Randomness and Non-Uniformity in Descriptive Complexity

Sujet de thèse

Damiano Mazza
CNRS, LIPN, Université Sorbonne Paris Nord

Background

Descriptive complexity. The goal of computational complexity is to understand how hard it is to solve decision problems (which are defined as subsets of $\{0,1\}^*$). Descriptive complexity aims at understanding, instead, how hard it is to *describe* such problems. In other words, instead of asking "How many computational resources does a Turing machine need to solve this problem?", we ask "What logical primitives (operators, quantifiers, etc.) are necessary for a formula to describe this problem?".

That these two questions are in fact closely related was first realized by Fagin, who proved that NP corresponds to problems definable in second-order existential logic [Fag74]. After this initial result, descriptive complexity grew to become a very rich field, providing logical characterizations of nearly every natural complexity class [Imm99].

To have an idea of how this works, observe that the set $\{0,1\}^*$ of binary strings may be seen as the set of finite models (modulo isomorphism) of the theory Str with one binary relation symbol $\leq$, axiomatized to be a total order, and a unary relation symbol isOne. For example, the model $\{0 < 1 < 2\}$ in which $\mathsf{isOne} = \{1,2\}$ corresponds to the string 011. In this setting, a logical formula on the language of Str induces a decision problem (*i.e.*, a subset of $\{0,1\}^*$), namely the set of its finite models. For example, the formula

$$\forall x.\forall y.\mathsf{Plus}(x,y,\mathsf{Max}) \Rightarrow (\mathsf{isOne}(x) \Leftrightarrow \mathsf{isOne}(y)),$$

where we stipulate that $\mathsf{Plus}(x,y,z)$ holds exactly when $x+y=z$ and that the constant Max is the maximum element of the finite model, describes the set of palindromes. This shows, in particular, that the palindromes problem may be described in first-order logic, using only universal quantifiers and some basic arithmetic symbols.

Randomness and non-uniformity. Randomized algorithms have become increasingly important in computational complexity, especially (but not exclusively) because of their applications to cryptography. The class BPP, which is composed of those problems solvable in polynomial time by a probabilistic machine with bounded error, is by many considered as legitimate as P as the class of "feasible" problems and is the subject of deep conjectures (the most prominent of which states that, in fact, $\mathsf{P} = \mathsf{BPP}$).

On the other hand, non-uniform computation has been, since the 1970s, an important topic in complexity because of its highly algebraic and combinatorial nature, which bypasses some apparent difficulties in dealing with computability assumptions. A possible, very useful

definition is in terms of families of *Boolean circuits*: a single Boolean circuit with n inputs decides only a subset of $\{0,1\}^n$, so in order to decide a problem we take a family of circuits $(C_n)_{n\in\mathbb{N}}$, where each C_n has n inputs and is in charge of dealing with strings of length n. Notice that we do not ask any relationship whatsoever between the different circuits: the family may very well be uncomputable. This is non-uniformity. Of course, *any* problem becomes decidable in this way, so we need to put some constraints on the families we consider. A typical constraint is to ask that the size of C_n grows polynomially in n. This defines a class called P/poly, a central object of complexity theory.

Non-uniformity is quite powerful: for example, a fundamental result of complexity theory states that BPP $\subseteq$ P/poly [Adl78], that is, randomness may be converted to non-uniformity. Nevertheless, it is still widely believed that NP $\not\subset$ P/poly, a conjecture which implies P $\neq$ NP and is therefore considered to be out of reach of current techniques. At any rate, non-uniform complexity is a subject of active research, especially for what concerns "small" circuit classes (*i.e.*, which are conjectured to be well below P/poly).

Objectives and Methodology

In spite of their importance, randomness and non-uniformity are currently completely disregarded by descriptive complexity: no characterization of any probabilistic class is known, and no available result gives a descriptive perspective on non-uniformity.

In broad terms, the objective of this thesis is to develop a descriptive complexity approach to randomness and non-uniformity. Methodologically, there are two starting points:

- Davoli's master thesis work with Dal Lago, in which they provide for the first time a logical account, in terms of an arithmetic theory, of probabilistic computation [Dav22].
- My own recent work on categorical descriptive complexity, which has in turn two key features:
  - generalizing formulas to functors, it allows to describe "generalized problems" with properties which usual problems cannot (or are not know to) have. For example, one can find "generalized problems" which are complete for classes not known to have complete problems, such as BPP; this is what would make possible, at least theoretically, the characterization of such classes, which is currently out of reach for traditional descriptive complexity.
  - It provides a mathematically sophisticated approach to non-uniformity, in which the problems defined by circuit familes may be seen as certain colimits.

Davoli's work suggests that randomness may be successfully incorporated into a logical framework, provided that the semantics (*i.e.*, the model theory) is suitably adjusted. Such adjustments should be compatible with the categorical framework I am developing (which is based on categorical logic, where one is free to consider models taken elsewhere than in the category of sets), suggesting the possibility of blending the two approaches.

More concretely, the thesis will address the following questions:

1. in (categorical) descriptive complexity, it is very easy to describe non-deterministic polytime Turing machines and, therefore, to characterize the class NP. The randomized machines used to define probabilistic classes are all based on non-deterministic polytime machines, so the first step will be to understand how probabilities may be added in the logical framework.

2. Once the above point is developed, it should be possible to immediately attempt a characterization of one-sided-error probabilistic classes (RP), as well as majority-based classes (PP).

3. In a later stage, we will look for the possibility of describing two-sided-error classes, most notably BPP.

4. For all of the above, we will systematically look for the existence of "generalized problems" which are complete for the various classes (in the case of PP, "generalized problems" are not needed, because "usual" complete problems are already known).

5. The colimits describing non-uniform computation, as mentioned above, always exist, but it is not clear whether they have the property of being *locally representable*, which makes them more easily manageable. This will be studied, especially by seeking sufficient conditions ensuring local representability.

6. Once the basic concepts have been imported in the theory, the relationship between randomness and non-uniformity will be studied and understood through the lense of (categorical) descriptive complexity. In particular, a natural goal is to reprove Adleman's theorem ($\mathsf{BPP} \subseteq \mathsf{P/poly}$) within the theory.

It is important to stress that, although the logical and computer-science-oriented aspects are emphasized in this thesis, using and applying the functorial framework will require the development of certain basic knowledge of more mathematically-oriented topics, like algebra and category theory. This is, however, to a level which will be readily achievable during the thesis by a student with excellent theoretical background such as Davoli.

Plan. We expect the work to be developed along roughly the following plan:
Year 1: introduction to categorical language and functorial descriptive complexity; start working on objective (1) and (5), which are independent.
Year 2: finalize objective (1) and work on objectives (2) and (3), while systematically considering question (4); finalize objective (5).
Year 3: work on objective (6) and finalize any remaining objective; write the manuscript.

References

[Adl78] Leonard Adleman. Two theorems on random polynomial time. In *Proceedings of FOCS*, 1978.

[Dav22] Davide Davoli. *A Bounded Arithmetic for Probabilistic Polynomial Time*. Master thesis, Universitá di Bologna, 2022.

[Fag74] Ronald Fagin. Generalized first-order spectra and polynomial-time recognizable sets. In *Proceedings of SIAM–AMS 1973*, 1974.

[Imm99] Neil Immerman. *Descriptive Complexity*. Springer, 1999.