L3 Internship Proposal: Tractable grammars on probabilistic words

Antoine Amarilli, Mikaël Monet, Sylvain Salvati

This internship is about studying a problem in formal language theory, namely, the computational complexity of *weighted counting for context-free languages*.

Fix the alphabet $\Sigma = \{0, 1\}$. A *probabilistic word* is just a finite sequence $\pi = p_1, \ldots, p_n$ of probability values in $[0, 1]$. We see a probabilistic word of length n as a probability distribution on the set Σ^n of words of length n: for $w = b_1 \cdots b_n$ in Σ^n, the probability $\pi(w)$ of w according to π is the product for $1 \leq i \leq n$ of p_i if $b_i = 1$ and $1 - p_i$ if $b_i = 0$. For instance, if $b_i = 1/2$ for all i, then all 2^n words of Σ^n have probability $1/2^n$. Likewise, if $p_i \in \{0, 1\}$ for each $1 \leq i \leq n$, then all words of Σ^n have probability zero except the one which is identical to π.

The *weighted counting problem* that we study is the following. Fix a language L over Σ, for instance given by a regular expression or a context-free grammar. The *weighted counting problem for* L, written $P(L)$ is intuitively defined as follows: we are given a probabilistic word π and want to compute the probability to obtain a word of L in the distribution represented by π. Formally:

- Input: a probabilistic word $\pi = p_1, \ldots, p_n$
- Output: the total probability according to π of the words of Σ^n that are in L, i.e., $\sum_{w \in L \cap \Sigma^n} \pi(w)$.

This problem can be tractable (in polynomial time) for some languages (for instance $L = \Sigma^*$), and it can be shown to be in PTIME for any regular language L (left as an exercise to the reader). However, this problem can be shown to be intractable, specifically #P-hard[1], for some *context-free languages*. In other words, there is a language L defined by a fixed context-free grammar for which the problem $P(L)$ is #P-hard.

The question of the internship is to understand for which context-free languages L the problem $P(L)$ is tractable, and for which it is #P-hard. The goal is to show a dichotomy characterizing the tractable languages, or partial results towards a dichotomy. One first direction would be to study which results can be shown in the restricted setting of linear languages[2]. Another question is to understand the connections between this problem and existing work on the topic of counting the number of strings accepted by context-free languages [BGS91], which correspond to the case of probabilistic words with all probabilities being 1/2.

Supervision and environment. This L3 internship proposal is intended for a duration of 2–3 months, starting some time between January and May 2025. It will take place in the LINKS team of the Inria center at University of Lille, in the North of France. The LINKS team focuses on logics, algorithms, formal language theory, and database theory, and offers a dynamic environment for research on these topics. The internship will be co-supervised by Antoine Amarilli[3] (Advanced Research Position at Inria), Mikaël Monet[4] (Chargé de recherche Inria), and Sylvain Salvati[5] (Professor).

This internship focuses on mathematical research in theoretical computer science, i.e., the object of the internship is to study theoretical problems, prove mathematical results, and write proofs. No practical implementation work is intended. Candidates are expected to have a solid background in theoretical computer science, experience in mathematical writing, and willingness to learn about concepts in formal language theory and counting complexity.

Applications should be sent by email to the three supervisors: `a3nm@a3nm.net`, `mikael.monet@inria.fr`, and `sylvain.salvati@univ-lille.fr`.

References

[BGS91] Alberto Bertoni, Massimiliano Goldwurm, and Nicoletta Sabadini. The complexity of computing the number of strings of given length in context-free languages. *TCS*, 86(2), 1991.

[1] https://en.wikipedia.org/wiki/%E2%99%AFP-complete
[2] https://en.wikipedia.org/wiki/Linear_grammar
[3] https://a3nm.net/
[4] https://mikael-monet.net/
[5] https://pro.univ-lille.fr/sylvain-salvati