Irregular Behaviours for Probabilistic Automata

Nathanaël Fijalkow and Michał Skrzypczak

LIAFA, Paris 7,
University of Warsaw

Abstract. We consider probabilistic automata over finite words. Such an automaton defines the language consisting of the set of words accepted with probability greater than a given threshold. We show the existence of a universally non-regular probabilistic automaton, *i.e.* an automaton such that the language it defines is non-regular for every threshold. As a corollary, we obtain an alternative and very simple proof of the undecidability of determining whether such a language is regular.

1 Introduction

Rabin introduced probabilistic automata in 1963 [Rab63]. They have been studied ever since, with applications to different areas, such as Computational Linguistics and Biology. Despite its simplicity, this computational model is very powerful, and many decision problems for probabilistic automata are known to be undecidable.

A probabilistic automaton defines a probabilistic language through a threshold semantics, as defined by Rabin [Rab63]. The algorithmic properties of these languages are well-understood; we refer to the book of Paz [Paz71] and the survey of Condon [Con01] for a wealth of results about them.

The class of probabilistic languages strictly subsumes the class of regular languages; in this paper, we consider the decision problem of determining whether a probabilistic language is regular. This problem has been considered by Bertoni [Ber75], and proved undecidable. The aim of this paper is to give a different and simple proof of this result.

2 Preliminaries

Let Q be a finite set of states. A distribution over Q is a function $\delta : Q \to [0, 1]$ such that $\sum_{q \in Q} \delta(q) = 1$. We denote $\mathcal{D}(Q)$ the set of distributions over Q.

Definition 1 (Probabilistic automaton). *A probabilistic automaton $\mathcal{A}$ is given by a finite set of states Q, a transition function $\phi : A \to (Q \to \mathcal{D}(Q))$, an initial state $q_0 \in Q$, and a set of final states $F \subseteq Q$.*

All the numbers in the transition function of a probabilistic automaton are assumed to be rational numbers.

In a transition function ϕ, the quantity $\phi(a)(s, t)$ is the probability to go from the state $s \in Q$ to the state $t \in Q$ reading the letter a. A transition function naturally induces a morphism $\phi : A^* \to (Q \to \mathcal{D}(Q))$. We denote $\mathbb{P}_{\mathcal{A}}(s \xrightarrow{w} t)$ the probability to go from a state s to a state t reading w on the automaton $\mathcal{A}$, *i.e.* $\phi(w)(s, t)$.

The *acceptance probability* of a word $w \in A^*$ by $\mathcal{A}$ is $\sum_{t \in F} \phi(w)(q_0, t)$, which we denote $\mathbb{P}_{\mathcal{A}}(w)$.

The following threshold semantics was introduced by Rabin [Rab63].

Definition 2 (Probabilistic language). *Let $\mathcal{A}$ be a probabilistic automaton and $x \in (0, 1)$. This induces the probabilistic language*

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

The *emptiness problem* was considered by Rabin: given a probabilistic automaton $\mathcal{A}$, determine whether $L^{>\frac{1}{2}}(\mathcal{A})$ is non-empty, *i.e.* whether there exists a word w such that $\mathbb{P}_{\mathcal{A}}(w) > \frac{1}{2}$.

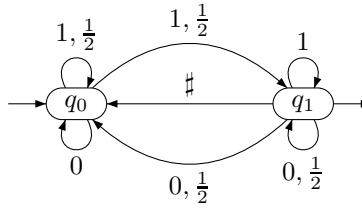**Theorem 1 ([Paz71]).** *The emptiness problem is undecidable.*

A simple undecidability proof was given by Gimbert and Oualhadj in [GO10].

3   A Universally Non-Regular Probabilistic Automaton

Theorem 2. *There is a probabilistic automaton $\mathcal{C}$ such that for every number x in $(0, 1)$, the language $L^{>x}(\mathcal{C})$ is non-regular.*

In the original paper introducing probabilistic automata, Rabin [Rab63] gave an example of a probabilistic automaton $\mathcal{A}$ such that $L^{>x}(\mathcal{A})$ is non-regular, for all irrational numbers x. The alphabet of the Rabin's automaton $\mathcal{A}$ is $\{0, 1\}$. The automaton $\mathcal{A}$ computes the binary decomposition function denoted bin, *i.e.* $\mathbb{P}_{\mathcal{A}}(u) = \mathrm{bin}(u)$, defined by $\mathrm{bin}(a_1 \cdots a_n) = \frac{a_1}{2^n} + \cdots + \frac{a_n}{2^1}$ (*i.e.* $0.a_n \ldots a_1$ in binary). We show that adding one letter and one transition to this probabilistic automaton makes it universally non-regular.

The automaton $\mathcal{C}$ is represented in Figure 1. The alphabet is $C = \{0, 1, \sharp\}$. The only difference between the automaton $\mathcal{A}$ proposed by Rabin [Rab63] and this one is the only transition over $\sharp$. As observed by Rabin, a simple induction shows that for u in $\{0, 1\}^*$, we have $\mathbb{P}_{\mathcal{C}}(u) = \mathrm{bin}(u)$.

Fig. 1. A universally non-regular probabilistic automaton. The initial state is marked by an ingoing arrow and the final state by an outgoing arrow. The first symbol over a transition is a letter (either 0, 1, or ♯). The second symbol (if given) is the probability of this transition. If there is only one symbol then the probability of the transition is 1.

We show that for all numbers x in $(0, 1)$, the language $L^{>x}(\mathcal{C})$ is non-regular. Let u, v in $\{0, 1\}^*$, observe that $\mathbb{P}_{\mathcal{C}}(u \cdot \sharp \cdot v) = \mathrm{bin}(u) \cdot \mathrm{bin}(v)$.

Fix x in $(0, 1)$. For every u, v in $\{0, 1\}^*$ such that $x < \mathrm{bin}(u) < \mathrm{bin}(v)$, there exists w in $\{0, 1\}^*$ such that $u \cdot \sharp \cdot w \notin L^{>x}(\mathcal{C})$ and $v \cdot \sharp \cdot w \in L^{>x}(\mathcal{C})$; it suffices to choose w such that $\mathrm{bin}(w)$ is in $\left(\frac{x}{\mathrm{bin}(v)}, \frac{x}{\mathrm{bin}(u)} \right)$. It follows that the left quotients $u^{-1} \cdot L^{>x}(\mathcal{C})$ and $v^{-1} \cdot L^{>x}(\mathcal{C})$ are distinct, so $L^{>x}(\mathcal{C})$ has an infinite number of pairwise distinct left quotients. By the Myhill-Nerode theorem $L^{>x}(\mathcal{C})$ is not regular.

4   Main Result

Theorem 3. *(Undecidability of the regularity problem) The regularity problem is undecidable for probabilistic automata.*

This result was originally proved in [Ber75].

Roughly speaking, the idea is to use the universally non-regular automaton given in Section 3 to "amplify" an irregular behaviour.

Proof. We construct a reduction from the emptiness problem to the regularity problem. Then, undecidability of the latter follows from Theorem 1.

Let $\mathcal{A}$ be a probabilistic automaton over an alphabet A. We construct a probabilistic automaton $\mathcal{B}$ such that:

$$L^{>\frac{1}{2}}(\mathcal{A}) \text{ is empty} \quad \text{if and only if} \quad L^{>\frac{1}{2}}(\mathcal{B}) \text{ is regular.}$$

The automaton $\mathcal{B}$ is over the alphabet $B = A \uplus C$ where $C = \{0, 1, \sharp\}$, and uses the automaton $\mathcal{C}$ from Section 3. It is obtained as the sequential composition of $\mathcal{A}$ and $\mathcal{C}$: it starts in $\mathcal{A}$ and from every final state of $\mathcal{A}$ moves by $\sharp$ to the initial

state of $\mathcal{C}$. The initial state of $\mathcal{B}$ is the initial state of $\mathcal{A}$, the only final state of $\mathcal{B}$ is the final state of $\mathcal{C}$.

For $u \in A^*$ and $v \in C^*$, we have $\mathbb{P}_{\mathcal{B}}(u \cdot \sharp \cdot v) = \mathbb{P}_{\mathcal{A}}(u) \cdot \mathbb{P}_{\mathcal{C}}(v)$. A word which is not in $A^* \cdot \sharp \cdot C^*$ has no accepting run, so is accepted with probability 0.

- Assume that $L^{>\frac{1}{2}}(\mathcal{A})$ is empty. Thanks to the above observation we have that $L^{>\frac{1}{2}}(\mathcal{B})$ is empty, so in particular it is regular.
- Conversely, assume that $L^{>\frac{1}{2}}(\mathcal{A})$ is non-empty. Let u be a word such that $\mathbb{P}_{\mathcal{A}}(u) > \frac{1}{2}$. Observe that $L^{>\frac{1}{2}}(\mathcal{B}) \cap (u \cdot \sharp \cdot C^*) = u \cdot \sharp \cdot L^{>x}(\mathcal{C})$, where $x = \frac{1}{2 \cdot \mathbb{P}_{\mathcal{A}}(u)}$ is in $(0,1)$. By Theorem 2, the language $L^{>x}(\mathcal{C})$ is non-regular, hence so is $u \cdot \sharp \cdot L^{>x}(\mathcal{C})$, implying that $L^{>\frac{1}{2}}(\mathcal{B})$ is also non-regular.

∎

References

[Ber75] A. Bertoni. Mathematical methods of the theory of stochastic automata. In A. Blikle, editor, *Mathematical Foundations of Computer Science*, volume 28 of *Lecture Notes in Computer Science*, pages 9–22. Springer Berlin Heidelberg, 1975.

[Con01] Anne Condon. *Bounded error probabilistic finite state automata*, volume II, chapter 1, pages 509–532. Handbook on Randomized Computing, Kluwer, 2001.

[GO10] Hugo Gimbert and Youssouf Oualhadj. Probabilistic automata on finite words: Decidable and undecidable problems. In Samson Abramsky, Cyril Gavoille, Claude Kirchner, Friedhelm Meyer auf der Heide, and PaulG. Spirakis, editors, *Automata, Languages and Programming*, volume 6199 of *Lecture Notes in Computer Science*, pages 527–538. Springer Berlin Heidelberg, 2010.

[Paz71] Azaria Paz. *Introduction to probabilistic automata*. Academic Press, 1971.

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