Computing Measure of MSO-Definable Sets of Infinite Trees

Damian Niwiński

Institute of Informatics, University of Warsaw, Poland

Paweł Parys[1]

Institute of Informatics, University of Warsaw, Poland

Michał Skrzypczak

Institute of Informatics, University of Warsaw, Poland

Abstract

This work addresses the problem of computing measures of recognisable sets of infinite trees. An algorithm is provided to compute the probability measure of a tree language recognisable by a nondeterministic parity automaton, or equivalently definable in monadic second-order logic. The measure is the uniform coin-flipping measure. As a tool, we develop a so-called unary μ-calculus, which is a version of μ-calculus with one implicit variable, and allows to compute fixed-point expressions in chain-complete orders.

2012 ACM Subject Classification Theory of computation → Automata over infinite objects; Mathematics of computing → Distribution functions

Keywords and phrases infinite trees, parity automata, coin-flipping measure

Related Version Full paper in preparation

Funding Work supported by the National Science Centre, Poland (grant no. 2021/41/B/ST6/03914).

1 Background

The non-emptiness problem asks if an automaton accepts at least one object. From a logical perspective, it is an instance of the consistency question: does a given specification have a model? Sometimes it is also relevant to ask a quantitative version of this question: whether a *non-negligible* set of models satisfy the specification. When taken to the realm of probability theory, this boils down to estimating the probability that a random object is accepted by a given automaton. In this paper, models under consideration are infinite binary trees labelled by a finite alphabet. Our main problem of interest is the following:

▶ **Problem 1.** *Given a regular tree language L, compute the probability that a randomly generated tree belongs to L.*

In other words, we ask for the probability measure of L. Here, the tree language L might be given by a formula of monadic second-order logic, but for complexity reasons it is more suitable to present it by a tree automaton. By default, the considered measure is the uniform *coin-flipping* measure, where each letter is chosen independently at random; but also more specific measures are of interest. If the computed probability is rational then it can be represented explicitly, but the measure can be irrational (see e.g. [6]), hence may require more complex representation. One of the possible choices, exploited in this paper, is a first-order formula over the field of reals $\mathbb{R}$.

Chen, Dräger, and Kiefer [2] addressed Problem 1 in the case where the tree language L is recognised by a deterministic top-down automaton and the measure is induced by a

[1] The author intending to present the paper on the FICS workshop

stochastic branching process, which then makes also a part of the input data. Their algorithm compares the probability with any given rational number in polynomial space and with 0 or 1 in polynomial time. The limitations of this result come from the deterministic nature of the considered automata: deterministic top-down tree automata are known to have limited expressive power within all regular tree languages.

Michalewski and Mio [6] stated Problem 1 explicitly and solved it for languages L given by so-called *game automata* and the coin-flipping measure. This class of automata subsumes deterministic ones and captures some important examples including the game languages [4], but even here the strength of non-determinism is limited; in particular, the class is not closed under finite union. The algorithm of Michalewski and Mio [6] reduces the problem to computing the value of a Markov branching play, and uses Tarski's decision procedure for the theory of reals. These authors also discover that the measure of a regular tree language can be irrational, which stays in contrast with the case of ω-regular languages, that is, regular languages of infinite words, where the coin-flipping measure is always rational [1].

Another step towards a solution to Problem 1 was made by Przybyłko and Skrzypczak [9], who proposed an algorithm to compute the coin-flipping measure of tree languages definable in fragments of first-order logic, and of tree languages recognised by safety automata, that is, non-deterministic automata with a trivial accepting condition. This result was subsumed by a work of Niwiński, Przybyłko, and Skrzypczak [7], who solve Problem 1 in the case where the language L is recognised by a weak alternating automaton or, equivalently, defined by a formula of weak monadic second-order logic.

An analogue of Problem 1 can be stated for ω-regular languages. As noted by Chen et al. [2], the problem then reduces to a well-known question in verification solved by Courcoubetis and Yannakakis [3] already in the 1990s, namely whether a run of a finite-state Markov chain satisfies an ω-regular property. The algorithm runs in single-exponential time with respect to the automaton (and linear with respect to the Markov chain). A related question was also studied by Staiger [10], who gave an algorithm to compute Hausdorff dimension and Hausdorff measure of a given ω-regular language.

At first sight, one may even wonder if Problem 1 is well-stated, as regular tree languages need not in general be Borel [8]. However, due to Gogacz, Michalewski, Mio, and Skrzypczak [5], we know that regular languages of trees are always universally measurable.

2 Our result

In the present paper, we solve Problem 1 in its full generality. The computed probability is presented by a first-order formula over the field of reals. Combined with the known decision procedures for the theory of reals, this gives the following:

▶ **Theorem 2.** *There is an algorithm that inputs a nondeterministic parity tree automaton $\mathcal{A}$ and a rational number q, and decides if the coin-flipping measure of the language of trees recognised by $\mathcal{A}$ is equal, smaller, or greater than q.*

3 Fixed-point techniques in the paper

It is known that parity automata are equivalent to an appropriate version of μ-calculus on trees. In this paper we indeed work with expressions of μ-calculus.

The key difficulty is that we work in a space of probability distributions $\mathcal{D}$, which is not a complete lattice. The fundamental obstacle lies in the following fact: having two random events A, B we can talk about their union $A \cup B$ and intersection $A \cap B$, but knowing only the probabilities $\mathbb{P}(A), \mathbb{P}(B)$ we cannot say what is the probability of $\mathbb{P}(A \cup B)$ or $\mathbb{P}(A \cap B)$. In consequence, having a monotone function F and an element x_0, we cannot ask for the least fixed point of F above x_0 (such a unique least fixed point need not to exist). But fixed points can be computed in a restricted case, namely if $F(x_0) \geq x_0$; then the unique least fixed point above x_0 is just the limit of repeatedly applying F to x_0. This time there is no obstacle: having an increasing chain of random events $A_1 \subseteq A_2 \subseteq \dots$, the probability of their union $\bigcup_i A_i$ is just the limit of their probabilities.

To be more abstract, we notice that the space $\mathcal{D}$ under consideration is a *chain-complete order*: every chain has a supremum and an infimum. Then, a large part of developments in our paper work for an arbitrary chain-complete order.

In this paper

- we define a so-called *unary μ-calculus*, which is a version of μ-calculus with one implicit variable;
- we provide a type system guarantying that an expression of the unary μ-calculus returns a well-defined result in a chain-complete order;
- we show an expression of the unary μ-calculus (and a type derivation for it) describing acceptance by a parity tree automaton.

References

1 Krishnendu Chatterjee, Marcin Jurdziński, and Thomas A. Henzinger. Quantitative stochastic parity games. In J. Ian Munro, editor, *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2004, New Orleans, Louisiana, USA, January 11-14, 2004*, pages 121–130. SIAM, 2004. URL: `http://dl.acm.org/citation.cfm?id=982792.982808`.

2 Taolue Chen, Klaus Dräger, and Stefan Kiefer. Model checking stochastic branching processes. In Branislav Rovan, Vladimiro Sassone, and Peter Widmayer, editors, *Mathematical Foundations of Computer Science 2012 - 37th International Symposium, MFCS 2012, Bratislava, Slovakia, August 27-31, 2012. Proceedings*, volume 7464 of *Lecture Notes in Computer Science*, pages 271–282. Springer, 2012. `doi:10.1007/978-3-642-32589-2_26`.

3 Costas Courcoubetis and Mihalis Yannakakis. The complexity of probabilistic verification. *J. ACM*, 42(4):857–907, 1995. `doi:10.1145/210332.210339`.

4 Alessandro Facchini, Filip Murlak, and Michał Skrzypczak. Index problems for game automata. *ACM Trans. Comput. Log.*, 17(4):24, 2016. `doi:10.1145/2946800`.

5 Tomasz Gogacz, Henryk Michalewski, Matteo Mio, and Michał Skrzypczak. Measure properties of regular sets of trees. *Inf. Comput.*, 256:108–130, 2017. `doi:10.1016/j.ic.2017.04.012`.

6 Henryk Michalewski and Matteo Mio. On the problem of computing the probability of regular sets of trees. In Prahladh Harsha and G. Ramalingam, editors, *35th IARCS Annual Conference on Foundation of Software Technology and Theoretical Computer Science, FSTTCS 2015, December 16-18, 2015, Bangalore, India*, volume 45 of *LIPIcs*, pages 489–502. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2015. `doi:10.4230/LIPIcs.FSTTCS.2015.489`.

7 Damian Niwiński, Marcin Przybyłko, and Michał Skrzypczak. Computing measures of Weak-MSO definable sets of trees. In Artur Czumaj, Anuj Dawar, and Emanuela Merelli, editors, *47th International Colloquium on Automata, Languages, and Programming, ICALP 2020, July 8-11, 2020, Saarbrücken, Germany (Virtual Conference)*, volume 168 of *LIPIcs*, pages 136:1–136:18. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020. `doi:10.4230/LIPIcs.ICALP.2020.136`.

8 Damian Niwiński and Igor Walukiewicz. A gap property of deterministic tree languages. *Theor. Comput. Sci.*, 303(1):215–231, 2003. `doi:10.1016/S0304-3975(02)00452-8`.

9 Marcin Przybyłko and Michał Skrzypczak. The uniform measure of simple regular sets of infinite trees. *Inf. Comput.*, 278:104595, 2021. `doi:10.1016/j.ic.2020.104595`.

10 Ludwig Staiger. The Hausdorff measure of regular omega-languages is computable. *Bull. EATCS*, 66:178–182, 1998.