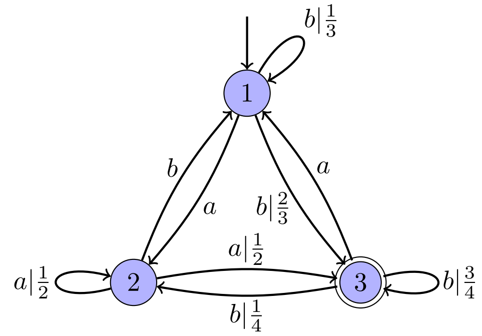
Probabilistic automata and Markov chains

Amaury Pouly
Université de Paris, CNRS, IRIF

Version 0.9993



Lectures notes of the Master Parisien de Recherche en Informatique

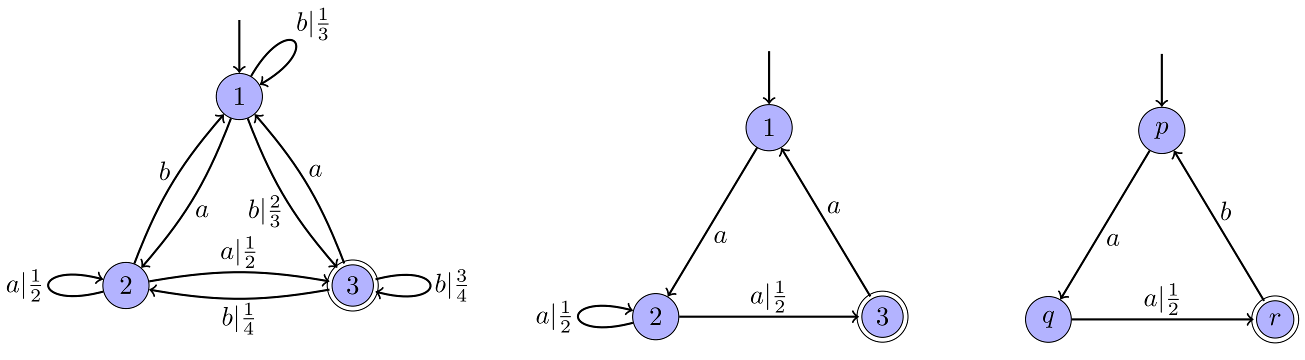
Course 2.16 – Finite automata based computation models

Academic year 2021 – 2022

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 former part of the course on weighted automata and transducers [Sak18].

[1] Version 0.9993—ad3070e, compiled on November 10, 2022



(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: $\mu(w_1 \cdots w_n) = \mu(w_1)\cdots\mu(w_n)$. To every word $w \in A^*$, we can now associate the *probability of acceptance* $\mathcal{A}(w) = S\mu(w)T$.

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 $\mathcal{A}\left(q \xrightarrow{w} q'\right) = \mu(w)_{q,q'}$.

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 is 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. 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 3. Show that the product of two stochastic matrices is stochastic.

Exercise 4. 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 $S\mu(w)$?

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 5. 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

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

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. More generally, given $\bowtie \in \{\geqslant, >, =, \neq, <, \leqslant\}$ a comparison operator, we can consider

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

Example 6 (Automaton of Figure 1a)**.** We have seen in Example 1 that $\mathcal{A}(bb) = \frac{13}{18}$ 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 7 (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 8 (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 9. Prove that every regular language is stochastic. *Hint: take a finite automaton 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 10. Let A be a finite alphabet, prove that the collection of regular languages over A^* is countable.

When comparing regular languages to other classes, the following characterization will be very useful. It provides not only a criterion to decide whether a language is regular, but also to estimate the number of states of a minimal automaton.

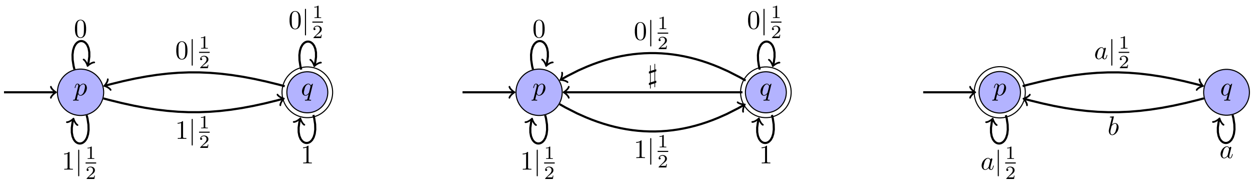
Theorem 11 (*Myhill-Nerode*)**.** *Let* L *be a language, we say that two words* u *and* v *are* right equivalent *for* L*, and write* $u \equiv_L v$*, if for every* $w \in A^*$*, we have* $uw \in L$ *if and only if* $vw \in L$*. Prove that* $\equiv_L$ *is an equivalence relation. Show that a language* 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 12. Prove Theorem 11. 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.

1.1.1 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 13. *Stochastic languages strictly contains regular languages.*



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

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

Proof. Every regular language is stochastic, see Exercise 9. 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 $\mathcal{A}(w) = [w]$. 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 $\lambda < \mathcal{A}(w) \leqslant \mu$ 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. □

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 (in fact for almost all λ). Surprisingly, a small modification of this automaton given by [FS15] allows us to strengthen this statement.

Theorem 14. *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 13 with an extra transition from q to p labelled by $\sharp$. Note that when reading a $\sharp$, the automaton must be in state q for the word to be accepted with positive probability. Therefore for all $u, v \in A^*$, we have that

$$\mathcal{B}(u\sharp v) = \mathcal{B}\left(p \xrightarrow{u} q\right) \mathcal{B}\left(p \xrightarrow{v} q\right) = \mathcal{A}(u)\mathcal{A}(v) = [u][v].$$

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 $\mathcal{B}(u\sharp w) = [u][w] < \lambda$ whereas $\mathcal{B}(v\sharp w) = [v][w] > \lambda$. 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 11. □

Exercise 15. 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 $x(n_1, \ldots, n_k) := a^{n_1} b a^{n_2} \cdots a^{n_k} b$ we have $\mathcal{C}(x(n_1, \ldots, n_k; m)) = 2^{-m} \prod_{i=1}^{k} (1 - 2^{-n_i})$. Show that if $u = x(n_1, \ldots, n_k)$ and $w = x(n_{k+1}, \ldots, n_\ell)$ then $\mathcal{C}(uw) = \mathcal{C}(u)\mathcal{C}(w)$. Show that $\{\mathcal{C}(x(n_1, \ldots, n_k)) : n_1, \ldots, n_k \in \mathbb{N}, k \in \mathbb{N}\}$ is dense in $[0, 1]$. Conclude that $\mathcal{C}$ is universally non-regular. *Hint:* use the same proof idea as Theorem 14.

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

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

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

Theorem 16. *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. For simplicity, we assume that there is a unique initial state and a unique final state (which are distinct). Write $\mathcal{A} = \langle A, Q, S, \mu, T\rangle$ where $Q = \{s_1, \dots, s_n\}$, s_1 is the unique initial state 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 $\equiv_L$-unequal words (*i.e.* $x_i \not\equiv_L x_j$ for $i \neq j$, see Theorem 11). Then by definition, for every $i \neq j$, there exist $y \in A^*$ such that $x_iy \in L$ but $x_jy \notin L$ (or the other way around). Since λ is isolated we must have that

$$\mathcal{A}(x_iy) - \mathcal{A}(x_jy) \geqslant 2\delta.$$

Let $(\xi^i_1, \dots, \xi^i_n)$ 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 $\mathcal{A}(x_iy) = S\mu(x_iy)T = S\mu(x_i)\mu(y)T = \xi^i_1\eta_1 + \cdots \xi^i_n\eta_n$ and thus

$$\mathcal{A}(x_iy) - \mathcal{A}(x_jy) = (\xi^i_1 - \xi^j_1)\eta_1 + \cdots (\xi^i_n - \xi^j_n)\eta_n \geqslant 2\delta.$$

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

$$|\xi^i_1 - \xi^j_1| + \cdots |\xi^i_n - \xi^j_n| \geqslant 2\delta, \qquad \text{for } i \neq j. \tag{1}$$

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 $x \in B_\delta(\xi^i)$ 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 11. 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 $c' = \frac{\sqrt{n}}{(n-1)!}$).

We leave the general case of a several initial/final state to the reader, since the proof is similar. □

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 is lower bounded by a function of n. The next theorem shows that, in fact, it is not.

Theorem 17. *There exists a probabilistic automaton $\mathcal{A}$ with only two states and a sequence $(\lambda_m)_{m\in\mathbb{N}}$ of isolated cut points such that $\mathcal{L}_\mathcal{A}(\lambda_m)$ cannot be recognized by a deterministic finite automaton with less than m 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

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

Contrary to the previous automaton, the set $P = \{\mathcal{A}(w) : w \in A^*\}$ 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 in their ternary (base 3) expansion. In fact, it can be seen that C is the topological closure of P, therefore if $\lambda \notin C$ then λ must be an isolated cut-point of $\mathcal{A}$. Now fix $m \in \mathbb{N}$ and consider the cut point

$$\lambda_m = 0.2222\cdots 2211 = \sum_{i=1}^{m-1} 2 \cdot 3^{-i} + 3^{-m} + 3^{-m-1}.$$

Note that $\lambda_m \notin C$ because its ternary expansions[1] contain the digit 1. Then the word $2^m \in \mathcal{L}_{\mathcal{A}}(\lambda_m)$ since it has probability of acceptance

$$\mathcal{A}(2^m) = \sum_{i=1}^{n} 2 \cdot 3^{i-k-1} > \lambda_m.$$

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

$$\mathcal{A}(w) \leqslant \sum_{i=1}^{|w|} 2 \cdot 3^{-i-|w|-1} \leqslant \sum_{i=1}^{m-1} 2 \cdot 3^{-i} \leqslant \lambda_m.$$

It follows that $\mathcal{A}(\lambda_m)$ is nonempty and must reject all words of length less than m. Therefore any deterministic finite automaton that recognizes this language must have at least m states (see Exercise 18). □

Exercise 18. Let L be a nonempty regular language that contains no words of length less than m. Show that any deterministic finite automaton that recognizes L must have at least m states. *Hint: use Theorem 11.*

Theorem 17 shows that the number of states of a deterministic automaton must be a function of the isolation threshold, even when the number of states of the probabilistic automaton is fixed. Conversely, it is natural to ask about the dependence on the number n of states of the probabilistic automaton when the isolation threshold is fixed. The following result shows an almost matching depencen with Theorem 16.

Theorem 19. *There exists $\delta > 0$ such that for infinitely many n, there exists a regular language recognized by a probabilistic automaton with n states and an isolated cut-point with isolation threshold at least δ, such that the smallest deterministic finite automaton recognizing it has $\Omega(2^{\frac{n \ln\ln n}{\ln n}})$ states.*

Proof. See Exercise 92. □

1.2 Operations on probabilistic automata

Probabilistic automata naturally define functions from Σ^* to $[0,1]$. This gives us more structure to work with than classical automata and begs the question of which operations can be done effectively. Natural operations include:

- convex combinations: $\alpha\mathcal{A}(w) + (1-\alpha)\mathcal{B}(w)$;
- complement: $1 - \mathcal{A}(w)$;
- product: $\mathcal{A}(w)\mathcal{B}(w)$;
- changing the probability of the empty word.

We will now see that all these operations are effective.

Lemma 20. *For any two probabilistic automata $\mathcal{A}$ and $\mathcal{B}$ over the same alphabet, and $\alpha \in [0,1]$, there exists an automaton $\alpha\mathcal{A} + (1-\alpha)\mathcal{B}$ that satisfies $(\alpha\mathcal{A} + (1-\alpha)\mathcal{B})(w) = \alpha\mathcal{A}(w) + (1-\alpha)\mathcal{B}(w)$ for every word w.*

[1] Beware that a number can have two ternary expansions, for example $\frac{1}{3} = 0.1 = 0.0222\cdots$ in base 3. In this case, $\lambda_m = 0.2222\cdots 2211 = 0.2222\cdots 2210222\cdots$ that both use the digit 1.

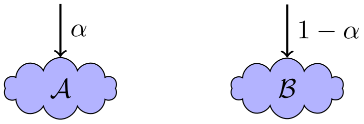
Proof. Let $\mathcal{A} = \langle A, Q, S, \mu, T\rangle$ and $\mathcal{B} = \langle A, Q', S', \mu', T'\rangle$. Define $\mathcal{C} = \langle A, Q \cup Q', S'', \mu'', T''\rangle$ where

$$S'' = \begin{bmatrix}\alpha S & (1-\alpha)S'\end{bmatrix}, \qquad \mu''(a) = \begin{bmatrix}\mu(a) & 0 \\ 0 & \mu'(a)\end{bmatrix}, \qquad T'' = \begin{bmatrix}T \\ T'\end{bmatrix}.$$

One can then check that

$$\mathcal{C}(w) = \begin{bmatrix}\alpha S & (1-\alpha)S'\end{bmatrix} \begin{bmatrix}\mu(w) & 0 \\ 0 & \mu'(w')\end{bmatrix} \begin{bmatrix}T \\ T'\end{bmatrix} = \alpha S\mu(w)T + (1-\alpha)S'\mu'(w)T'.$$

Graphically, this construction corresponds to the following:



□

Lemma 21. *For any probabilistic automaton $\mathcal{A}$, there exists $\mathcal{A}^c$ such that $\mathcal{A}^c(w) = 1 - \mathcal{A}(w)$ for every word w.*

Proof. Let $\mathcal{A} = \langle A, Q, S, \mu, T\rangle$ be a probabilistic automaton and define $\mathcal{A}^c = \langle A, Q, S, \mu, T'\rangle$ where $T'_i = 1 - T_i$, *i.e.* S is the "complement" over T. Then for every word $w \in A^*$, we have that

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

by stochasticity of $S\mu(w)$. □

Lemma 22. *For any two probabilistic automata $\mathcal{A}$ and $\mathcal{B}$ over the same alphabet, there exists a product automaton $\mathcal{A} \cdot \mathcal{B}$ that satisfies $\mathcal{A} \cdot \mathcal{B}(w) = \mathcal{A}(w)\mathcal{B}(w)$ for every word w.*

Proof. 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 $M \in \mathbb{R}^{I \times J}$ and $M' \in \mathbb{R}^{I' \times J'}$ then $M \otimes M' \in \mathbb{R}^{(I \times I') \times (J \times J')}$ is defined by

$$(M \otimes M')_{(i,i'),(j,j')} = M_{i,j} M'_{i',j'}.$$

We check that $\otimes$ preserves stochasticity: for every $(i, i') \in I \times I'$,

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

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

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD).$$

Indeed, if $A \in \mathbb{R}^{I \times K}, B \in \mathbb{R}^{I' \times K'}, C \in \mathbb{R}^{K \times J}, D \in \mathbb{R}^{K' \times J'}$,

$$\begin{aligned}
(A \otimes B)(C \otimes D)_{(i,i'),(j,j')} &= \sum_{k \in K} \sum_{k' \in K'} (A \otimes B)_{(i,i'),(k,k')} (C \otimes D)_{(k,k'),(j,j')} \\
&= \sum_{k \in K} \sum_{k' \in K'} A_{i,k} B_{i',k'} C_{k,j} D_{k',j'} \\
&= \sum_{k \in K} A_{i,k} C_{k,j} \sum_{k' \in K'} B_{i',k'} D_{k',j'} \\
&= (AC)_{i,j} (BD)_{i',j'} \\
&= ((AC) \otimes (BD))_{(i,i'),(j,j')}.
\end{aligned}$$

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

$$\begin{aligned}
\mathcal{C}(w) &= (S \otimes S')\mu''(a)(T \otimes T') \\
&= (S \otimes S')(\mu(a_1) \otimes \mu'(a_1)) \cdots (\mu(a_{|a|}) \otimes \mu'(a_{|a|}))(T \otimes T') \\
&= (S\mu(a_1) \cdots \mu(a_n)T) \otimes (S\mu(a_1) \cdots \mu(a_n)T) && \text{by the mixed-product property} \\
&= \mathcal{A}(w)\mathcal{B}(w).
\end{aligned}$$

□

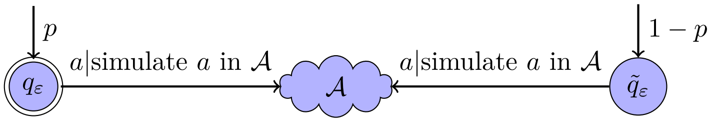
In a number of proofs, we will want to specifically change the probability of the empty word after a construction, typically to change it to zero so that it is rejected.

Lemma 23. *For any probabilistic automaton $\mathcal{A}$ and probability p, there exists an automaton $\mathcal{A}[\varepsilon \leftarrow p]$ that satisfies $\mathcal{A}[\varepsilon \leftarrow p](\varepsilon) = p$ and $\mathcal{A}[\varepsilon \leftarrow p](w) = \mathcal{A}(w)$ for any non-empty word w.*

Proof. Let $\mathcal{A} = \langle A, Q, S, \mu, T\rangle$ and let $q_\varepsilon, \tilde{q}_\varepsilon \notin Q$ be a fresh states. Define $\mathcal{B} = \langle A, \{q_\varepsilon, \tilde{q}_\varepsilon\} \cup Q, S', \mu', T'\rangle$ where

$$S' = \begin{bmatrix} p & 1-p & \mathbf{0} \end{bmatrix}, \qquad \mu'(a) = \begin{bmatrix} 0 & 0 & S\mu(a) \\ 0 & 0 & S\mu(a) \\ \mathbf{0} & \mathbf{0} & \mu(a) \end{bmatrix}, \qquad T' = \begin{bmatrix} 1 \\ 0 \\ T \end{bmatrix}.$$

Graphically, this construction corresponds to the following:


p
a|simulate a in $\mathcal{A}$
q_ε
$\mathcal{A}$
a|simulate a in $\mathcal{A}$
$1-p$
$\tilde{q}_\varepsilon$

It is clear that $\mu'(a)$ is stochastic and furthermore we have that

$$\mathcal{B}(\varepsilon) = S'T' = \begin{bmatrix} p & 1-p & \mathbf{0} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ T \end{bmatrix} = p.$$

Furthermore, check by induction that for every non-empty word w, we have

$$\mu'(w) = \begin{bmatrix} 0 & 0 & S\mu(w) \\ 0 & 0 & S\mu(w) \\ \mathbf{0} & \mathbf{0} & \mu(aw) \end{bmatrix}$$

and hence

$$\mathcal{B}(w) = S'\mu(w)T' = \begin{bmatrix} p & 1-p & \mathbf{0} \end{bmatrix} \begin{bmatrix} 0 & 0 & S\mu(w) \\ 0 & 0 & S\mu(w) \\ \mathbf{0} & \mathbf{0} & \mu(aw) \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ T \end{bmatrix} = pS\mu(w)T + (1-p)S\mu(w)T = \mathcal{A}(w).$$

□

1.3 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 24 (*Strict Emptiness*)**.** *Given a probabilistic automaton $\mathcal{A}$ and a cut-point λ, decide whether there exists a word w such that $\mathcal{A}(w) > \lambda$.*

Problem 25 (*Emptiness*)**.** *Given a probabilistic automaton $\mathcal{A}$ and a cut-point λ, decide whether there exists a word w such that $\mathcal{A}(w) \geqslant \lambda$.*

Problem 26 (*Universality*)**.** *Given a probabilistic automaton $\mathcal{A}$ and a cut-point λ, decide whether it is true that $\mathcal{A}(w) \geqslant \lambda$ for all word w.*

In fact, all three 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 27 (*Equality*)**.** *Given a probabilistic automaton $\mathcal{A}$ and a cut-point λ, decide whether there exists a word w such that $\mathcal{A}(w) = \lambda$.*

It is clear that those problems are equivalent to asking whether $\mathcal{L}_\mathcal{A}^{>}(\lambda)$, $\mathcal{L}_\mathcal{A}^{\geqslant}(\lambda)$ and $\mathcal{L}_\mathcal{A}^{=}(\lambda)$ are empty. An important fact about stochastic language is that the cut-point is usually irrelevant because it can easily be changed, as follow lemma (left as an exercise) shows.

Exercise 28. Given $\mathcal{A}$ a probabilistic automaton and two rational cut-points $\lambda, \mu \in (0,1)$, show that there exists an automaton $\mathcal{B}$ such that $\mathcal{A}(w) \geqslant \lambda$ if and only if $\mathcal{B}(w) \geqslant \mu$.

We will start with the equality problem and reduction from the *Post Correspondence Problem* (PCP) given by Bertoni [BMT77]. Recall that PCP is a classical example of undecidable problem.

Problem 29 (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 $w \in A^* \setminus \{\varepsilon\}$ such that $\phi_1(w) = \phi_2(w)$.*

It will be usual, 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 initial and transition probability is in $\{0, \frac{1}{2}, 1\}$. Note that this is, in some sense, the weakest set of probabilities that one can use to produce nontrivial behaviours: an automaton using only $\{0,1\}$ would only recognize regular languages. The follow lemma, left as an exercise shows that this restriction is not as strong as it seems.

Exercise 30. 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 $\mathcal{A}$ over alphabet A where all transition probabilities are dyadic, build a simple automaton $\mathcal{B}$, also over alphabet A, such that $\{\mathcal{B}(w) : w \in A^*\} = \{0\} \cup \{\mathcal{A}(w) : w \in A^*\}$. *Hint: if p is the maximum dyadic exponent that appears in $\mathcal{A}$, build $\mathcal{B}$ such that $\mathcal{A}(w_1 \cdots w_k) = \mathcal{B}(w_1^{p+1} \cdots w_k^{p+1})$.*

Theorem 31. *The Equality Problem is undecidable, even for simple automata and cut-point $\frac{1}{2}$.*

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 1 after every letter of $\phi_i(a)$ so that $\varphi_i(a) \in \{01, 11\}^*$. 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 13, define

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

Check that $[\cdot]$ is injective over $\{01, 11\}^*$ and therefore for every words $w \in A^*$,

$$[\varphi_1(w)] = [\varphi_2(w)] \text{ if and only if } \varphi_1(w) = \varphi_2(w). \tag{2}$$

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

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

One checks that $\mu(a)$ is substochastic[2] by checking that $[w] \leqslant 1 - 2^{-|w|}$ for every $w \in \{1,0\}^*$. 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

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

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

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

Finally, we can build automaton $\mathcal{B} = \frac{1}{2}\mathcal{A}_1 + \frac{1}{2}\mathcal{A}_2^c$. We then obtain that every word $w \in A^*$,

$$\mathcal{B}(w) = \tfrac{1}{2} \quad \Longleftrightarrow \quad \tfrac{1}{2}\mathcal{A}_1(w) + \tfrac{1}{2}\mathcal{A}_2^c(w) = \tfrac{1}{2} \quad \Longleftrightarrow \quad [\varphi_1(w)] = [\varphi_2(w)] \quad \Longleftrightarrow \quad \varphi_1(w) = \varphi_2(w)$$

using (2). It remains to deal with the empty word since $\mathcal{B}(\varepsilon) = \frac{1}{2}$ but we do not want to accept the empty word. Therefore $\mathcal{B}[\varepsilon \leftarrow 0]$ 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 30). Furthermore, check that the construction of $\mathcal{B}[\varepsilon \leftarrow 0]$ preserves simplicity. □

[2] It is possible to build stochastic matrices directly by taking $S = \begin{bmatrix} 1 & 0 \end{bmatrix}$, $\mu_i(a) = \begin{bmatrix} 1 - [\varphi_i(a)] & [\varphi_i(a)] \\ 1 - [1\varphi_i(a)] & [1\varphi_i(a)] \end{bmatrix}$ 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 $\varphi_i(a) \in 1\{0,1\}^*$.

We will now extend this result to other problems by a clever trick to encode an equality problem into an inequality one.

Proposition 32. *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 $\mathcal{A}(w) = \frac{1}{2}$,*
- *there exists a word w such that $\mathcal{B}(w) \geqslant \frac{1}{4}$,*
- *there exists a word w such that $\mathcal{C}(w) > \frac{1}{8}$.*

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 $\mathcal{B} = \mathcal{A} \cdot \mathcal{A}^c$ (see Lemma 22), then $\mathcal{B}(w) = \mathcal{A}(w)(1-\mathcal{A}(w))$. By construction, all transition probabilities of $\mathcal{B}$ are already multiple of $\frac{1}{4}$ (see Exercise 33).

To build $\mathcal{C}$, we start by noticing that since all initial and transition probabilities of $\mathcal{B}$ are multiple of $\frac{1}{4}$, $\mathcal{B}(w)$ is a multiple of $4^{-|w|-1}$ for every word w. Thus $\mathcal{B}(w) \geqslant \frac{1}{4}$ if and only if $\mathcal{B}(w) > \frac{1}{4} - 4^{-|w|-1}$ if and only if $\frac{1}{2}\mathcal{B}(w) + \frac{1}{2}4^{-|w|-1} > \frac{1}{8}$. One easily builds an automaton $\mathcal{D}$ such that $\mathcal{D}(w) = 4^{-|w|-1}$ for any word w. Then $\mathcal{C} = \frac{1}{2}\mathcal{B} + \frac{1}{2}\mathcal{D}$ satisfies that

$$\mathcal{C}(w) > \tfrac{1}{8} \quad \Longleftrightarrow \quad \mathcal{B}(w) > \tfrac{1}{4} - 4^{-|w|-1} \quad \Longleftrightarrow \quad \mathcal{B}(w) \geqslant \tfrac{1}{4}.$$

Finally, we can make $\mathcal{B}$ and $\mathcal{C}$ simple by Exercise 30. □

Exercise 33. Show that if $\mathcal{A}$ and $\mathcal{B}$ are simple then all initial and transition probabilities of $\mathcal{A} \cdot \mathcal{B}$ are multiple of $\frac{1}{4}$.

As a consequence of Theorem 31 and Proposition 32, we get:

Theorem 34. *The emptiness and strict emptiness problems are undecidable, even for simple automata and a fixed dyadic cut-point in* $(0, 1)$.

1.4 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 35 (*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.

Remark 36. We will see an alternative proof of undecidability in Section 1.6.

Problem 37 (*$\omega-$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 38. 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 39. *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 40. *The isolation problem is undecidable, even for simple automata and a fixed dyadic cut-point in* $(0, 1)$.

Proof. We will show the result for the cut-point $\lambda = \frac{1}{2}$, this can be extended to any rational $\lambda \in (0,1)$ by Exercise 28.

The problem will essentially be the same as for Theorem 31 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 31, we define $[w] = \sum_{i=1}^{|w|} w_i 2^{-i}$ for every $w \in \{0,1\}^*$ and build a probabilistic automaton $\mathcal{C}$ such that $\mathcal{C}(w) = \frac{1}{2} + \frac{1}{2}([\phi_1(w)] - [\phi_2(w)])$ for every $w \in A^*$. Finally we let $\mathcal{B} = \mathcal{C}[\varepsilon \leftarrow 0]$ to avoid any problem with the empty word. 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, since w is non-empty,

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

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 39, 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 non-empty word w,

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

and the empty word has probability 0 so $|\mathcal{B}(\varepsilon) - \frac{1}{2}| = \frac{1}{2} \geqslant 2^{-n_0}$. Since n_0 is independent of w, this shows that $\frac{1}{2}$ is isolated. □

1.5 The value 1 problem

There is a slight discrepancy in Theorem 40 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

$$\operatorname{val}(\mathcal{A}) = \sup\{\mathcal{A}(w) : w \in A^*\}.$$

In other words, it is the supremum of the probability of acceptance over all possible input words. Note that this probability may not be achieved by any path, only as the limit of longer and longer paths.

Problem 41 (*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 $\mathcal{A}(w) > 1 - \varepsilon$.*

Remark 42. We will see an alternative proof of undecidability in Section 1.6.

Proposition 43. *Let $x \in (0,1)$, then the automaton $\mathcal{A}_x$ from Figure 3 has value* 1 *if $x > \frac{1}{2}$, and value $\frac{1}{2}$ otherwise.*

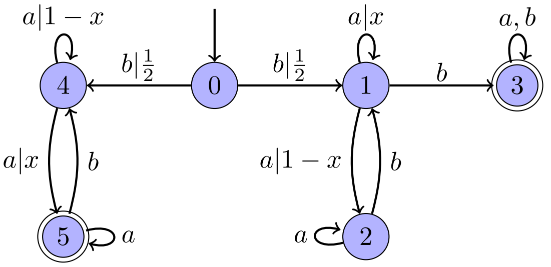


Figure 3: Auxiliary automaton for the value 1 problem.

Proof. Let $n \in \mathbb{N}$, check that if we are in state 4, the only way to reach state 4 by reading a^k is to use the self-loop k times. Furthermore, the only way to reach 4 from 4 by reading $a^k b$ is to reach 5 from 4 by reading a^k. But the only reachable states from 4 are 4 and 5 so by stochasticity, the probability to reach 5 is the complement of that of reaching 4. Therefore,

$$\mathcal{A}_x\left(4 \xrightarrow{a^n} 4\right) = (1-x)^n, \qquad \mathcal{A}_x\left(4 \xrightarrow{a^n b} 4\right) = \mathcal{A}_x\left(4 \xrightarrow{a^n} 5\right) = 1 - \mathcal{A}_x\left(4 \xrightarrow{a^n} 4\right) = 1-(1-x)^n.$$

Furthermore, for $n_0, \ldots, n_k \in \mathbb{N}$, since there is no transition from 4 labelled by b, the only way a path of the form $a^{n_0}ba^{n_1}b\cdots ba^{n_k}$ can be accepted from state 4 is by repeatedly reaching 4 from 4 when reading $a^{n_i}b$. Hence

$$\mathcal{A}_x\left(4 \xrightarrow{a^{n_0}b\cdots ba^{n_k}} 5\right) = \left(\prod_{i=0}^{k-1} \mathcal{A}_x\left(4 \xrightarrow{a^{n_i}b} 4\right)\right) \mathcal{A}_x\left(4 \xrightarrow{a^{n_k}} 5\right) = \prod_{i=0}^{k}(1-(1-x)^{n_i}).$$

A similar reasoning shows that

$$\mathcal{A}_x\left(1 \xrightarrow{a^n} 1\right) = x^n, \qquad \mathcal{A}_x\left(1 \xrightarrow{a^n} 2\right) = 1 - x^n, \qquad \mathcal{A}_x\left(1 \xrightarrow{a^n b} 1\right) = 1 - x^n.$$

Now note that the only way to reach 3 from 1 by reading $a^{n_0}b\cdots ba^{n_k}$ is to have already reached 3 after reading $a^{n_0}b\cdots a^{n_{k-1}}b$ from 1. But then, because of the last transition by b, one must be in state 1 or 3. By stochasticity, the probability to be in 3 in the complement of that of being in 1. Therefore,

$$\mathcal{A}_x\left(1 \xrightarrow{a^{n_0}b\cdots a^{n_{k-1}}b} 1\right) = \prod_{i=0}^{k-1} \mathcal{A}_x\left(1 \xrightarrow{a^{n_i}b} 1\right) = \prod_{i=0}^{k-1}(1-x^{n_i}),$$

and

$$\mathcal{A}_x\left(1 \xrightarrow{a^{n_0}b\cdots ba^{n_k}} 3\right) = \mathcal{A}_x\left(1 \xrightarrow{a^{n_0}b\cdots a^{n_{k-1}}b} 3\right) \mathcal{A}_x\left(3 \xrightarrow{a^{n_k}} 3\right) = 1 - \mathcal{A}_x\left(1 \xrightarrow{a^{n_0}b\cdots a^{n_{k-1}}b} 1\right) = 1 - \prod_{i=0}^{k-1}(1-x^{n_i}).$$

If $x > \frac{1}{2}$, by letting $n_0 = \cdots = n_k = n$ and $k = 2^n - 1$ for $n \geqslant 2$, we have

$$\begin{aligned}
\mathcal{A}_x\left(4 \xrightarrow{(a^n b)^k a^n} 5\right) &= (1-(1-x)^n)^{k+1} \\
&= e^{(k+1)\log(1-(1-x)^n)} && \text{but } (1-x)^n < \tfrac{1}{2} \\
&\geqslant e^{-2(k+1)(1-x)^n} && \text{using } \log(1-y) \geqslant -2y \text{ when } y \leqslant \tfrac{1}{2} \\
&= e^{-2(2-2x)^n} && \text{since } k-1 = 2^n \\
&\to 1 \text{ as } n \to \infty && \text{since } 2-2x < 1 \text{ since } x > \tfrac{1}{2}.
\end{aligned}$$

And similarly,

$$\begin{aligned}
\mathcal{A}_x\left(1 \xrightarrow{(a^n b)^k a^n} 3\right) &= 1-(1-x^n)^k \\
&= 1 - e^{k\log(1-x^n)} \\
&\geqslant 1 - e^{-kx^n} && \text{using } \log(1-y) \leqslant -y
\end{aligned}$$

$$\begin{aligned} &\geqslant 1 - e^{-\frac{1}{2}(2x)^n} && \text{since } k \geqslant 2^{n-1} \\ &\to 1 \text{ as } n \to \infty && \text{since } 2x > 1. \end{aligned}$$

It follows that

$$\mathcal{A}_x(b(a^n b)^{2^n} a^n) = \tfrac{1}{2}\mathcal{A}_x\left(4 \xrightarrow{(a^n b)^{2^n} a^n} 5\right) + \tfrac{1}{2}\mathcal{A}_x\left(1 \xrightarrow{(a^n b)^{2^n} a^n} 3\right) \to 1 \text{ as } n \to \infty$$

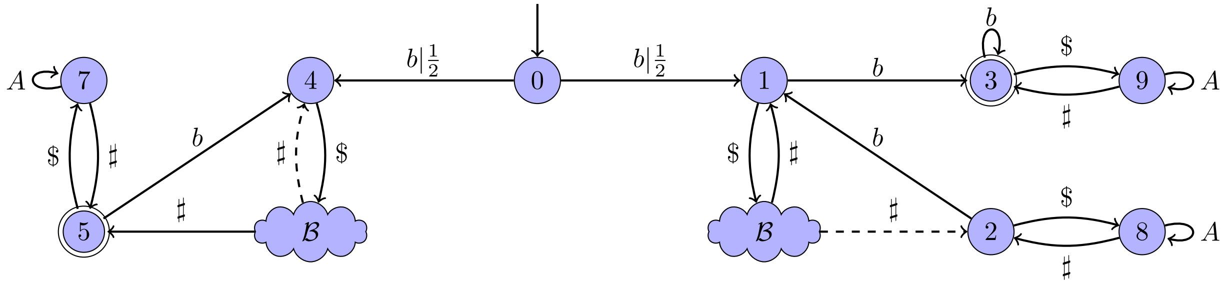
and therefore $\operatorname{val}(\mathcal{A}_x) = 1$ when $x > \frac{1}{2}$. On the other hand, if $x \leqslant \frac{1}{2}$, note that any word w that is accepted must start with a b and hence is of the form $w = ba^{n_0}ba^{n_1}b \cdots ba^{n_k}$ for some $n_0, \ldots, n_k \in \mathbb{N}$. Hence,

$$\begin{aligned} \mathcal{A}_x(w) &= \tfrac{1}{2}\mathcal{A}_x\left(1 \xrightarrow{a^{n_0}ba^{n_1}b\cdots ba^{n_k}} 3\right) + \tfrac{1}{2}\mathcal{A}_x\left(4 \xrightarrow{a^{n_0}ba^{n_1}b\cdots ba^{n_k}} 5\right) \\ &= \tfrac{1}{2} - \tfrac{1}{2}\prod_{i=0}^{k-1}(1 - x^{n_i}) + \tfrac{1}{2}\prod_{i=0}^{k}(1 - (1-x)^{n_i}) \\ &\leqslant \tfrac{1}{2} - \tfrac{1}{2}\prod_{i=0}^{k-1}(1 - x^{n_i}) + \tfrac{1}{2}\prod_{i=0}^{k}(1 - x^{n_i}) && \text{since } x^{n_i} \leqslant (1-x)^{n_i} \text{ when } x \leqslant \tfrac{1}{2} \\ &\leqslant \tfrac{1}{2}. \end{aligned}$$

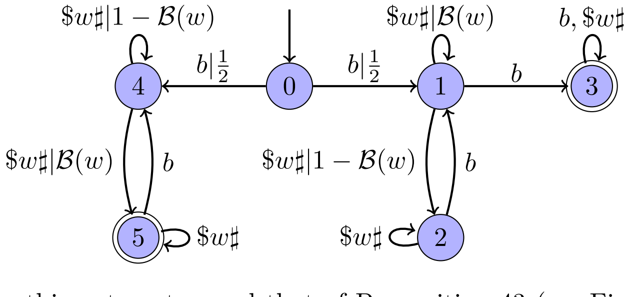
This proves that $\operatorname{val}(\mathcal{A}_x) \leqslant \frac{1}{2}$ when $x \leqslant \frac{1}{2}$. On the other hand, check that $\mathcal{A}_x(bb) = \frac{1}{2}$ so $\operatorname{val}(\mathcal{A}_x) \geqslant \frac{1}{2}$. □

Theorem 44. *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, b, \$ and $\sharp$. 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 X on the arrows means that there is a transition for every letter $a \in X$.



This automaton can be summarized with the following more informal picture, where we allow "meta transitions" of the form $\$w\sharp|p(w)$ which means that we take this transition by reading a word of the form $\$w\sharp$ for some $w \in A^*$ with has probability $p(w)$.



Note that close proximity between this automaton and that of Proposition 43 (see Figure 3). We now claim that $\mathcal{C}$ has value 1 if $\exists w \in A^*$ such that $\mathcal{B}(w) > \frac{1}{2}$, and otherwise it has value $\leqslant \frac{1}{2}$.

First, let us note that any word accepted by $\mathcal{C}$ with positive probability must be of the form

$$y = bu_1 b \cdots bu_k, \qquad \text{where } u_i = \$w_{i1}\sharp \cdots \$w_{in_i}\sharp \tag{3}$$

for some $k \in \mathbb{N}$, $n_i \in \mathbb{N}$ and $w_{ij} \in A^*$. To see that, we can partition the states of $\mathcal{C}$ into two groups:

- "type A" states: $0, 1, 2, 3, 4, 5$,

- "type B" states: $7, 8, 9$ and the states of the two copies of $\mathcal{B}$.

Then observe that (1) we start in a type A state, (2) accepting states have type A, (3) the only valid transitions in type A states are b and (possibly) \$, (4) reading \$ in a type A state leads to a type B state, (5) the only valid transitions in type B states are A and $\sharp$, (6) reading $\sharp$ in a type B state leads to a type A state. The general form (3) then follows from those observations.

Second, having summarized the automaton $\mathcal{C}$ as in the second picture, one can see that the exact same proof as that of Proposition 43 shows that if y is as in (3), then its probability of acceptance is

$$\mathcal{C}(y) = \tfrac{1}{2} + \tfrac{1}{2}\prod_{i=0}^{k}\left(1 - \prod_{j=1}^{n_i}(1 - \mathcal{B}(w_{ij}))\right) - \tfrac{1}{2}\prod_{i=0}^{k}\left(1 - \prod_{j=1}^{n_i}\mathcal{B}(w_{ij})\right). \tag{4}$$

Also recall that the probability of acceptance in $\mathcal{A}_x$ in Proposition 43 was

$$\mathcal{A}_x(a^{n_0}ba^{n_1}b\cdots ba^{n_k}) = \tfrac{1}{2} + \tfrac{1}{2}\prod_{i=0}^{k}(1 - x^{n_i}) - \tfrac{1}{2}\prod_{i=0}^{k-1}(1 - x^{n_i}) \tag{5}$$

We are now ready to show the result. Assume there exists a word w such that $\mathcal{B}(w) > \frac{1}{2}$ and let $x = \mathcal{B}(w)$. Let $\varepsilon > 0$, since $\mathcal{A}_x$ has value 1 by Proposition 43, there exists $n_1, \dots, n_k$ such that $\mathcal{A}_x(ba^{n_1}b\cdots a^{n_k}) \geqslant 1 - \varepsilon$. But now observe that by (4) and (5),

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

It follows that $\operatorname{val}(\mathcal{C}) = 1$.

Conversely, assume that $\mathcal{B}(w) \leqslant \frac{1}{2}$ for every word $w \in A^*$. Recall that only words y of the form (3) are accepted by $\mathcal{C}$ and let $x = \max_{i,j}\mathcal{B}(w_{ij}) \leqslant \frac{1}{2}$. Then

$$\begin{aligned}
\mathcal{C}(y) &= \tfrac{1}{2} + \tfrac{1}{2}\prod_{i=0}^{k}\left(1 - \prod_{j=1}^{n_i}(1 - \mathcal{B}(w_{ij}))\right) - \tfrac{1}{2}\prod_{i=0}^{k}\left(1 - \prod_{j=1}^{n_i}\mathcal{B}(w_{ij})\right) && \text{by (4)}\\
&\leqslant \tfrac{1}{2} + \tfrac{1}{2}\prod_{i=0}^{k}\left(1 - \prod_{j=1}^{n_i}(1 - x)\right) - \tfrac{1}{2}\prod_{i=0}^{k}\left(1 - \prod_{j=1}^{n_i}x\right) && \text{since } \mathcal{B}(w_{ij}) \leqslant x\\
&= \tfrac{1}{2} + \tfrac{1}{2}\prod_{i=0}^{k}(1 - (1 - x)^{n_i})) - \tfrac{1}{2}\prod_{i=0}^{k}(1 - x^{n_i})\\
&= \mathcal{A}_x(ba^{n_1}b\cdots ba^{n_k}) && \text{by (5)}\\
&\leqslant \tfrac{1}{2}
\end{aligned}$$

by Proposition 43 since $x \leqslant \frac{1}{2}$. Hence we have shown that $\mathcal{C}(y) \leqslant \frac{1}{2}$ for all words y, so in particular $\operatorname{val}(\mathcal{C}) \leqslant \frac{1}{2} < 1$. □

1.6 The value approximation problem

We have seen in the previous sections that it is undecidable to check whether a cut-point λ is isolated, even when $\lambda = 1$. Said differently, we cannot determine if the value (which is a limit) is bigger than a particular number. However, it seems reasonable to expect that this limit is at least *approximable* with small or even arbitrarily small error. A very surprising result by Condon and Lipton is that even computing an approximation with error strictly less than $\frac{1}{2}$ is impossible [CL89]. In fact, Fijalkow showed an even stronger result: in some sense, we cannot approximate it even if we allow the algorithm to be incorrect or not terminating sometimes [Fij17].

Theorem 45. *There is no algorithm such that given a probabilistic automaton $\mathcal{A}$,*

- *if* $\operatorname{val}(\mathcal{A}) = 1$*, then the algorithm outputs "yes",*
- *if* $\operatorname{val}(\mathcal{A}) \leqslant \frac{1}{2}$*, then the algorithm outputs "no",*
- *otherwise, the algorithm can output anything or not terminate.*

Proof. In fact we have already proven this fact: in the proof of Theorem 44, we have seen that the automaton $\mathcal{C}$ that has value either 1 or $\frac{1}{2}$ depending on whether $\mathcal{L}_{\mathcal{B}}(\frac{1}{2}) = \varnothing$ or not. Hence if an algorithm as described in the statement of the theorem existed, it would decide if $\mathcal{L}_{\mathcal{B}}(\frac{1}{2}) = \varnothing$ which is not possible by Theorem 44. □

We now explore some consequences of it. Essentially, this theorem subsumes all the undecidability results that we have seen so far.

Corollary 46. *For any fixed $\lambda \in (0,1]$, the problems of deciding, given an automaton $\mathcal{A}$, whether* $\operatorname{val}(\mathcal{A}) > \lambda$ *(resp.* $\operatorname{val}(\mathcal{A}) \geqslant \lambda$*) is undecidable. In particular, the value* 1 *problem is undecidable.*

Proof. We look at the strict problem first: if $1 > \lambda \geqslant \frac{1}{2}$ then this is immediate. Indeed, if the problem was decidable, it would give us an algorithm that, in particular, outputs "yes" when $\operatorname{val}(\mathcal{A}) = 1$ and "no" when $\operatorname{val}(\mathcal{A}) \leqslant \frac{1}{2}$, which contradicts Theorem 45. When $\lambda < \frac{1}{2}$, assume that we have an algorithm that decides whether $\operatorname{val}(\mathcal{A}) > \lambda$. Let $k \in \mathbb{N}$ be such that $2^k\lambda \in [\frac{1}{2}, 1)$ and consider the algorithm that given $\mathcal{A}$, builds $\mathcal{B}$ such that $\mathcal{B}(w) = 2^{-k}\mathcal{A}(w)$ and runs the algorithm on $\mathcal{B}$. Then $\operatorname{val}(\mathcal{B}) = 2^{-k}\operatorname{val}(\mathcal{A})$ and therefore the algorithm accepts if and only if $\operatorname{val}(\mathcal{A}) > 2^k\lambda$. But since $2^k\lambda \geqslant \frac{1}{2}$, we are back to the previous case where we have shown that such an algorithm cannot exists.

The non-strict problem is exactly is the same except that the case distinction is on $\lambda > \frac{1}{2}$. The value 1 problem is clearly equivalent to $\operatorname{val}(\mathcal{A}) \geqslant 1$. □

Corollary 47. *The emptiness, strict-emptiness and universality problems are undecidable.*

Proof. Let $\mathcal{A}$ be an automaton and $\lambda \in (0,1)$ a cut-point, then observe that there exists w such that $\mathcal{A}(w) > \lambda$ if and only if $\operatorname{val}(\mathcal{A}) > \lambda$. But checking if the value of an automaton is strictly bigger than λ is undecidable by the previous corollary. The emptiness problem reduces to the strict emptiness problem using a similar construction to that of Proposition 32, and universality is simply the emptiness of the complement. □

Corollary 48. *The isolation problem is undecidable for any cut-point.*

Proof. Let $\lambda \in [0,1]$: if $\lambda = 0$ then note that 0 is isolated for $\mathcal{A}$ if and only if 1 is isolated for $1 - \mathcal{A}$, so we can always assume that $\lambda > 0$. Assume that there is an algorithm to decide isolated of λ, we will show that there is an algorithm that satisfies Theorem 45 and reach a contradiction. Indeed, given $\mathcal{A}$, we can build $\mathcal{B} = \lambda\mathcal{A}$, run the isolation algorithm on $\mathcal{B}$ for λ and output the opposite. If $\operatorname{val}(\mathcal{A}) = 1$ then $\operatorname{val}(\mathcal{B}) = \lambda$ so λ is not isolated so the algorithm outputs "yes". If $\operatorname{val}(\mathcal{A}) \leqslant \frac{1}{2}$ then $\operatorname{val}(\mathcal{B}) = \frac{\lambda}{2} < \lambda$ so λ is isolated and we output "no". The other cases don't matter, and we have indeed reached a contradiction. □

1.7 The density problem

We have seen in Section 1.1.3 that isolated cut-points yield regular languages but that deciding whether a given cut-point is isolated is undecidable (Section 1.4 for cut-points in $(0,1)$ and Section 1.5 for 0 and 1). A potential much weaker is to ask whether a given automaton has *any* isolated cut-point, or equivalenty whether the set of acceptance probabilities is dense in $[0,1]$.

Problem 49 (Density)**.** *Given a probabilistic automaton $\mathcal{A}$ over A, decide whether $\{\mathcal{A}(w) : w \in A^*\}$ is dense in $[0,1]$.*

Surprisingly, this problem is also undecidable. However this result will not follow from Theorem 45 and will require a completely different proof. Intuitively, this is related to whether the smallest set stable under a certain linear map is dense in $[0,1]$. We will start with a warm-up to give the intuition of the construction, which we then modify in a similar fashion to Section 1.5.

Lemma 50. *Let $u \in [0, \frac{1}{4}]$ and let $D_u \subseteq [0,1]$ be the smallest set such that $0 \in D_u$ and if $x \in D_u$ then $f_i(u,x) \in D_u$ for all $i \in \{0,1,2,3\}$, where $f_i(u,x) := \frac{1-u}{3}i + ux$. Then D_u is dense in $[0,1]$ if and only if $u = \frac{1}{4}$.*

Proof. If $u = \frac{1}{4}$ then $0 \in D_u$ and for all $i \in \{0,1,2,3\}$, $f_i(u,x) = \frac{1-u}{3}i + ux = \frac{i+x}{4} \in D_u$. In other words, D_u contains all 4$-$adic rationals which are clearly dense in $[0,1]$.

Conversely, let $u < \frac{1}{4}$ and $\varepsilon = \frac{1}{4} - u > 0$. We will show that show that $D_u \subseteq X$, where $X = [0, \frac{1}{4} - \varepsilon] \cup [\frac{1}{4}, 1]$. Observe that

$$f_0(u,X) = uX \subseteq [0,u] \subseteq [0, \tfrac{1}{4} - \varepsilon] \subseteq X$$

and for $i \in \{1,2,3\}$,

$$f_i(u,X) = \tfrac{1-u}{3}i + uX \subseteq \tfrac{1-u}{3}i + [0,u] \subseteq [\tfrac{1-u}{3}i, 1] \subseteq [\tfrac{1-u}{3}, 1] \subseteq [\tfrac{1}{4}, 1]$$

since $\frac{1-u}{3} \geqslant \frac{1}{4}$ since $u < \frac{1}{4}$. It follows that $0 \in X$ and $f_i(X) \subseteq X$, therefore $D_u \subseteq X$ since D_u is the smallest such set. Clearly X is not dense in $[0,1]$ so D_u is not either. □

We will now see how to encode the set D_u defined in the previous lemma in a probabilistic automaton. Intuitively, this is possible because the maps f_i are linear and endomorphisms of $[0,1]$.

Lemma 51. *Let $u \in [0, \frac{1}{4}]$ and D_u be defined as in Lemma 50. Let $\mathcal{A}_u = (A, Q, S, T, \mu)$ where $A = \{0, 1, 2, 3\}$, $Q = \{1, 2\}$,*

$$S = \begin{bmatrix} 1 & 0 \end{bmatrix}, \qquad \mu(i) = \begin{bmatrix} 1 - a_i & a_i \\ 1 - b_i & b_i \end{bmatrix}, \qquad T = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

where $a_i = \frac{1-u}{3} i$ and $b_i = u + a_i$. Then $\mathcal{A}_u$ is a probabilistic automaton and $\{\mathcal{A}_u(w) : w \in A^*\} = D_u$.

Proof. Let $u \in [0, \frac{1}{4}]$ and D_u and f_u be as in Lemma 50. First note that $a_i, b_i \in [0, 1]$ so $\mathcal{A}_u$ is indeed a probabilistic automaton when $u \in [0, \frac{1}{4}]$. Then check that for every i, $f_i(u, x) = \frac{1-u}{3} i + ux = a_i(1 - x) + b_i x$. Let $X = \{\mathcal{A}_u(w) : w \in A^*\}$, we claim that $X = D_u$. First note that $\mathcal{A}_u(0) = 0$ so $0 \in X$. If $x \in X$ then by stochasticity and the definition of S, there exists $w \in A^*$ such that $S\mu(w) = \begin{bmatrix} 1 - x & x \end{bmatrix}$. But then for any $i \in A$,

$$\mathcal{A}_u(wi) = \mu(w)\mu(i)T = \begin{bmatrix} 1 - x & x \end{bmatrix} \begin{bmatrix} a_i \\ b_i \end{bmatrix} = a_i(1 - x) + b_i x = f_i(u, x)$$

and hence $f_i(u, x) \in X$. This shows that $D_u \subseteq X$. But conversely, by construction, any $x \in X$ is of the form $x = \mathcal{A}_u(w)$ for some $w \in A^*$. By induction on $|w|$, one shows that $\mathcal{A}_u(w) \in D_u$ since for $|w| = 0$, $\mathcal{A}_u(w) = 0 \in D_u$ and if we have shown that $\mathcal{A}_u(w) \in D_u$ then $\mathcal{A}_u(wi) = f_i(u, \mathcal{A}_u(w)) \in D_u$ for any $i \in A$. This shows that $X \subseteq D_u$ and completes the proof. □

Corollary 52. *Let $\mathcal{A}_u$ be as in Lemma 51, then $\mathcal{A}_u$ has an isolated cut-point if and only if $u < \frac{1}{4}$.*

We will now see how this construction can be modified so that u is replaced by an arbitrary automaton $\mathcal{E}$. If we can ensure that the probabilities of the words are in $[0, \frac{1}{4}]$, the construction will have an isolated cut-point if and only if $\frac{1}{4}$ is isolated in $\mathcal{E}$. The first step is to modify Lemma 51 and replace the a_i and b_i by arbitrary automata.

Lemma 53. *Let $\mathcal{B}, \mathcal{C}$ be arbitrary automata over some alphabet Σ. Let $\sharp \notin \Sigma$, then there exists a probabilistic automaton $\mathcal{D}$ over $\Sigma' := \Sigma \cup \{\sharp\}$ such that for all $w^{(1)}, \dots, w^{(k)} \in \Sigma^*$,*

$$\mathcal{D}(\sharp w^{(1)} \sharp w^{(2)} \sharp \cdots \sharp w^{(k)} \sharp) = \begin{bmatrix} 1 & 0 \end{bmatrix} \prod_{i=1}^{k} M(w^{(i)}) \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad \text{where} \quad M(w) := \begin{bmatrix} 1 - \mathcal{B}(w) & \mathcal{B}(w) \\ 1 - \mathcal{C}(w) & \mathcal{C}(w) \end{bmatrix} \qquad \forall w \in \Sigma^*.$$

Proof. We write $\mathbf{1}$ (resp. $\mathbf{0}$) for the all-one (resp. all-zero) vector. For any vector x, we let $x^c := \mathbf{1} - x$.

Write $\mathcal{B} = (\Sigma, Q_1, S_1, \mu_1, T_1)$ and $\mathcal{C} = (\Sigma, Q_2, S_2, \mu_2, T_2)$. Let $\mathcal{D} = (A', Q', S', \mu', T')$ where $Q' = Q_1 \cup Q_2$,

$$S' = \tfrac{1}{\alpha} \begin{bmatrix} (T_1^c)^T & (T_2^c)^T \end{bmatrix}, \quad \mu'(\sigma) = \begin{bmatrix} \mu_1(\sigma) & \mathbf{0} \\ \mathbf{0} & \mu_2(\sigma) \end{bmatrix}, \quad \mu'(\sharp) = \begin{bmatrix} T_1^c & T_1 \\ T_2^c & T_2 \end{bmatrix} \begin{bmatrix} S_1 & \mathbf{0} \\ \mathbf{0} & S_2 \end{bmatrix} = \begin{bmatrix} T_1^c S_1 & T_1 S_2 \\ T_2^c S_1 & T_2 S_2 \end{bmatrix}, \quad T' = \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}$$

and $\alpha = (T_1^c + T_2^c)^T \mathbf{1}$ is such that S' is stochastic. Note that $\mu'(\sigma)$ is stochastic and $\mu'(\sharp)$ is the product of stochastic matrices, so is stochastic. Let $w^{(1)}, \dots, w^{(k)} \in A^*$, then

$$\mathcal{D}(\sharp c^{(1)} \sharp c^{(2)} c \cdots \sharp c^{(k)} c) = S' \mu'(\sharp) \mu'(w^{(1)}) \mu'(\sharp) \cdots \mu'(\sharp) \mu'(w^{(k)}) \mu'(\sharp) = S' \begin{bmatrix} T_1^c & T_1 \\ T_2^c & T_2 \end{bmatrix} A_1 \cdots A_k \begin{bmatrix} S_1 & \mathbf{0} \\ \mathbf{0} & S_2 \end{bmatrix} T'$$

where

$$\begin{aligned} A_i &= \begin{bmatrix} S_1 & \mathbf{0} \\ \mathbf{0} & S_2 \end{bmatrix} \begin{bmatrix} \mu_1(w^{(i)}) & \mathbf{0} \\ \mathbf{0} & \mu_2(w^{(i)}) \end{bmatrix} \begin{bmatrix} T_1^c & T_1 \\ T_2^c & T_2 \end{bmatrix} \\ &= \begin{bmatrix} S_1 \mu_1(w^{(i)}) & \mathbf{0} \\ \mathbf{0} & S_2 \mu_2(w^{(i)}) \end{bmatrix} \begin{bmatrix} T_1^c & T_1 \\ T_2^c & T_2 \end{bmatrix} \\ &= \begin{bmatrix} S_1 \mu_1(w^{(i)}) T_1^c & S_1 \mu_1(w^{(i)}) T_1 \\ S_2 \mu_2(w^{(i)}) T_2^c & S_2 \mu_2(w^{(i)}) T_2 \end{bmatrix} \\ &= \begin{bmatrix} 1 - \mathcal{B}(w^{(i)}) & \mathcal{B}(w^{(i)}) \\ 1 - \mathcal{C}(w^{(i)}) & \mathcal{C}(w^{(i)}) \end{bmatrix}. \end{aligned}$$

Furthermore,

$$S' \begin{bmatrix} T_1^c & T_1 \\ T_2^c & T_2 \end{bmatrix} = \tfrac{1}{\alpha} \begin{bmatrix} (T_1^c)^T & (T_2^c)^T \end{bmatrix} \begin{bmatrix} T_1^c & T_1 \\ T_2^c & T_2 \end{bmatrix} = \tfrac{1}{\alpha} \begin{bmatrix} (T_1^c)^T T_1^c + (T_2^c)^T T_2^c & (T_1^c)^T T_1 + (T_2^c)^T T_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

Indeed, if $v \in \{0, 1\}^n$ then $v^T v^c = 0$ and $v^T v = v^T(\mathbf{1} - v^c) = v^T \mathbf{1} - v^T v^c = v^T \mathbf{1}$, therefore $(T_1^c)^T T_1^c + (T_2^c)^T T_2^c = (T_1^c)^T \mathbf{1} + (T_2^c)^T \mathbf{1} = \alpha$. Finally,

$$\begin{bmatrix} S_1 & \mathbf{0} \\ \mathbf{0} & S_2 \end{bmatrix} T' = \begin{bmatrix} S_1 & \mathbf{0} \\ \mathbf{0} & S_2 \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix} = \begin{bmatrix} S_1 \mathbf{0} \\ S_2 \mathbf{1} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

by stochasticity of S_1 and S_2. This shows the result. □

We will now instantiante this construction for a particular choice of $\mathcal{B}$ and $\mathcal{C}$ that mimics the choice of a_i and b_i in Lemma 51.

Lemma 54. *Let $\mathcal{E}$ be an arbitrary automaton over some alphabet Γ. There exists automata $\mathcal{B}$ and $\mathcal{C}$ over $\Sigma := A \cup \Gamma$, where $A = \{1, 2, 3, 4\}$, such that for any word $w \in \Gamma^*$ and $i \in A$,*

$$\mathcal{B}(iw) = \frac{1 - \mathcal{E}(w)}{3} i, \qquad \mathcal{C}(iw) = \mathcal{E}(w) + \mathcal{B}(iw),$$

and $\mathcal{B}(u) = \mathcal{C}(u) = 0$ for all $u \notin A\Gamma^*$. In particular, if f_i is defined as in Lemma 50 and M is defined as in Lemma 53, then for all $i \in A$, $w \in \Gamma^*$ and $x \in [0, 1]$ we have

$$\begin{bmatrix} 1 - x & x \end{bmatrix} M(iw) = \begin{bmatrix} 1 - f_i(\mathcal{E}(w), x) & f_i(\mathcal{E}(w), x) \end{bmatrix}.$$

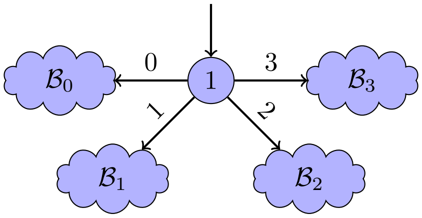
and for all $w' \notin A\Gamma^*$ we have

$$\begin{bmatrix} 1 - x & x \end{bmatrix} M(w') = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

Proof. For every $i \in A$, let $\mathcal{B}_i$ and $\mathcal{C}_i$ be such that

$$\mathcal{B}_i(w) = \tfrac{1 - \mathcal{E}(w)}{3} i, \qquad \mathcal{C}_i(w) = \mathcal{E}(w) + \mathcal{B}_i(w).$$

This is possible because for each i, $1 - \mathcal{E}(w)$ corresponds to the complement, and multiplying by $\frac{i}{3}$ can be done trivially by a convex combination of $1 - \mathcal{E}(w)$ and the automaton that has constant probability 0 for all words. Hence $\mathcal{B}_i$ is immediately seen to be a probabilistic automaton. Similarly, observe that $\mathcal{C}_i(w) = \frac{i}{3} + (1 - \frac{i}{3})\mathcal{E}(w)$ and hence is also a convex combination of $\mathcal{E}$ and the automaton that has constant probability 1. From this, we construct $\mathcal{B}$ such that $\mathcal{B}(iw) = \mathcal{B}_i(w)$ as follows.



It is sub-stochastic and can be made stochastic with a sink state. Clearly $\mathcal{B}(\varepsilon) = 0$ and $\mathcal{B}(iw) = \mathcal{B}_i(w)$ for all $w \in \Gamma^*$. Furthermore, if $w \notin \Gamma^*$ then $\mathcal{B}(iw) = 0$ since there $\mathcal{B}_i$ only has letters labelled by Γ (ie the transition will lead to the sink state). Finally, if the first letter is not in A, it will also lead to a sink state. The construction for $\mathcal{C}$ is exactly the same.

Let f_i be defined as in Lemma 50 and M as in Lemma 53. For all $i \in A$ and $w \in \Gamma^*$ we have ($\star$ denotes that the value is computed by stochasticity)

$$\begin{aligned}
\begin{bmatrix} 1 - x & x \end{bmatrix} M(iw) &= \begin{bmatrix} 1 - x & x \end{bmatrix} \begin{bmatrix} 1 - \mathcal{B}(iw) & \mathcal{B}(iw) \\ 1 - \mathcal{C}(iw) & \mathcal{C}(iw) \end{bmatrix} \\
&= \begin{bmatrix} 1 - x & x \end{bmatrix} \begin{bmatrix} 1 - \mathcal{B}_i(w) & \mathcal{B}_i(w) \\ 1 - \mathcal{C}_i(w) & \mathcal{C}_i(w) \end{bmatrix} \\
&= \begin{bmatrix} \star & (1 - x)B_i(w) + C_i(w) \end{bmatrix} \\
&= \begin{bmatrix} \star & B_i(x) + (C_i(w) - B_i(w))x \end{bmatrix} \\
&= \begin{bmatrix} \star & \frac{1 - \mathcal{E}(w)}{3} i + \mathcal{E}(w)x \end{bmatrix} \\
&= \begin{bmatrix} \star & f_i(\mathcal{E}(w), x) \end{bmatrix}.
\end{aligned}$$

For any $w' \notin A\Gamma^*$, we have

$$\begin{aligned}
\begin{bmatrix} 1 - x & x \end{bmatrix} M(w') &= \begin{bmatrix} 1 - x & x \end{bmatrix} \begin{bmatrix} 1 - \mathcal{B}(w') & \mathcal{B}(w') \\ 1 - \mathcal{C}(w') & \mathcal{C}(w') \end{bmatrix} \\
&= \begin{bmatrix} 1 - x & x \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \end{bmatrix}.
\end{aligned}$$

□

Contrary to Lemma 53, we do not necessarily have that the set of acceptance probabilities of $\mathcal{D}$ is of the form D_u for some u. However, it will contain many such sets, in fact at least one per acceptance probability of $\mathcal{E}$.

Lemma 55. *If $\mathcal{D}$ is defined as in Lemma 53 with $\mathcal{B}, \mathcal{C}, \mathcal{E}$ from Lemma 54, then for all $w \in \Gamma^*$, $D_{\mathcal{E}(w)} \subseteq \{\mathcal{D}(v) : v \in \Sigma'^*\}$.*

Proof. We show a slightly stronger result. Let $X = \{\mathcal{D}(v) : v \in \sharp(\Sigma^*\sharp)^*\}$, then $0 \in X$ since $\mathcal{D}(\sharp) = 0$ by Lemma 53 Furthermore, if $x \in X$ then there exists $v \in \sharp(\Sigma^*\sharp)^*$ such that $S'\mu'(v) = \begin{bmatrix}1-x & x\end{bmatrix}$. But then, by Lemma 53 and Lemma 54, for every $i \in A$,

$$\mathcal{D}(viw\sharp) = S'\mu'(viw\sharp)T' = \begin{bmatrix}1-x & x\end{bmatrix} M(iw) \begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}1-f_i(\mathcal{E}(w),x) & f_i(\mathcal{E}(w),x)\end{bmatrix} \begin{bmatrix}0\\1\end{bmatrix} = f_i(\mathcal{E}(w),x).$$

It follows that $f_i(\mathcal{E}(w),x) \in X$. But $D_{\mathcal{E}(w)}$ is the smallest set that is stable under those operations, hence $D_{\mathcal{E}(w)} \subseteq X$. □

We now want to argue that if $\mathcal{E}$ accepts words with probabilities arbitrarily close to $\frac{1}{4}$ (i.e. $\frac{1}{4}$ is not isolated) then the sets D_u "converge" to $D_{1/4}$ as u gets close to $\frac{1}{4}$.

Lemma 56. *If $(u_n) \in [0,1]^{\mathbb{N}}$ converges to some u^* then $\overline{\bigcup_{n=0}^{\infty} D_{u_n}}$ contains $\overline{D_{u^*}}$.*

Proof. Let $X = \{x : \exists (s_n)_n$ such that $x = \lim_{n\to\infty} s_n$ and $s_n \in D_{u_n}$ for all $n\}$. Clearly $X \subseteq Y := \overline{\bigcup_{n=0}^{\infty} D_{u_n}}$. Therefore if we show that $D_{u^*} \subseteq X$, we will have $D_{u^*} \subseteq Y$ and hence $\overline{D_{u^*}} \subseteq \overline{Y} = Y$ since Y is closed; which shows the result.

It remains to see that $D_{u^*} \subseteq X$: clearly $0 \in X$ since $0 \in D_{u_n}$ for all n. Let $x \in X$ and write $x = \lim_{n\to\infty} s_n$ where $s_n \in D_{u^*}$. Then that for any $i \in A$,

$$f_i(u^*, x) = f_i(u^*, \lim_n s_n) = \lim_n f_i(u^*, s_n)$$

by continuity of f_i. Furthermore, by the uniform continuity of f_i (continuity over the compact set $[0,1]^2$), there exists α such that for any n, $|f_i(u^*, s_n) - f_i(u_n, s_n)| \leqslant \alpha|u^* - u_n|$. Hence we can write $f_i(u^*, s_n) = f_i(u_n, s_n) + \varepsilon_n$ where $|\varepsilon_n| \leqslant \alpha|u^* - u_n| \to 0$ as $n \to \infty$. Now let $s'_n = f_i(u_n, s_n)$, then $s'_n \in D_{u_n}$ since $s_n \in D_{u_n}$ and $(s'_n)_n$ has a limit since $s'_n = f_i(u^*, s_n) - \varepsilon_n \to f_i(u^*, s)$ as shown above. Therefore,

$$f_i(u^*, x) = \lim_n f_i(u^*, s_n) = \lim_n s'_n + \varepsilon_n = \lim_s s'_n \in X.$$

This shows that X is stable under the application of the $f_i(u^*, \cdot)$, hence it contains D_{u^*} which is the smallest set to satisfy this condition. □

Corollary 57. *If $\mathcal{D}$ is defined as in Lemma 53 with $\mathcal{B}, \mathcal{C}, \mathcal{E}$ from Lemma 54, and $\frac{1}{4}$ is not an isolated cut-point of $\mathcal{E}$ then $\mathcal{D}$ has no isolated cut-points.*

Proof. If $\frac{1}{4}$ is not isolated then there exists a sequence $(w_n)_n$ of words such that $\mathcal{E}(w_n) \to \frac{1}{4}$. But then by Lemma 55, we have that

$$\bigcup_n D_{\mathcal{E}(w_n)} \subseteq \{\mathcal{D}(v) : v \in \Gamma'^*\}$$

and hence, by Lemma 56,

$$\overline{D_{1/4}} \subseteq \overline{\bigcup_n D_{\mathcal{E}(w_n)}} \subseteq \overline{\{\mathcal{D}(v) : v \in \Gamma'^*\}}.$$

But $\overline{D_{1/4}} = [0,1]$ by Lemma 50, hence $\overline{\{\mathcal{D}(v) : v \in \Gamma'^*\}} = [0,1]$ and therefore $\mathcal{D}$ cannot have an isolated cut-point. □

We will now show that the converse holds: if $\frac{1}{4}$ is isolated in $\mathcal{E}$ then $\mathcal{D}$ has isolated points. This requires to generalize the argument that we used in Lemma 50.

Lemma 58. *Let $u < \frac{1}{4}$ and D'_u be the smallest set such that $0 \in D'_u$ and for all $x \in D'_u$, $i \in A$ and $u' \leqslant u$, $f_i(u', x) \in D'$. Then D'_u is not dense in $[0,1]$.*

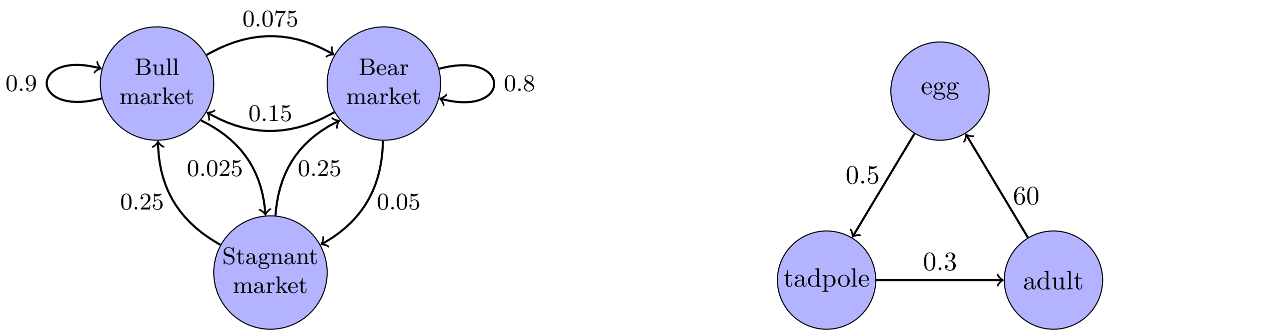
Proof. The proof is essentially the same as that of Lemma 50. Let $\varepsilon = \frac{1}{4} - u > 0$ and $X = [0, \frac{1}{4} - \varepsilon] \cup [\frac{1}{4}, 1]$. Observe that for all $u' \leqslant u$,

$$f_0(u', X) = u'X \subseteq [0, u'] \subseteq [0, u] \subseteq [0, \tfrac{1}{4} - \varepsilon] \subseteq X$$

and for $i \in \{1,2,3\}$,

$$f_i(u', X) = \tfrac{1-u'}{3} i + u'X \subseteq \tfrac{1-u'}{3} i + [0, u'] \subseteq [\tfrac{1-u'}{3} i, 1] \subseteq [\tfrac{1-u'}{3}, 1] \subseteq [\tfrac{1}{4}, 1]$$

since $\frac{1-u'}{3} \geqslant \frac{1}{4}$ since $' \leqslant u < \frac{1}{4}$. It follows that $0 \in X$ and $f_i(X) \subseteq X$, therefore $D'_u \subseteq X$ since D'_u is the smallest such set. Clearly X is not dense in $[0,1]$ so D'_u is not either. □



(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

Corollary 59. *If $\mathcal{D}$ is defined as in Lemma 53 with $\mathcal{B}, \mathcal{C}, \mathcal{E}$ from Lemma 54, and if there exists $\varepsilon > 0$ such that $\mathcal{E}(w) \leqslant \frac{1}{4} - \varepsilon$ for all $w \in \Sigma^*$, then $\mathcal{D}$ has at least one isolated cut-point.*

Proof. Let $X = \{\mathcal{D}(w) : w \in \Sigma'^*\}$. By Lemma 53 and Lemma 54, we have that every $x \in X$ is either 0 or of the form $f_i(\mathcal{E}(v), x')$ for some $v \in \Gamma^*$ and $x' \in X$. But $\mathcal{E}(v) \leqslant \frac{1}{4} - \varepsilon$, hence by letting $u = \frac{1}{4} - \varepsilon < \frac{1}{4}$, we get that $X \subseteq D'_u$ where D'_u is defined as in Lemma 58. It follows by Lemma 58 that X is not dense in $[0, 1]$. Consequently, there is an open interval I in $[0, 1]$ that does not intersect X and the center of this interval is an isolated cut-point of $\mathcal{D}$. □

Theorem 60. *The density problem is undecidable.*

Proof. We show that the problem is undecidable by reducing from the problem of deciding whether $\frac{1}{2}$ is an isolated cut-point of a given automaton. The will show the result since the latter is an undecidable problem.

Let $\mathcal{F}$ be any automaton on alphabet Γ. We can build $\mathcal{E}$ such that $\mathcal{E}(w) = \mathcal{F}(w) \cdot (1 - \mathcal{F}(w))$ for all $w \in \Gamma^*$ by the product construction. Then observe that $\mathcal{E}(w) \leqslant \frac{1}{4}$ for all w. Furthermore, $\frac{1}{4}$ is isolated for $\mathcal{E}$ if and only if $\frac{1}{2}$ is isolated for $\mathcal{F}$. We now build $\mathcal{D}$ as done in Lemma 53 and Lemma 54. By Corollary 57 and Corollary 59, we have that $\mathcal{D}$ has an isolated cut-point if and only if $\frac{1}{4}$ is isolated for $\mathcal{E}$. Therefore we have reduced the problem of whether $\frac{1}{2}$ is isolated for $\mathcal{F}$ to the problem of deciding whether $\mathcal{D}$ has an isolated cut-point. □

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

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

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 61. 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 S, 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 62. 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 $u_n = \mathcal{M}(n)$. 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 $a_0, \dots, a_{k-1} \in \mathbb{R}$. 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 63 (*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 $I_d, A, \dots, A^{d-1}$.*

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 64. *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 S and A are integers, then it is an integer LRS. In particular, if $\mathcal{M}$ is a Markov chain, then $(\mathcal{M}(n))_{n\in\mathbb{N}}$ 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 $I_d, A, \dots, A^{d-1}$ so we can find $a_0, \dots, a_{d-1} \in \mathbb{Q}$ 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 S and A are rational, then the characteristic polynomial p of A has rational entries thus the coefficients a_i are rationals. Similarly if S and A are rational, then the coefficients of p are integers.

Conversely, if $(u_n)_n$ is a LRS of order d, let $a_0, \dots, a_{d-1}$ 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 $n \in \mathbb{N}$,

$$A\begin{bmatrix} u_n \\ u_{n+1} \\ \vdots \\ u_{n+d-1} \end{bmatrix} = \begin{bmatrix} u_{n+1} \\ \vdots \\ u_{n+d-1} \\ a_0u_n + \cdots + a_{d-1}u_{n+d-1} \end{bmatrix} = \begin{bmatrix} u_{n+1} \\ \vdots \\ u_{n+d-1} \\ u_{n+d} \end{bmatrix} \quad \text{and thus} \quad A^nT = \begin{bmatrix} u_n \\ u_{n+1} \\ \vdots \\ u_{n+d-1} \end{bmatrix} \tag{6}$$

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

Remark 65. The proof of Proposition 64 could give the impression that any Markov chain or LDS $\langle S, A, T\rangle$ verifies equation (6), *i.e.* it shifts consecutive terms by one. *This is not the case in general*, see Exercise 66.

Exercise 66. 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 64, 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 67. *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. For Markov chains, this is a special case of Section 1.2, and in fact the same proofs works for non-stochastic systems as well. We redo the proof for completeness.

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 64, 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 64, $(u_n + v_n)_n$ 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 proof of Lemma 22). 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 64, $(u_n v_n)_n$ 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 68. *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*

$$u_n = A_1(n)\lambda_1^n + \dots + A_d(n)\lambda_d^n. \tag{7}$$

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[3]. Conversely, any sequence of this form is a LRS.

Proof. Put A in Jordan Normal Form (see Proposition 69) 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. □

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

Proposition 69 (*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.

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 (7).

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 Decision problems

Recall that in Section 1.3, 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 70 (*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 71 (*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.*

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 72 (*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 73 (*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 74. 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 75. 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 [Aks+15] 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 there exists $n \in \mathbb{N}$ such that $eA^n y = 1$ where $e = \begin{bmatrix}1 & 0 & \dots & 0\end{bmatrix}$.*

Proposition 76. *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 75), 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 $eP^n v = e(A^+ - A^-)^n x = A^n_{1,2}$ by showing, by induction on n, that

$$P^n v = \begin{bmatrix}(A^+ - A^-)^n x \\ -(A^+ - A^-)^n x\end{bmatrix}.$$

The second step is to rescale the matrix to make it stochastic[4], 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 77. *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 markov chain $\mathcal{M}''$ such that $eA^n y = 1$ if and only if $\mathcal{M}''(n+1) = \frac{1}{2}$ and $\mathcal{M}''(0) = 0$. This will give us an instance of the Markov reachability problem.

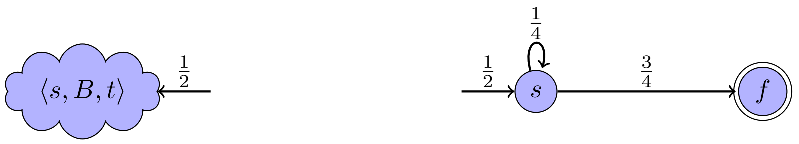
First, we need to put y in the matrix itself since we cannot have a value of 2 in the vector T and ensure that the resulting matrix is stochastic. Define

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

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

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

Next, we will use the following automaton, which is essentially the same as in the proof of Proposition 32, to "compensate" for the 4^{-n} factor.



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

Formally, let $\mathcal{M} = \langle s, B, t\rangle$ be the Markov chain above, and $\mathcal{M}'$ be the Markov chain such that $\mathcal{M}'(n) = 1 - 4^{-n}$ for all $n \in \mathbb{N}$, which is easy to define. Finally, let $\mathcal{M}'' = \frac{1}{2}\mathcal{M} + \frac{1}{2}\mathcal{M}'$. It follows that for $n \geqslant 1$,

$$\mathcal{M}''(n) = \tfrac{1}{2} \;\Leftrightarrow\; \mathcal{M}(n) = \mathcal{M}'(n) \;\Leftrightarrow\; sB^{n-1}t = 4^{-n} \;\Leftrightarrow\; eA^{n-1}y = 1.$$

Furthermore, $\mathcal{M}''(0) = \frac{1}{2}\mathcal{M}(0) = \frac{1}{2}st = 0$ by definition of s and t. □

We can now show the main result of this section.

Theorem 78. *The following problem are interreducible*[5]*:*

- *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 76 and Proposition 77, 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 64, 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 64. □

Theorem 79. *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 78 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 80. Show that A is periodic of period r if and only if there exists a finite set $P \subseteq \{0, \ldots, r-1\}$ 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 81. *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.*

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

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

Given a polynomial $P(x) = a_0 + a_1 x + \cdots + a_n x^d$ with rational coefficients, we define its valuation to be $v_p^j(P) = \min\{v_p(a_j), \dots, v_p(a_n)\}$ 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 82. *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. Write $P(x) = a_0 + a_1 x + \cdots + a_n x^n$ and $R(x) = b_0 + b_1 x + \cdots + b_{n+1} x^{n+1}$, with $a_n, b_{n+1} \neq 0$. By definition of R, we get that $b_{n+1} = a_n$, $b_{i+1} = a_i - m a_{i+1}$ and $b_0 = -m a_0$. It follows that

$$a_i = b_{i+1} + m b_{i+2} + \cdots + m^{n-i} b_{n+1}.$$

But then $v_p(a_i) \geqslant \min(v_p(b_{i+1}), \dots, v_p(b_{n+1})) = v_p^{i+1}(P)$ for all i. This implies that $v_p^i(R) \geqslant v_p^{i+1}(P)$. □

Proposition 83. *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. Assume that $\{n : b_n = 0\}$ is infinite, we will show that b_n is identically zero. It is enough to show for all $n, u \in \mathbb{N}$ that $v_p(b_n) \geqslant u$. For any $n \in \mathbb{N}$, let

$$R_n(x) = \sum_{i=0}^{n} \frac{d_i p^i}{i!} x(x-1)\cdots(x-i+1).$$

It follows that for all $n \geqslant m$, $R_n(m) = b_m$. Furthermore, $v_p^i(R_n) \geqslant i - \frac{i}{p-i}$ for all $i \in \mathbb{N}$. Indeed, if $R_n(x) = \sum_{i=0}^{n} a_i^{(n)} x^i$, then $a_i^{(n)}$ is a linear combination of $\frac{d_j p^j}{j!}$ for $j \geqslant i$. But

$$v_p\left(\tfrac{d_j p^j}{j!}\right) = v_p\left(d_j p^j\right) - v_p\left(j!\right) \geqslant v_p\left(p^j\right) - \tfrac{j}{p-1} \geqslant j - \tfrac{j}{p-1}$$

and thus $v_p(a_i^{(n)}) \geqslant i - \frac{i}{p-1}$ for all i.

Now fix $n, u \in \mathbb{N}$, and let i such that $i - \frac{i}{p-1} \geqslant u$. Let $m_1, \dots, m_i$ be distincts elements such that $b_{m_j} = 0$, and let $n_0 \geqslant \max(n, m_1, \dots, m_i)$. Then $R_{n_0}(m_j) = b_{m_j} = 0$ for all j as we have seen before (since $n_0 \leqslant m_j$). It follows that $R_{n_0}(x) = (x - m_1)\cdots(x - m_i)P(x)$ for some polynomial P. Thus

$$\begin{aligned}
v_p(b_n) &= v_p(R_{n_0}(q)) && \text{since } n \geqslant n_0 \\
&= v_p(P(q)) && \text{since } R_{n_0}(q) = P(q)y \text{ for some } y \in \mathbb{N} \\
&\geqslant v_p^0(P) && \\
&\geqslant v_p^i(R_{n_0}) && \text{by Lemma 82} \\
&\geqslant i - \tfrac{i}{p-1} \geqslant u && \text{by assumption on } i.
\end{aligned}$$

□

Proposition 84. *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 $r < p^{d^2}$ when d is the dimension of A.*

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 $\mathbb{F}$. Since $\mathrm{GL}_k(\mathbb{F})$ is finite of cardinal less than p^{d^2}, it follows that there exists $r < p^{d^2}$ such that $\tilde{A}^r = I$ and thus $A^r = I + pM$ where M is an integer coefficient matrix.

Let $j \in \{0, \dots, r-1\}$ 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 83 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 85 (Skolem–Mahler–Lech)**.** *Let $(u_n)_n$ be a LRS, then the set $Z = \{n : u_n = 0\}$ is a ultimately-periodic.*

Proof. We will show this result in the case of rational LRS only, and admit the general case. By Proposition 64, 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,

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

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 86). Then apply Proposition 84 to $\langle S', B', T'\rangle$ to conclude. Note that the resulting set is only ultimately periodic and not quasi-periodic, because of the shift $d + n$ introduced to make A invertible. □

Exercise 86. 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

[Aks+15] S. Akshay et al. “Reachability problems for Markov chains”. In: *Information Processing Letters* 115.2 (2015), pp. 155–158. ISSN: 0020-0190. DOI: https://doi.org/10.1016/j.ipl.2014.08.013. URL: http://www.sciencedirect.com/science/article/pii/S0020019014001781.

[Amb96] Andris Ambainis. “The Complexity of Probabilistic versus Deterministic Finite Automata”. In: *Proceedings of the 7th International Symposium on Algorithms and Computation.* ISAAC ’96. Berlin, Heidelberg: Springer-Verlag, 1996, pp. 233–238. ISBN: 3540620486.

[Bal14] Kaspars Balodis. “Counting with Probabilistic and Ultrametric Finite Automata”. In: *Computing with New Resources: Essays Dedicated to Jozef Gruska on the Occasion of His 80th Birthday.* Ed. by Cristian S. Calude, Rūsiņš Freivalds, and Iwama Kazuo. Cham: Springer International Publishing, 2014, pp. 3–16. ISBN: 978-3-319-13350-8. DOI: 10.1007/978-3-319-13350-8_1. URL: https://doi.org/10.1007/978-3-319-13350-8_1.

[BMT77] Alberto Bertoni, Giancarlo Mauri, and Mauro Torelli. “Some recursively unsolvable problems relating to isolated cutpoints in probabilistic automata”. In: *International Colloquium on Automata, Languages, and Programming.* Springer. 1977, pp. 87–94.

[CL89] A. Condon and R. J. Lipton. “On the Complexity of Space Bounded Interactive Proofs”. In: *Proceedings of the 30th Annual Symposium on Foundations of Computer Science.* SFCS ’89. Washington, DC, USA: IEEE Computer Society, 1989, pp. 462–467. ISBN: 0-8186-1982-1. DOI: 10.1109/SFCS.1989.63519. URL: https://doi.org/10.1109/SFCS.1989.63519.

[Fij17] Nathanaël Fijalkow. “Undecidability results for probabilistic automata”. In: *SIGLOG News* 4.4 (2017), pp. 10–17. URL: http://siglog.org/download/14th-newsletter-october-2017/?wpdmdl=395.

[FS15] Nathanaël Fijalkow and Michał Skrzypczak. “Irregular Behaviours for Probabilistic Automata”. In: *Reachability Problems.* Ed. by Mikolai Bojanczyk, Slawomir Lasota, and Igor Potapov. Cham: Springer International Publishing, 2015, pp. 33–36. ISBN: 978-3-319-24537-9.

[GO10] Hugo Gimbert and Youssouf Oualhadj. “Probabilistic Automata on Finite Words: Decidable and Undecidable Problems”. In: *Automata, Languages and Programming.* Ed. by Samson Abramsky et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 527–538. ISBN: 978-3-642-14162-1.

[Han86] G. Hansel. “Une démonstration simple du théorème de Skolem-Mahler-Lech”. In: *Theoretical Computer Science* 43 (1986), pp. 91–98. ISSN: 0304-3975. DOI: https://doi.org/10.1016/0304-3975(86)90168-4. URL: http://www.sciencedirect.com/science/article/pii/0304397586901684.

[OW12] Joël Ouaknine and James Worrell. “Decision Problems for Linear Recurrence Sequences”. In: *Reachability Problems.* Ed. by Alain Finkel, Jérôme Leroux, and Igor Potapov. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 21–28. ISBN: 978-3-642-33512-9.

[Rab63] Michael O. Rabin. “Probabilistic automata”. In: *Information and Control* 6.3 (1963), pp. 230–245. ISSN: 0019-9958. DOI: https://doi.org/10.1016/S0019-9958(63)90290-0. URL: http://www.sciencedirect.com/science/article/pii/S0019995863902900.

[Sak18] Jacques Sakarovitch. *Five lectures in the theory of Weighted Automata and Transducers.* 2018. URL: https://perso.telecom-paristech.fr/jsaka/ENSG/MPRI/FLAT.html.

A Exercises

Exercise 87. Consider the following language over alphabet $A = \{a, b\}$:

$$L = \{a^{n_1}ba^{n_2}b \cdots a^{n_k}ba^* : k > 1, \exists i > 1, n_1 = n_i\}.$$

(a) Show that L is a *context-free*[6] language. *If you don't know context-free languages, you can ignore the question.*

We now assume that $L = \mathcal{L}_{\mathcal{A}}(\lambda)$ for some probabilistic automaton $\mathcal{A} = \langle A, Q, S, \mu, T \rangle$. The goal of this exercise is to reach a contradiction, therefore showing that L is not stochastic. Let $P(x) = c_0 + c_1 x + \cdots + c_d x^n$ be the characteristic polynomial of $\mu(a)$. Recall that by Theorem 63, $P(\mu(a)) = 0$.

(b) Recall why 1 is an eigenvalue of $\mu(a)$. Show that $c_0 + \cdots + c_d = 0$ and that for any word w, $\sum_{i=0}^{d} c_i \mathcal{A}(a^i w) = 0$.

Let Pos $:= \{i : c_i > 0\}$ and NonPos $:= \{i : c_i \leqslant 0\}$. Define $w = ba^{i_1}b \cdots ba^{i_k}b$ where $\{i_1, \ldots, i_k\} =$ Pos.

(c) Let $i \in \{0, \ldots, n\}$, when is $a^i w \in L$? Show that $\sum_{i=0}^{d} c_i \mathcal{A}(a^i w) > \sum_{i=0}^{d} c_i \lambda$. Why is this a contradiction?

Exercise 88. Consider the following language over alphabet $A = \{a, b, c\}$:

$$L = \{a^n b^n c^n : n > 0\}.$$

It is a classical result that L is a not a context-free language. The goal of this exercise is to show that L is stochastic.

(a) Show that $L = L_1 \cap L_2$ where $L_1 = \{a^n b^n c^+ : n > 0\}$ and $L_2 = \{a^+ b^n c^n : n > 0\}$ where $x^+ := xx^*$.

(b) Build an automaton $\mathcal{A}_1$ such that $\mathcal{A}_1(b^m c) = 2^{-m}$ if $m > 0$ and 0 otherwise[7]. Then modify it into $\mathcal{B}_1$ such that $\mathcal{B}_1(a^* b^m c^+) = 2^{-m}$ if $m > 0$ and 0 otherwise.

(c) Build an automaton $\mathcal{A}_2$ such that $\mathcal{A}_2(a^n b) = 1 - 2^{-n}$ if $n > 0$ and 0 otherwise. Then modify it into $\mathcal{B}_2$ such that $\mathcal{B}_2(a^n b^+ c^+) = 1 - 2^{-n}$ if $n > 0$ and 0 otherwise.

(d) Build an automaton $\mathcal{C}_1$ such that $\mathcal{C}_1(a^n b^m c^+) = \frac{1}{2}(1 - 2^{-n} + 2^{-m})$ if $n, m > 0$ and 0 otherwise.

(e) Show that $L_1 = \mathcal{L}^{=}_{\mathcal{C}_1}(\frac{1}{2})$.

(f) Show that for all $x, y \in [0, 1]$, $x = \frac{1}{2} \wedge y = \frac{1}{2}$ if and only if $\frac{1}{2}x(1 - x) + \frac{1}{2}y(1 - y) = \frac{1}{4}$.

(g) Show that for any two automata $\mathcal{A}$ and $\mathcal{B}$, there exists an automaton $\mathcal{C}$ such that $\mathcal{L}^{=}_{\mathcal{C}}(\frac{1}{4}) = \mathcal{L}^{=}_{\mathcal{A}}(\frac{1}{2}) \cap \mathcal{L}^{=}_{\mathcal{B}}(\frac{1}{2})$.

(h) Conclude.

Exercise 89. We will now consider various operations on stochastic languages.

(a) Show that if L is regular then there exists an automaton $\mathcal{A}$ such that $\mathcal{A}(w) = 1$ if $w \in L$ and 0 if $w \notin L$.

(b) Let L be a regular language, $\mathcal{A}$ be a probabilistic automaton and $\lambda < 1$ a threshold, show that there exists $\mathcal{B}$ and μ, δ such that $\mathcal{L}_{\mathcal{B}}(\mu) = \mathcal{L}_{\mathcal{A}}(\lambda) \cap L$ and $\mathcal{L}_{\mathcal{B}}(\delta) = \mathcal{L}_{\mathcal{A}}(\lambda) \cup L$.

(c) Show that if $L = \mathcal{L}^{=}_{\mathcal{A}}(\frac{1}{2})$ for some automaton $\mathcal{A}$ then L is stochastic.

Consider $L = \{a^{n_1}b \cdots ba^{n_k}b : k > 1 \wedge n_1 = n_k\}$. We will show that LL' is not stochastic for some regular language L'.

(d) Build an automaton $\mathcal{A}$ such that $\mathcal{A}(a^{n_1}b \cdots ba^{n_k}b) = 1 - 2^{1-k-n_1}$ if $k \geqslant 1$.

(e) Build an automaton $\mathcal{B}$ such that $\mathcal{B}(a^{n_1}b \cdots ba^{n_k}b) = 2^{1-k-n_k}$ if $k > 1$.

(f) Show that L is a stochastic language.

(g) Show that $LA^* = \{a^{n_1}ba^{n_2}b \cdots a^{n_k}ba^* : k > 1, \exists i > 1, n_1 = n_i\}$ where $A = \{a, b\}$.

(h) Conclude using Exercise 87.

(i) Show that LcA^* is stochastic. Find a homomorphism $h : \{a, b, c\} \to A^*$ such that $h(LcA^*)$ is not stochastic.

(j) **(not easy)** Using a similar technique as in Exercise 87, show that L^* is not stochastic by consider the word $w = ba^{i_1}b(a^{i_2}b)^2 \cdots (a^{i_k}b)^2$.

[6] Also known as algebraic languages, see `https://fr.wikipedia.org/wiki/Langage_alg%C3%A9brique`.

[7] In other words, any word not of the form $b^m c$ must have probability of acceptance 0.

Exercise 90. Let $\mathcal{A}$ and $\mathcal{B}$ be two probabilistic automata. We say that they are *equivalent* if for every word w, $\mathcal{A}(w) = \mathcal{B}(w)$.

(a) Write $\mathcal{A} = \langle A, Q_1, S_1, \mu_1, T_1 \rangle$ and $\mathcal{B} = \langle A, Q_2, S_2, \mu_2, T_2 \rangle$. Recall the construction of $\mathcal{C} = \langle A, Q, S, \mu, T \rangle$ such that $\mathcal{C}(w) = \frac{1}{2}\mathcal{A}(w) + \frac{1}{2}\mathcal{B}(w)$. Find a vector $\tilde{T}$ such that for every for word w, $\mathcal{A}(w) = \mathcal{B}(w)$ if and only if $S\mu(w)\tilde{T} = 0$.

(b) Define $V_n = \operatorname{span}\{S\mu(w) : |w| \leqslant n\}$ for all $n \in \mathbb{N}$. Show that if $V_n = V_{n+1}$ for some n, then $V_{n+1} = V_{n+2}$.

(c) Define $V = \operatorname{span}\{S\mu(w) : w \in A^*\}$, show that $V = V_{d+d'}$ where d (resp. d') is the number of states of $\mathcal{A}$ (resp. $\mathcal{B}$).

(d) Show that $\mathcal{A}$ and $\mathcal{B}$ are equivalent if and only if $v\tilde{T} = 0$ for all $v \in V$.

(e) Show that if $\mathcal{A}$ and $\mathcal{B}$ are not equivalent then there exists w of length at most $d + d'$ such that $\mathcal{A}(w) \neq \mathcal{B}(w)$.

(f) Show that the equivalence problem is in coNP.

Exercise 91 ([Bal14])**.** Let $\Sigma = \{a\}$ be a unary alphabet. For any $n \in \mathbb{N}$, let $C_n = \{a^n\}$ be the language consisting of a single word a^n.

(a) By using Myhill-Nerode theorem, show that for any $n \in \mathbb{N}$, the smallest deterministic complete finite automaton recognizing C_n has exactly $n + 2$ states.

(b) Let $\delta \in [0, 1]$, build a probabilistic automaton $\mathcal{A}$ with 3 states (including any sink state) such that $\mathcal{A}(\varepsilon) = 0$ and $\mathcal{A}(a^\ell) = (1 - \delta)^{\ell-1}\delta$ for any $\ell \geqslant 1$.

(c) Modify your automaton (still with 3 states) so that $\mathcal{A}(a^\ell) = (1 - \delta)^{\ell-1}\delta\ell$ for any $\ell \geqslant 1$.

(d) Show that for any $n \in \mathbb{N}$, there exists a choice of δ such that $\ell \mapsto \mathcal{A}(a^\ell)$ has unique maximum at $\ell = n$.

(e) Show that for any $n \in \mathbb{N}$, the language C_n is recognized by a 3−states probabilistic automaton with an isolated cut-point. What is your conclusion?

(f) Show that the isolation threshold at the previous question is equivalent to $(2ne)^{-1}$ as n goes infinity.

The automaton $\mathcal{A}$ is small but has one drawback: the isolation threshold of the cut-point decreases to 0 as n increases. We will now see how to build an automaton with a constant isolation threshold. Let $p_1, p_2, \ldots,$ be the infinite sequence of primes in increasing order. Fix $n, k \in \mathbb{N}$, $\delta \in [0, 1]$. and consider the automaton $\mathcal{B}_{n,k}$ in Figure 5. It has states q^i_j for $i = 1, \ldots, k$ and $j = 0, \ldots, p_i - 1$, and an extra sink state s. For each i, j, there is a transition from q^i_j to $q^i_{(j+1) \bmod p_i}$ with probability $1 - \delta$, and a transition to s with probability δ. The initial probability distribution is $\frac{1}{k}$ in each of the states $q^1_0, \ldots, q^k_0$. The state q^i_j is final if and only if $n \equiv j \bmod p_i$. Finally, we set $\delta = 1 - \varepsilon^{1/n}$ where $\varepsilon \in (0, 1)$ is fixed.

(g) Let $\ell \in \mathbb{N}$, show that $\mathcal{B}_{n,k}(a^\ell) = \dfrac{1}{k}\displaystyle\sum_{i=1}^{k} \mathbb{1}\left[n \equiv \ell \bmod p_i\right](1 - \delta)^\ell$ and $\mathcal{B}_{n,k}(a^\ell) = \varepsilon$.

(h) Show that $\mathcal{B}_{n,k}(a^\ell) \leqslant \varepsilon^2$ for all $\ell \geqslant 2n$.

We now assume k is chosen so that $p_1 \cdots p_k > n$ and we let $r(n) = \min\{i : p_1 \cdots p_i \geqslant n\}$.

(i) Let $0 \leqslant \ell < 2n$ be distinct from n and $m = |\{i : n \equiv \ell \bmod p_i\}|$. Show that $\mathcal{B}_{n,k}(a^\ell) \leqslant \frac{m}{k}$ and if $m > r(n)$ then $\ell \geqslant 2n$.

(j) Show that for a certain choice of ε and α, C_n is recognised by $\mathcal{B}_{n,\alpha r(n)}$ with a constant isolation threshold.

Some well-known results on the distribution of primes imply that $r(n) = O\left(\frac{\ln n}{\ln \ln n}\right)$ and that $\sum_{i=1}^{k} p_i \leqslant k \ln k$.

(k) Show that the automaton $\mathcal{B}_{n,\alpha r(n)}$ has $O\left(\frac{\ln^2 n}{\ln \ln n}\right)$ states.

Exercise 92. The goal of this exercise is to show Theorem 19, following [Amb96]. Let $m \geqslant 1$ and $A = \{a_1, \ldots, a_m\}$ be an alphabet of size m. We consider the language L_m over A that contains each letter of the alphabet exactly m times:

$$L_m = \{w \in A^* : \forall a \in A, |w|_a = m\}.$$

(a) By using Myhill-Nerode theorem, show that the smallest deterministic complete finite automaton recognizing L_m has exactly $(m + 1)^m + 1$ states.

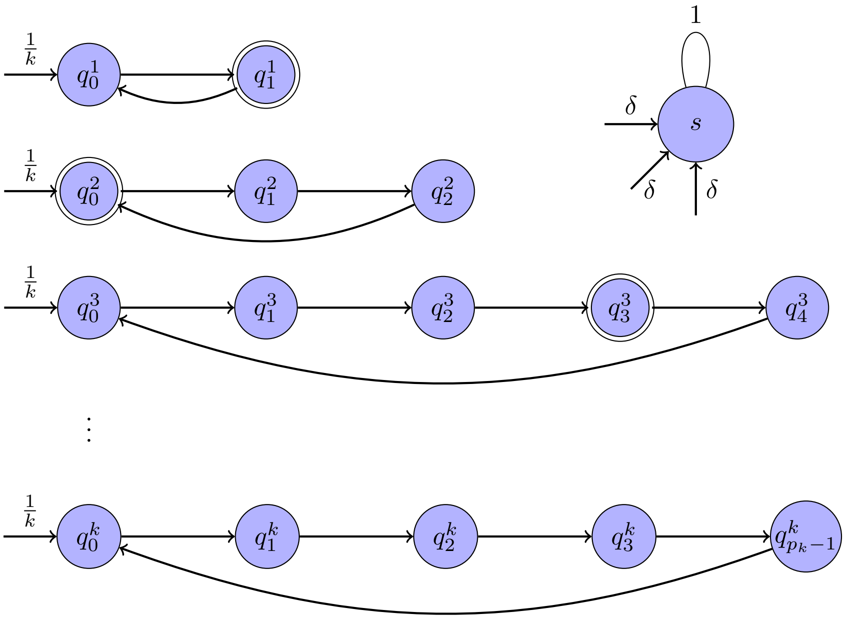


Figure 5: Automaton $\mathcal{B}_{n,k}$; for readability all edges have probability $1-\delta$ when there is no label.

Let $\Sigma = \{x\}$ be a unary alphabet. For any $n \in \mathbb{N}$, let $C_n = \{x^n\}$ be the language over Σ consisting of a single word x^n. We admit the following result, proven in Exercise 91.

Lemma 93. *There exists $\delta > 0$ such that for every n, there exists a probabilistic automaton $\mathcal{A}_n$ with $O\left(\frac{\ln^2 n}{\ln \ln n}\right)$ states that recognizes C_n with an isolated cut-point and isolation threshold at least δ.*

(b) Explain why we can assume that the cut-point Lemma 93 is 9/10, *i.e.* $C_n = \mathcal{L}_{\mathcal{A}_n}(9/10)$. You will need to justify that this cut-point is still isolated with constant threshold δ' independent of n.

Let p be a prime greater than αm for some $\alpha \in \mathbb{N}$ to fix later. For any $i \in \{1, \ldots, \alpha m\}$ and $v \in \mathbb{N}$, consider the language

$$L'_{i,v} = \left\{ w \in A^* : \sum_{t=1}^{m} (i^{t-1} \bmod p)|w|_{a_t} = v \right\}.$$

(c) By using Lemma 93, show that for any i and v, there exists a probabilistic automaton $\mathcal{B}_{i,v}$ with $O\left(\frac{\ln^2 v}{\ln \ln v}\right)$ states that recognizes $L'_{i,v}$ with isolated cut-point 9/10 and isolation threshold at least δ. *Hint: reading one letter in $\mathcal{B}_{i,v}$ corresponds to reading several letters at once in $\mathcal{A}_v$.*

(d) Show that for every $i \in \{1, \ldots, \alpha m\}$, $L_m \subseteq L'_{m,i,v_i}$ for a certain value $v_i \in \{0, \ldots, m^2 p\}$ that you will identify.

We admit the following lemma, whose proof is deferred to question (j)

Lemma 94. *Let $y_1, \ldots, y_m \in \{1, \ldots, \alpha m\}$ be pairwise distinct, then the vectors $z_0, \ldots, z_{m-1}$ defined by*

$$z_j = (y_1^j \bmod p, y_2^j \bmod p, \ldots, y_m^j \bmod p)$$

are linearly independent.

(e) Let $S \subseteq \{1, \ldots, \alpha m\}$. Show that if $|S| \geqslant m$ then $\bigcap_{i \in S} L'_{m,i,v_i} \subseteq L_m$ using Lemma 94.

We now consider the probabilistic automaton $\mathcal{C}_m = \frac{1}{\alpha m} \sum_{i=1}^{\alpha m} \mathcal{B}_{i,v_i}$ where v_i is defined as in question (d).

(f) Show that if $w \in L_m$ then $\mathcal{C}(w) \geqslant 9/10 + \delta$.

(g) Show that if $w \notin L_m$ then $\mathcal{C}(w) \leqslant \frac{9}{10} - \delta + \frac{1/10+\delta}{\alpha}$.

(h) Show that there exists a choice of α, independent of m, such that $\mathcal{C}_m$ recognizes L_m with an isolated cut-point and isolation threshold at least $\delta/2$. Show that $\mathcal{C}_m$ has $O\left(m\frac{\ln^2 n}{\ln\ln n}\right)$ states. *Hint: you can use the fact that we can choose p such that $p = \alpha n + o(\alpha n)$.*

(i) Show that there exists $\delta > 0$ such that for infinitely many n, there exists a regular language recognized by a probabilistic automaton with n states and an isolated cut-point with isolation threshold at least δ, such that the smallest deterministic finite automaton recognizing it has $\Omega(2^{\frac{n\ln\ln n}{\ln n}})$ states. Compare with the result in the course about isolated cut-points.

(j) Let $y_1, \dots, y_m$ be as in Lemma 94 and assume that $z_1, \dots, z_m$ are linearly dependent. Show that there exists $c_0, \dots, c_{m-1}$ not all zero such that $c_0 + c_1 x + \cdots + c_{m-1}x^{m-1} = 0 \bmod p$ for all $x \in \{y_1, \dots, y_m\}$. Prove Lemma 94. *Hint: you can use the fact that a degree d polynomial with integer coefficients has at most d distinct roots modulo any prime number $p > d$.*

B Solutions to exercises

Exercise 2. 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 3. 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 4. 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. Indeed, this is true when w is just one letter, by definition. Let $w \in \Sigma^*$ and $a \in \Sigma$ then any path $q \xrightarrow{wa|x} q'$ is of the form $q \xrightarrow{w|y} q'' \xrightarrow{a|z} q$ where $q'' \in A$, $z=\mu(a)_{q'',q}$ and $x=yz$. Summing over all such paths with fixed q'' gives a probability of $\mu(w)_{q,q''}\mu(a)_{q'',q}$ by induction. Therefore the sum of all paths from q to q' labelled by wa is

$$\sum_{q''\in Q}\mu(w)_{q,q''}\mu(a)_{q'',q'}=(\mu(w)\mu(a))_{q,q'}=\mu(wa)_{q,q'}.$$

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 5. 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 7. Check that $\mathcal{A}(bba)=\frac{1}{12}$ and $\mathcal{A}(abb)=\frac{2}{3}$. 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 $\mathcal{A}((ab)^n b)=\frac{2}{3}$ for every $n\in\mathbb{N}$, thus $(ab)^*b \subseteq \mathcal{L}_{\mathcal{A}}(\lambda)$ for all $\lambda<\frac{2}{3}$.

Exercise 8. 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 9. TODO

Exercise 10. 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 12. 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 15. 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,

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

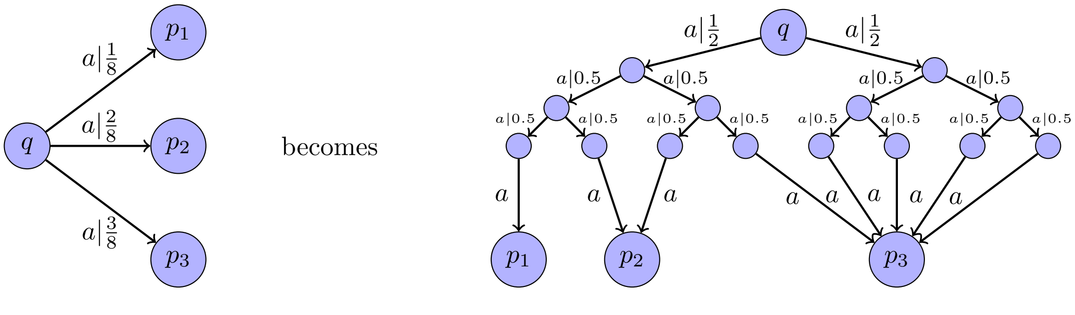
Let $u = x(n_1, \dots, n_k)$ and $w = x(n_{k+1}, \dots, n_\ell)$ then $uw = x(n_1, \dots, n_\ell)$ thus $\mathcal{C}(uw) = \mathcal{C}(u)\mathcal{C}(w)$ by a straightforward calculation. To see the density, fix $\lambda \in (0, 1)$ and let $\mu_\infty = \log \lambda < 0$. Now consider the sequence defined 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 $e^{\mu_k} = \prod_{i=1}^{k}(1 - 2^{-n_i}) = \mathcal{C}(x(n_1, \dots, n_k))$. The proof that $\mathcal{C}$ is universally non-regular is then the same as for Theorem 14.

Exercise 18. 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 11, any deterministic finite automaton that recognizes L must therefore have at least n states.

Exercise 28. Let $\mathcal{A}$ and $\lambda, \mu \in (0, 1)$. There are two cases depending on whether $\lambda \geqslant \mu$ or not. If $\lambda \geqslant \mu$ then we can let $\mathcal{B} = \frac{\mu}{\lambda}\mathcal{A}$ and then for any word w, $\mathcal{B}(w) \geqslant \mu$ if and only if $\mathcal{A}(w) \geqslant \lambda$. Note that we indeed have $\frac{\mu}{\lambda} \in [0, 1]$ so $\mathcal{B}$ is a stochastic automaton. If $\lambda < \mu$, define $\mathcal{B} = (1 - \alpha) + \alpha\mathcal{A}$ where $\alpha = \frac{1-\mu}{1-\lambda} \in [0, 1]$ Then check that for any word w,

$$\mathcal{B}(w) \geqslant \lambda \;\Leftrightarrow\; \alpha\mathcal{A}(w) \geqslant \lambda + \alpha - 1 \;\Leftrightarrow\; \mathcal{A}(w) \geqslant \lambda.$$

Exercise 30. 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 already. We now give the intuition: for each state q and letter a, we will build a tree of height p such that each leaf has probability 2^{-p} to be reached from q after reading a^p. But since p is such that $2^p\mu(a)$ has integer entries, it means that we simply need to choose $(2^p\mu(a))_{q,p}$ leaves for each p and put a transition with probability 1 from this leaf to p. Graphically, for example,



Exercise 33. 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 $\frac{1}{4}$. The same is true for the initial probabilities.

Exercise 38. TODO

Exercise 66. 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

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

Exercise 87. (a) We show L can be recognized by a nondeterministic automaton with one counter. The automaton first counts to n_1 until it reaches the first b. It then guesses the occurrence of a b and starts decreasing the counter for every a until the next b. If the counter reaches 0, the word is accepted.

It is also possible to write a grammar for this language:

$$S \to RB \qquad R \to aRa \,|\, Bb \,|\, b \qquad B \to Ba \,|\, Bb \,|\, b\,.$$

Indeed let $A = \{a, b\}$ then $L(B) = bA^*$ thus $L(R) = \bigcup_{n>0} a^n(L(B)b + b)a^n = \bigcup_{n>0} a^n(bA^*b + b)a^n$ and therefore

$$L(S) = \bigcup_{n>0} a^n(bA^*b + b)a^n bA^*.$$

(b) Since $\mu(a)$ is stochastic, we get that

$$\mu(a)\begin{bmatrix}1\\ \vdots \\ 1\end{bmatrix} = \begin{bmatrix}\mu(a)_{1,1} + \cdots + \mu(a)_{1,d} \\ \vdots \\ \mu(a)_{d,1} + \cdots + \mu(a)_{d,d}\end{bmatrix} = \begin{bmatrix}1\\ \vdots \\ 1\end{bmatrix}$$

which shows that 1 is an eigenvalue. It follows that $P(1) = 0$ since P is the characteristic polynomial and thus $P(\mu(a)) = 0$ by Theorem 63. Unfolding the definition, we get that $c_0 I_d + c_1\mu(a) + \cdots + c_d\mu(a)^d = 0$. Using that μ is a morphism, we then have that

$$\sum_{i=0}^{d} c_i\mathcal{A}(a^i w) = \sum_{i=0}^{d} c_i S\mu(a)^i\mu(w)T = S\left(\sum_{i=0}^{d} c_i\mu(a)^i\right)\mu(w)T = 0.$$

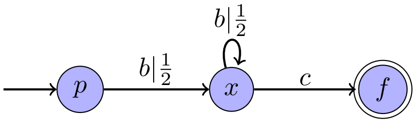
(c) By definition of L, $a^iw \in L$ if and only if $i \in \mathrm{Pos}$. Then

$$\begin{aligned}
\sum_{i=0}^{d} c_i\mathcal{A}(a^iw) &= \sum_{i\in\mathrm{Pos}}^{d} c_i\mathcal{A}(a^iw) + \sum_{i\in\mathrm{NonPos}} c_i\mathcal{A}(a^iw) \\
&> \sum_{i\in\mathrm{Pos}} c_i\lambda + \sum_{i\in\mathrm{NonPos}} c_i\mathcal{A}(a^iw) && \text{since } a^iw \in L \text{ thus } \mathcal{A}(a^iw) > \lambda \\
&> \sum_{i\in\mathrm{Pos}} c_i\lambda + \sum_{i\in\mathrm{NonPos}} c_i\lambda && \text{since } a^iw \notin L \text{ thus } \mathcal{A}(a^iw) \leqslant \lambda \text{ and } c_i \leqslant 0 \\
&= \sum_{i=0}^{d} c_i\lambda = 0.
\end{aligned}$$

This is a contradiction because have seen that $\sum_{i=0}^{d} c_i\mathcal{A}(a^iw) = 0$ and $\sum_{i=0}^{d} c_i\lambda = \left(\sum_{i=0}^{d} c_i\right)\lambda = 0$ thus $0 > 0$.

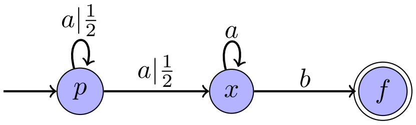
Exercise 88. (a) Clearly $L_1 \cap L_2 \subseteq a^+b^+c^+$. Then $a^nb^mc^p \in L_1 \cap L_2$ if and only if $n = m$ (by virtue of being in L_1) and $m = p$ (by virtue of being in L_2). Thus $L = L_1 \cap L_2$.

(b) Consider the automaton $\mathcal{A}_1$ below, it is substochastic only but can trivially be made stochastic.



It is clear that any word not in the language b^+c has probability of acceptance 0. Furthermore, a direct computation shows that $\mathcal{A}_1(b^mc) = 2^{-m}$ if $m > 0$. We then add a self-loop to p to accept any a and a self-loop to f to accept any c to obtain $\mathcal{B}_1$.

(c) Consider the automaton $\mathcal{A}_2$ below, again it is substochastic only.



Check that any word not in a^+b has probability of acceptance 0. Then observe that $\mathcal{A}_2\left(p \xrightarrow{a^n} p\right) = 2^{-n}$ and by stochasticity $\mathcal{A}_2\left(p \xrightarrow{a^n} x\right) = 1 - \mathcal{A}_2\left(p \xrightarrow{a^n} p\right) = 1 - 2^{-n}$ since after reading a^n, the automaton is either is state p or x. Finally, only state x leads to f when reading b therefore $\mathcal{A}_2(a^nb) = \mathcal{A}_2\left(p \xrightarrow{a^n} x\right) = 1 - 2^{-n}$.

Another clever solution found by a student is the following: take $\mathcal{A}_1$, replace b by a and b by c and take the complement. This automaton satisfies $\mathcal{A}_2(a^nb) = 1 - 2^{-n}$ but has probability of acceptance 1 for the other words. But notice that a^+b^+ is regular so we can build $\mathcal{C}$ such that $\mathcal{C}(w) = 1$ if $w \in a^+b^+$ and 0 otherwise. Then the product of the two automata gives the result.

We then obtain obtain $\mathcal{B}_2$ by making f non-final, adding a state f' with an arrow from f to f' labelled by c, adding a self-loop to f to accept any b and a self-loop to f' to accept any c.

(d) Build $\mathcal{C}_1$ such that $\mathcal{C}_1(w) = \frac{1}{2}\mathcal{B}_1(w) + \frac{1}{2}\mathcal{B}_2(w)$, then

$$\mathcal{C}_1(a^nb^mc^+) = \tfrac{1}{2}\mathcal{B}_1(a^{n-1}b^mc^+) + \tfrac{1}{2}\mathcal{B}_2(a^{n-1}b^mc^+) = \tfrac{1}{2}2^{-m} + \tfrac{1}{2}(1 - 2^{-n}) = \tfrac{1}{2}(1 + 2^{-m} - 2^{-n}).$$

(e) Any word not in $a^+b^+c^+$ has probably of acceptance 0, and by the previous computation, $\mathcal{C}_1(a^nb^mc^+) = \frac{1}{2}$ if and only if $n = m$. Therefore $\mathcal{L}^{=}_{\mathcal{C}_1}(\frac{1}{2}) = \{a^nb^mc^+ : n = m\} = L_1$.

(f) Clearly if $x = y = \frac{1}{2}$ then $\frac{1}{2}x(1-x) + \frac{1}{2}y(1-y) = \frac{1}{4}$. Conversely, observe that $x(1-x) \geqslant \frac{1}{4}$ for all x, and similarly for y. Thus if $\frac{1}{2}x(1-x) + \frac{1}{2}y(1-y) \neq \frac{1}{4}$ then either $x(1-x) > \frac{1}{4}$ or $y(1-y) > \frac{1}{4}$ and thus either $x \neq \frac{1}{2}$ or $y \neq \frac{1}{2}$.

(g) Without loss of generality we can assume that both automata have the same alphabet by taking the intersection (since any word in the resulting intersection must be in both). Write $\mathcal{A} = \langle A, Q_1, S_1, \mu_1, T_1 \rangle$ and $\mathcal{B} = \langle A, Q_2, S_2, \mu_2, T_2 \rangle$, define $\mathcal{C} = \langle A, Q', S', \mu', T' \rangle$ where

$$S = \tfrac{1}{2}\begin{bmatrix} S_1 & S_2 \end{bmatrix}, \qquad \mu(a) = \begin{bmatrix} \mu_1(a) & 0 \\ 0 & \mu_2(a) \end{bmatrix}, \qquad T = \begin{bmatrix} T_1 & T_2 \end{bmatrix}.$$

Note that S is indeed stochastic. Then one easily checks that $\mathcal{C}(w) = S\mu(w)T = \frac{1}{2}(S_1\mu_1(w)T_1 + S_2\mu_2(w)T_2) = \frac{1}{2}(\mathcal{A}(w) + \mathcal{B}(w))$. Using the observation of (f), we get the result.

(h) Similarly to $\mathcal{C}_1$, we can build $\mathcal{C}_2$ such that $\mathcal{L}^{=}_{\mathcal{C}_2}(\frac{1}{2}) = L_2$. Then there exists $\mathcal{C}$ such that $\mathcal{L}^{=}_{\mathcal{C}}(\frac{1}{4}) = \mathcal{L}^{=}_{\mathcal{C}_1}(\frac{1}{2}) \cap \mathcal{L}^{=}_{\mathcal{C}_2}(\frac{1}{2}) = L_1 \cap L_2 = L$.

Exercise 89.

(a) Trivial.

(b) Let $\mathcal{B}$ be as above for L, then build automaton $\mathcal{C}$ such that $\mathcal{C}(w) = \frac{1}{2}\mathcal{A}(w) + \frac{1}{2}\mathcal{B}(w)$. Using that $\mathcal{B}(w) \in \{0,1\}$ we get that

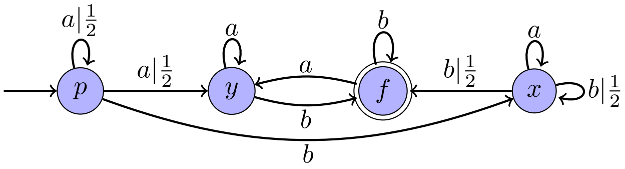
$$\mathcal{C}(w) > \tfrac{\lambda}{2} \Leftrightarrow \mathcal{B}(w) + \mathcal{A}(w) > \lambda \Leftrightarrow \mathcal{B}(w) = 1 \vee \mathcal{A}(w) > \lambda \Leftrightarrow w \in L \vee w \in \mathcal{L}_{\mathcal{A}}(\lambda).$$

Similarly,

$$\mathcal{C}(w) > \tfrac{1+\lambda}{2} \Leftrightarrow \mathcal{B}(w) + \mathcal{A}(w) > 1 + \lambda \Leftrightarrow \mathcal{B}(w) = 1 \wedge \mathcal{A}(w) > \lambda \Leftrightarrow w \in L \wedge w \in \mathcal{L}_{\mathcal{A}}(\lambda).$$

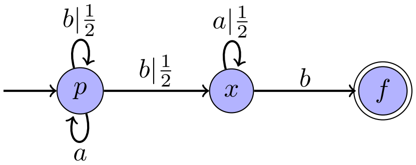
(c) Use that $x(1-x) \geqslant \frac{1}{4}$ if and only if $x = \frac{1}{2}$. See Proposition 32, it works even for non-simple automata with a few tweaks.

(d) Check that the automaton below is stochastic.



Check that $\mathcal{A}(a^{n_1}b \cdots ba^{n_k}b) = 1 - 2^{1-k-n_1}$ if $k \geqslant 1$. Indeed, after reading $a^{n_1}b \cdots ba^{n_k}b$ with $k \geqslant 1$, the automaton can only be in f or x. But since it is stochastic, the probability of acceptance (which is the probability of being in f), is 1 minus the probability of being in x, which is 2^{1-k-n_1}. Any other word has probability of acceptance 0.

(e) Check that the following automaton satisfies $\mathcal{B}(a^{n_1}b \cdots ba^{n_k}b) = 2^{1-k-n_k}$ if $k > 1$.



(f) Build $\mathcal{C}$ such that $\mathcal{C}(w) = \frac{1}{2}(\mathcal{A}(w) + \mathcal{B}(w))$. Then $\mathcal{C}(a^{n_1}b \cdots ba^{n_k}b) = \frac{1}{2}(1 + 2^{1-k-n_1} - 2^{1-k-n_k})$. It follows that $\mathcal{L}^{=}_{\mathcal{C}}(\frac{1}{2}) = L$ and thus L is stochastic.

(g) Trivial.

(h) LA^* is not stochastic, thus concatenation of a stochastic and a regular language is not necessarily stochastic.

(i) Clearly cA^* is regular and since c is not in the alphabet of L, LcA^* is stochastic: the letter c acts as a reset to go from one automaton to another. The morphism $h(a) = a$, $h(b) = b$, $h(c) = \varepsilon$ is such that $h(LcA^*) = LA^*$ which is not stochastic.

(j) Assume that $L^* = \mathcal{L}_{\mathcal{A}}(\lambda)$ for some probabilistic automaton $\mathcal{A} = \langle A, Q, S, \mu, T\rangle$. Let $P(x) = c_0 + c_1x + \cdots + c_dx^n$ be the characteristic polynomial of $\mu(a)$. Recall that by Theorem 63, $P(\mu(a)) = 0$. Since 1 is an eigenvalue of $\mu(a)$, $c_0 + \cdots + c_d = 0$ and for any word w, $\sum_{i=0}^{d} c_i\mathcal{A}(a^iw) = 0$. Define $w = ba^{i_1}b(a^{i_2}b)^2 \cdots (a^{i_k}b)^2$ where $\{i_1, \ldots, i_k\} = \{i : c_i > 0\}$, then $a^iw \in L$ if and only if $i \in \{i_1, \ldots, i_k\}$ but similarly reasoning to Exercise 87 shows that $\sum_{i=0}^{d} c_i\mathcal{A}(a^iw) > \sum_{i=0}^{d} c_i\lambda$ which is absurd.

Exercise 90.

(a) Define $\mathcal{C} = \langle A, Q', S', \mu', T'\rangle$ where

$$S = \tfrac{1}{2}\begin{bmatrix} S_1 & S_2 \end{bmatrix}, \qquad \mu(a) = \begin{bmatrix} \mu_1(a) & 0 \\ 0 & \mu_2(a) \end{bmatrix}, \qquad T = \begin{bmatrix} T_1 & T_2 \end{bmatrix}, \qquad \tilde{T} = \begin{bmatrix} T_1 & -T_2 \end{bmatrix},$$

Note that S is indeed stochastic. Then check that

$$\mathcal{C}(w) = S\mu(w)T = \tfrac{1}{2}(S_1\mu_1(w)T_1 + S_2\mu_2(w)T_2) = \tfrac{1}{2}(\mathcal{A}(w) + \mathcal{B}(w)).$$

On the other hand, we have that

$$S\mu(w)\tilde{T} = \tfrac{1}{2}(S_1\mu_1(w)T_1 - S_2\mu_2(w)T_2) = \tfrac{1}{2}(\mathcal{A}(w) - \mathcal{B}(w))$$

thus $S\mu(w)\tilde{T}$ if and only if $\mathcal{A}(w) = \mathcal{B}(w)$.

(b) Observe[8] that for any word w and letter a, $S\mu(wa) = (S\mu(w))\mu(a) \in V_{|w|}\mu(a)$. Thus $V_{n+1} = V_n \cup \bigcup_{a\in A} V_n\mu(a)$. Therefore if $V_n = V_{n+1}$ for some n, then

$$V_{n+2} = V_{n+1} \cup \bigcup_{a\in A} V_{n+1}\mu(a) = V_n \cup \bigcup_{a\in A} V_n\mu(a) = V_{n+1}.$$

(c) By the previous question, for every n, either $V_n = V_{n+1}$ or $\dim V_n < \dim V_{n+1}$. But $V_n \subseteq \mathbb{R}^{d+d'}$ thus $\dim V_n \leqslant d+d'$. It follows that the sequence $V_n = V_{d+d'}$ for all $n \geqslant d+d'$. But since $V = \bigcup_{n\in\mathbb{N}} V_n$, we get that $V = V_{d+d'}$.

(d) If $\mathcal{A}$ and $\mathcal{B}$ are equivalent then $S\mu(w)\tilde{T} = 0$ for all w. By linearity, and since the $S\mu(w)$ span V, it follows that $v\tilde{T} = 0$ for all $v \in V$. Conversely if $\mathcal{A}$ and $\mathcal{B}$ are not equivalent, then there exists $w \in A^*$ such that $S\mu(w)\tilde{T} \neq 0$. But $S\mu(w) \in V$ so $\exists v \in V$ such that $v\tilde{T} \neq 0$.

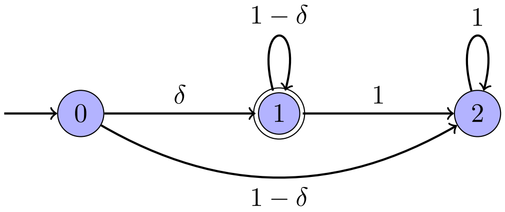
(e) If $\mathcal{A}$ and $\mathcal{B}$ are not equivalent, then there exists $v \in V$ such that $v\tilde{T} \neq 0$. But $V = V_{d+d'} = \operatorname{span}\{S\mu(w) : |w| \leqslant d+d'\}$ thus there must exists w with $|w| \leqslant d+d'$ and $S\mu(w)\tilde{T} \neq 0$ (otherwise by linearity every vector $v \in V$ would satisfy $v\tilde{T} = 0$). Finally $S\mu(w)\tilde{T} \neq 0$ implies that $\mathcal{A}(w) \neq \mathcal{B}(w)$.

(f) To show that the equivalence problem is in coNP, it is equivalent to show that the disequivalence problem (decide whether two automata are *not* equivalent) is in NP. Thanks to the previous question, this is equivalent to searching a word w of linear size $(d+d')$ such that $\mathcal{A}(w) \neq \mathcal{B}(w)$. This can be done in NP by guessing such a word, computing $\mathcal{A}(w)$ and $\mathcal{B}(w)$ and comparing them. Note that we can compute $\mathcal{A}(w)$ in polynomial time because the numbers are rational and their size remains bounded by a polynomial.

Exercise 91.

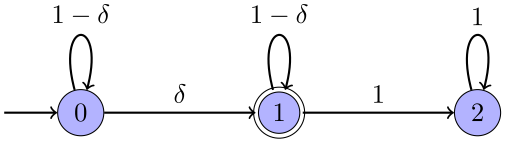
(a) First observe that for any $i < j \leqslant n+1$, $a^i \not\equiv_L a_j$, since for $u = a^{n-i}$ (note that $i \leqslant n$ so $n-i \geqslant 0$), $a^i u \in C_n$ but $a^j u \notin L$. On the other hand, for any $i \geqslant n+1$, $a^{n+1} \equiv_L a^i$ since for all $u \in \Sigma^*$, $a^{n+1}u \notin L$ and $a^i u \notin L$. Hence, there are exactly $n+2$ equivalence classes and by Myhill-Nerode theorem, that's exactly the number of states of a minimal DFA recognizing C_n.

(b) Consider the following automaton:



It is clear that the probability of a^ℓ being accepted is the probability of going from state 0 to state 1 (δ) and the probability of staying $\ell - 1$ times state 1 $(1-\delta)$.

(c) Consider the following automaton:



An accepting run for a^ℓ first stay i times in state 0 then transitions to state 1 and stays $\ell - i - 1$ times in state 1. Therefore the probability is

$$\sum_{i=0}^{\ell-1} (1-\delta)^i \delta (1-\delta)^{n-1-i} = (1-\delta)^{\ell-1}\delta \sum_{i=0}^{\ell-1} 1 = (1-\delta)^{\ell-1}\delta\ell.$$

(d) Let $f(\ell) = (1-\delta)^{\ell-1}\delta\ell$, then $f'(\ell) = \log(1-\delta)(1-\delta)^{\ell-1}\delta\ell + (1-\delta)^{\ell-1}\delta = (1-\delta)^{\ell-1}(1 + \ell\log(1-\delta))$. The sign of the derivative is given by $1 + \ell\log(1-\delta)$, a linear function of ℓ. Since the derivative at 0 is positive and negative at infinity, the maximum is attained once at ℓ such that $1 + \ell\log(1-\delta) = 0$. Hence we choose δ such that the maximum is attained at $\ell = n$, that is $1 + n\log(1-\delta)$, so $\delta = 1 - e^{-1/n}$.

[8] If X is a set and M a matrix, $XM := \{xM : x \in X\}$.

(e) If we let $\delta = 1 - e^{-1/n}$, then $\mathcal{A}(a^\ell)$ is increasing from $\ell = 0$ to $\ell = n$ and decreasing from $\ell = n$ to infinity. Hence if we take $\lambda = \frac{1}{2}(\mathcal{A}(\ell^n) + \max(\mathcal{A}(a^{n-1}), \mathcal{A}(a^{n+1})))$ then $\mathcal{L}_{\mathcal{A}}(\lambda) = C_n$ and by construction λ is isolated. There is a striking comparison between DFAs and PFAs since the former requires n states to recognize C_n whereas the latter only requires 3 independently of n.

(f) First note that

$$\frac{\mathcal{A}(a^{n+1})}{\mathcal{A}(a^{n-1})} = \frac{(1-\delta)^2(n+1)}{n-1} = \frac{e^{-2/n}(n+1)}{n-1} \geqslant \frac{\left(1 - \frac{2}{n} + \frac{4}{n^2}\right)(n+1)}{n-2} = \frac{n - 1 + \frac{2}{n} + \frac{4}{n^2}}{n-1} > 1.$$

Therefore $\mathcal{A}(a^{n+1}) > \mathcal{A}(a^{n-1})$ so the isolation threshold is

$$\begin{aligned}
\tfrac{1}{2}\left(\mathcal{A}(a^n) - \mathcal{A}(a^{n-1})\right) &= \tfrac{1}{2}\left((1-\delta)^{n-1}\delta n - (1-\delta)^n\delta(n+1)\right)\\
&= \left(n - (1-\delta)^2(n+1)\right)(1-\delta)^{n-1}\tfrac{\delta}{2}\\
&= \left(n - e^{-\frac{2}{n}}(n+1)\right)e^{-\frac{n-1}{n}}\frac{1-e^{-\frac{1}{n}}}{2}\\
&\sim \tfrac{1}{2ne} && \text{as } n \to \infty.
\end{aligned}$$

(g) Starting from each initial state q_0^i, the automaton goes in a loop: at each step there is a probability $1-\delta$ to continue, and otherwise we go to a sink state. Hence, the probability to stay in the loop ℓ times is $(1-\delta)^\ell$ and we end up in state $q^i_{\ell \bmod p_i}$ which is final if and only if $\ell \equiv \ell \bmod p_i$. The result follows since each initial state has probability $1/k$. It follows that $\mathcal{B}_{n,k}(a^n) = (1-\delta)^n = \varepsilon$.

(h) If $\ell \geqslant 2n$ then $(1-\delta)^\ell \leqslant (1-\delta)^{2n}$ and hence

$$\mathcal{B}_{n,k}(a^\ell) \leqslant (1-\delta)^\ell < (1-\delta)^{2n} = \varepsilon^2.$$

(i) By the previous questions we have that

$$\mathcal{B}_{n,k}(a^\ell) = \frac{1}{k}\sum_{i=1}^{k} \mathbb{1}\left[n \equiv \ell \bmod p_i\right](1-\delta)^\ell \leqslant \frac{1}{k}\sum_{i=1}^{k}\mathbb{1}\left[n \equiv \ell \bmod p_i\right] = \frac{m}{k}.$$

Assume that $m > r(n)$ and let $X = \{i : n \equiv \ell \bmod p_i\}$. Note that $m = |X|$. By the Chinese remainder theorem, we must have $n \equiv \ell \bmod N$ where $N = \prod_{p\in X} p$. Since $(p_i)_i$ is the sequence of all primes in increasing order, we must have have $N \geqslant \prod_{i=1}^{m} p_i = (\prod_{i=1}^{r(n)} p_i)\prod_{i=r(n)+1}^{m} p_i > \prod_{i=1}^{r(n)} p_i \geqslant n$ by definition of $r(n)$. Since $n \neq \ell$, we must have in particular that $|n - \ell| \geqslant N > n$ so $\ell > 2n$.

(j) By the previous questions, we have $\mathcal{B}_{n,k}(a^n) = \varepsilon$ and $\mathcal{B}_{n,k}(a^\ell) \leqslant \max(\varepsilon^2, r(n)/k)$ for all $\ell \neq n$. If we choose $k = 3r(n)$ (*i.e.* $\alpha = 3$) and $\varepsilon = \frac{1}{2}$ then $\varepsilon^2 < r(n)/k = \frac{1}{3}$ so we can choose $\lambda = \frac{5}{12}$ and we have an isolation threshold of $\frac{1}{12}$.

(k) The automaton $\mathcal{B}_{n,k}$ has $N_n := 1 + \sum_{i=1}^{k} p_i$ states. Since $k = \alpha r(n)$, by the bounds we admitted, we have

$$\begin{aligned}
N_n &\leqslant 1 + \sum_{i=1}^{\alpha r(n)} p_i\\
&\leqslant 1 + \alpha r(n)\ln(\alpha r(n))\\
&= O(r(n)\ln r(n)) && \text{since } r(n) \to \infty\\
&= O\left(\frac{\ln n}{\ln\ln n}\ln\frac{\ln n}{\ln\ln n}\right)\\
&= O\left(\frac{\ln^2 n}{\ln\ln n}\right).
\end{aligned}$$

Exercise 92.

(a) Intuitively, the automaton needs to count each of the m letters up to m, and as soon as one goes above m, we can reject. Hence we need $(m+1)^m$ states to count $\{0,\ldots,m\}^m$, and one extra state to reject.

Let $\equiv_{L_m}$ denote the Myhill-Nerode equivalence relation for L_m. For any $k_1,\ldots,k_m \in \mathbb{N}$, define $w(k_1,\ldots,k_m) = a_1^{k_1}\cdots a_m^{k_m}$. Let $(k_1,\ldots,k_m) \neq (k'_1,\ldots,k'_m) \in \{0,\ldots,m\}^m$, then $w(k_1,\ldots,k_m) \not\equiv_{L_m} w(k'_1,\ldots,k'_m)$. Indeed, on the one hand we have $w(k_1,\dots,k_m)w(m-k_1,\dots,m-k_m) \in L_m$ since each letter a_i appears $k_i + m - k_i = m$ times (note that we used that $k_i \leqslant m$ for $m - k_i$ to be nonnegative). On the other hand, there is i such that $k_i \neq k'_i$ and therefore the word $w(k'_1,\dots,k'_m)w(m-k_1,\dots,m-k_m) \notin L_m$ because it contains $k'_i + m - k_i \neq m$ times the letter a_i. Furthermore, for any $(k_1,\dots,k_m) \in \{0,\dots,m\}^m$, $w(k_1,\dots,k_m) \not\equiv_{L_m} a_1^{m+1}$. Indeed, we have seen that $w(k_1,\dots,k_m)w(m-k_1,\dots,m-k_m) \in L_m$ but $a_1^{m+1}w(m-k_1,\dots,m-k_m) \notin L_m$ because the letter a_1 appears at least $m+1+m-k_1 > m$ times (since $k_1 \leqslant m$).

We have therefore shown that $\equiv_{L_m}$ has *at least* $(m+1)^m + 1$ equivalences classes ($(m+1)^m$ is the cardinal of $\{0,\dots,m\}^m$ and the $+1$ is for the word a_1^{m+1}). By the Myhill-Nerode theorem, any DFA that recognizes L_m has at least that many states On the other hand, it is trivial to build a DFA with $(m+1)^m + 1$ states that recognizes L_m by counting the number of each letters up to m and adding one extra state to reject as soon as a letter appears $> m$ times.

(b) Let $\mathcal{A}_n$ be the automaton of the lemma, then $C_n = \mathcal{L}_{\mathcal{A}_n}(\lambda_n)$ for some λ_n. Clearly $\lambda_n \neq 0, 1$ because C_n is not the empty language, nor the universal one. There are two cases:

- If $\lambda_n \geqslant 9/10$ then we can let $\mathcal{A}'_n = \frac{9}{10\lambda_n}\mathcal{A}_n$ by multiplying the probability of the intial states of $\mathcal{A}_n$ by $\frac{9}{10\lambda_n} \in [0,1]$. We immediately have that $\mathcal{L}_{\mathcal{A}'_n}(9/10) = \mathcal{L}_{\mathcal{A}_n}(\lambda_n) = C_n$. Furthermore, for all $w \in A^*$,

  $$|\mathcal{A}'_n(w) - \tfrac{9}{10}| = |\tfrac{9}{10\lambda_n}\mathcal{A}_n(w) - \tfrac{9}{10}| = \tfrac{9}{10\lambda_n}|\mathcal{A}_n(w) - \lambda_n| \geqslant \tfrac{9\delta}{10\lambda_n} \geqslant \tfrac{9}{10}\delta$$

  since $\lambda_n \leqslant 1$, and is therefore independent of n.
- If $\lambda_n < 9/10$ then we can let $\mathcal{A}'_n = \alpha\mathcal{A}_n + (1-\alpha)$, where $\alpha = \frac{1-9/10}{1-\lambda_n} = \frac{1}{10(1-\lambda_n)}$, by doing a convex combination with the automata that accepts all words. Note there that $\alpha \in [0,1]$ because $0 < \lambda_n < 9/10$. A small computation shows that $\mathcal{L}_{\mathcal{A}'_n}(9/10) = \mathcal{L}_{\mathcal{A}_n}(\lambda_n) = C_n$. Furthermore, for all $w \in A^*$,

  $$|\mathcal{A}'_n(w) - \tfrac{9}{10}| = |\alpha\mathcal{A}_n(w) + (1-\alpha) - \tfrac{9}{10}| = \alpha|\mathcal{A}_n(w) - \lambda_n| \geqslant \alpha\delta \geqslant \tfrac{\delta}{10}$$

  since $\lambda_n \geqslant 0$, and is therefore independent of n.

In summary, we have shown that the isolation threshold is always at least $\frac{\delta}{10}$ which is independent of n.

(c) We consider the automaton $\mathcal{B}_{i,v}$ that has the same states as $\mathcal{A}_v$ (including the same initial and final states). We modify the transitions so that for any pair of states q, q' and letter $a_t \in A$,

$$\mathcal{B}_{i,v}\left(q \xrightarrow{a_t} q'\right) = \mathcal{A}_v\left(q \xrightarrow{x^{\ell_{a_t}}} q'\right) \text{ where } \ell_{a_t} = i^{t-1} \bmod p.$$

Technically, this can be done by defining the transition matrix of a_t in $\mathcal{B}_{i,v}$ to be equal to $\mu^{\ell_{a_t}}$ where μ is the transition matrix of $\mathcal{A}_v$. In other words, reading a_t in $\mathcal{B}_{i,v}$ is like reading $x^{\ell_{a_t}}$ in $\mathcal{A}_v$. Note that ℓ_{a_t} only depends on t (and i is fixed) and is positive (since $p \nmid i^{t-1}$ by primality of p and the fact that $i \leqslant \alpha m < p$) so this is well-defined. Now given a word $w \in A^*$, it follows that the probability of acceptance of w is $\mathcal{B}_{i,v}(w) = \mathcal{A}_v(x^M)$ where

$$M = \sum_{k=1}^{|w|} \ell_{w_k} = \sum_{t=1}^{m} \ell_{a_t}|w|_{a_t} = \sum_{t=1}^{m}(i^{t-1} \bmod p)|w|_{a_t}.$$

Again, technically, this can be shown by using the matrix definition above (call S and T the initial and final vectors of both $\mathcal{A}_v$ and $\mathcal{B}_{i,v}$):

$$\mathcal{B}_{i,v}(w) = S\mu^{\ell_{w_1}}\cdots\mu^{\ell_{w_{|w|}}}T = S\mu^{\sum_{i=1}^{|w|}\ell_{w_i}}T = \mathcal{A}_v(x^M).$$

Finally, we conclude by the fact that $\mathcal{A}_v$ only recognizes those words x^M such that $M = v$. Note that this construction has the same cut-point and isolation threshold as $\mathcal{A}_v$. By question (b), we can assume that the $\mathcal{A}_v$ have cut-point 9/10.

(d) If $w \in L_m$ then $|w|_{a_t} = m$ for all t. Therefore, for all i,

$$\sum_{t=1}^{m}(i^{t-1} \bmod p)|w|_{a_t} = m\sum_{t=1}^{m}(i^{t-1} \bmod p).$$

Hence if we let v_i be the right-hand side, we indeed have that $w \in L'_{i,v_i}$. We finally check that

$$v_i = m\sum_{t=1}^{m}(i^{t-1} \bmod p) \leqslant m^2 p.$$

(e) We will show the result for $|S| = m$. This will imply the result for all $|S| \geqslant m$ since having more elements only makes the intersection smaller. Denote the elements of S by $y_1, \dots, y_m$. If $w \in \bigcap_{i \in S} L'_{m,i,v_i}$ then, by definition, $\sum_{t=1}^m (i^{t-1} \bmod p)|w|_{a_t} = v_i = m \sum_{t=1}^m (i^{t-1} \bmod p)$ for all $i \in S$. Therefore

$$\sum_{t=1}^m (|w|_{a_t} - m)(i^{t-1} \bmod p) = 0$$

for all $i \in S$. Using the notation of Lemma 94, this can be written as $\sum_{t=1}^m (|w|_{a_t} - m)(z_t)_j = 0$ for all $j \in \{1, \dots, m\}$ since $S = \{y_1, \dots, y_m\}$. Therefore $\sum_{t=1}^m (|w|_{a_t} - m)z_t = 0$. But the y_i are pairwise distinct by definition, so by Lemma 94, $z_0, \dots, z_{m-1}$ are linearly independent, hence $|w|_{a_t} - m = 0$ for all t. This shows that $w \in L_m$.

(f) If $w \in L_m$ then $w \in L'_{i,v_i}$ for all $i = 1, \dots, \alpha m$, by question (d). By question (c), $\mathcal{B}_{i,v_i}$ recognizes L'_{i,v_i} so $\mathcal{B}_{i,v_i}(w) \geqslant 9/10 + \delta$ since the cut-point has isolation threshold δ. By construction of $\mathcal{C}_m$, it immediately follows that $\mathcal{C}_m(w) \geqslant 9/10 + \delta$.

(g) Let $w \notin L_m$ and let $S = \{i : w \in L'_{i,v_i}\} \subseteq \{1, \dots, \alpha m\}$. By question (e), $|S| < m$ for otherwise we would have $w \in L_m$. For $i \in S$, we have $\mathcal{B}_{i,v_i} \leqslant 1$ since it is a probability. But since $\mathcal{B}_{i,v_i}$ has isolation threshold δ by question (c), if $i \notin S$, then $\mathcal{B}_{i,v_i} \leqslant 9/10 - \delta$. Therefore,

$$\begin{aligned}
\mathcal{C}(w) &= \frac{1}{\alpha m} \sum_{i=1}^{\alpha m} \mathcal{B}_{i,v_i}(w) \\
&= \frac{1}{\alpha m} \left(\sum_{i \in S} \mathcal{B}_{i,v_i}(w) + \sum_{i \notin S} \mathcal{B}_{i,v_i}(w) \right) \\
&\leqslant \frac{1}{\alpha m} \left(|S| + |\{1, \dots, \alpha m\} \setminus S| (\tfrac{9}{10} - \delta) \right) \\
&\leqslant \frac{1}{\alpha m} \left(m + (\alpha - 1)m(\tfrac{9}{10} - \delta) \right) \\
&= \frac{9}{10} - \delta + \frac{1/10 + \delta}{\alpha}.
\end{aligned}$$

(h) It suffices to choose α such that $-\delta + \frac{1/10+\delta}{\alpha} \leqslant 0$ which is always possible because $\frac{1/10+\delta}{\alpha} \to 0$ as $\alpha \to \infty$. Note that this choice does not depend on m. The number of states of $\mathcal{C}$ is the sum of the number of states of the $\mathcal{B}_{i,v_i}$ for $i = 1, \dots, \alpha m$. Automaton $\mathcal{B}_{i,v_i}$ has as many states as $\mathcal{A}_{v_i}$ which is $O\left(\frac{\ln^2 v_i}{\ln \ln v_i}\right)$. On the other hand, $v_i \leqslant m^2 p$ by question (d). By the distribution of primes, we can always choose $p = \alpha m + o(\alpha m)$ and α was chosen to be a constant that only depends on δ and is independent of m. Therefore $v_i = O(m^3)$ and $\mathcal{C}$ has

$$\alpha m \cdot O\left(\frac{\ln^2 O(m^3)}{\ln \ln O(m^3)} \right) = O\left(m \frac{\ln^2 m}{\ln \ln m} \right)$$

states.

(i) Putting questions (a) and (h) together, for every m, we have found a language L_m recognized by a probabilistic automaton with $n = O\left(m \frac{\ln^2 m}{\ln \ln m}\right)$ states, but whose smallest DFA that recognizes it has $N = (m+1)^m$ states. First observe that $N = (m+1)^m = 2^{O(m \ln m)}$ and that $n \frac{\ln \ln m}{\ln m} = O(m \ln m)$. Furthermore, observe that

$$n = O\left(m \frac{\ln^2 m}{\ln \ln m} \right) \;\Rightarrow\; n = \Omega(m) \text{ and } n = O(m^2) \;\Rightarrow\; \ln n = \Theta(\ln m) \;\Rightarrow\; \ln \ln n = \Theta(\ln \ln m).$$

It follows that

$$N = 2^{O(n \frac{\ln \ln m}{\ln m})} = 2^{O(n \frac{\ln \ln n}{\ln n})}.$$

By comparison, Theorem 16 says that for a cut-point language with isolation threshold δ, the number of states for DFA is bounded by $(1 + \frac{r}{\delta})^{n-1}$ where r is the number of accepting states. Clearly r is smaller than the number of states which is $O(m^2)$, and recall that δ is constant, therefore the upper bound of the theorem is

$$2^{(n-1)\ln(1+\frac{r}{\delta})} = 2^{O(n \ln m)} = 2^{O(n \ln n)}.$$

Therefore there is a still a gap between this upper bound and what we obtain but the two bounds are quite close.

(j) If there are linearly dependent, there exists $c_0, \ldots, c_{m-1}$ such that $c_0 z_0 + \cdots + c_{m-1} z_{m-1} = 0$. Therefore, for all t,

$$0 = (c_0 z_0 + \cdots + c_{m-1} z_{m-1})_t = c_0 y_t^0 + c_1 y_t^1 \cdots + c_{m-1} y_t^{m-1} \bmod p.$$

Since $y_1, \ldots, y_m \in \{1, \ldots, \alpha m\}$ then in particular $y_i < p$ so the y_i are pairwise distinct *modulo* p. Therefore the polynomial $P(x) = c_0 + c_1 x + \cdots + c_{m-1} x^{m-1}$, which has degree at most $m - 1$, has at least m distinct roots modulo p. This is a contradiction with the hint since $p > m$.

Note: the hint can be proven by induction on the degree of P. If P has degree 1 then $P(x) = a + bx$ for some a and $b \neq 0 \bmod p$ (otherwise this is trivial). If x, y are such that $P(x) = P(y) = 0 \bmod p$ then $a + bx = a + by \bmod p$ so $x = y \bmod p$ (b is invertible modulo p, by primality of p) so P has only one root modulo p. Now if P has degree $d > 1$, assume that P has at least one root modulo p (otherwise the result is proved already): $P(x_0) = 0 \bmod p$ for some x_0. Then we can write $P(x) = (x - x_0)Q(x) \bmod p$ for some polynomial Q of degree $d - 1$ (simply consider the expansion of $P(x_0 + x) \bmod p$ to find Q). But now, if y is such that $y \neq x_0 \bmod p$ and $P(y) = 0 \bmod p$ then it must be the case that $Q(y) = 0 \bmod p$ (again by primality of p). By induction, Q has at most $d - 1$ solutions modulo p, therefore there can only be d roots of P modulo p.

This can also be shown more abstractly: any nonzero polynomial $P \in R[x]$ of degree d, where R is an (integral) domain, has at most d roots in R. In fact this is a characterization of integral domains.