left(4 \xrightarrow{w} 6\right) &= \sum_{i=0}^{k-1} \mathbb{P}_{\mathcal{A}_x}\left(4 \xrightarrow{a^{n_1}b\cdots a^{n_i}b} 4\right)\mathbb{P}_{\mathcal{A}_x}\left(4 \xrightarrow{a^{n_{i+1}}b} 6\right)\mathbb{P}_{\mathcal{A}_x}\left(6 \xrightarrow{a^{n_{i+2}}b\cdots a^{n_k}b} 6\right) \\
&= \sum_{i=0}^{k-1}\prod_{j=1}^{i}(1-(1-x)^{n_j})\mathbb{P}_{\mathcal{A}_x}\left(4 \xrightarrow{a^{n_{i+1}}} 6\right) = \sum_{i=0}^{k-1}(1-(1-x)^{n_j})\prod_{j=1}^{i}(1-x)^{n_j} \\
&= 1 - \prod_{j=1}^{k}(1-(1-x)^{n_j}) \qquad \text{by induction on } k \\
&\leqslant \sum_{i=1}^{k}(1-x)^{n_i} \qquad \text{by induction on } k.
\end{aligned}$$

Exercise 45. If we follow the proof of the course then we need find $a, b \in \mathbb{Q}$ such that

$$A^2 = aA^1 + bA^0 \quad \Leftrightarrow \quad \begin{bmatrix}-1 & 2\\ -2 & 3\end{bmatrix} = \begin{bmatrix}b & a\\ -a & 2a+b\end{bmatrix} \quad \Leftrightarrow \quad a = 2 \wedge b = -1.$$

Therefore we get that $u_{n+2} = 2u_{n+1} - u_n$, $u_0 = SA^0T = 0$ and $u_1 = SA^1T = 1$. It is not hard to see that $u_n = n$. It is also immediate that

$$A^n \begin{bmatrix}u_n\\ u_{n+1}\end{bmatrix} = \begin{bmatrix}u_{n+1}\\ 2u_{n+1} - u_n\end{bmatrix} = \begin{bmatrix}u_{n+1}\\ u_{n+2}\end{bmatrix}.$$

One easily checks by induction that for every $n \in \mathbb{N}$,

$$B^n = \begin{bmatrix}1 & n\\ 0 & 1\end{bmatrix}$$

and thus $SB^nT = n = u_n$, but

$$B^n \begin{bmatrix}u_n\\ u_{n+1}\end{bmatrix} = \begin{bmatrix}u_n + u_{n+1}\\ u_{n+1}\end{bmatrix} \neq \begin{bmatrix}u_{n+1}\\ u_{n+2}\end{bmatrix}.$$