Theoretical Computer Science 570 (2015) 70–101

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Associative and commutative tree representations for Boolean functions [☆]

Antoine Genitrini [a,b,*], Bernhard Gittenberger [c], Veronika Kraus [c], Cécile Mailler [d]

[a] Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, F-75005, Paris, France
[b] CNRS, UMR 7606, LIP6, F-75005, Paris, France
[c] Technische Universität Wien, Wiedner Hauptstrasse 8-10/104, A-1040 Wien, Austria
[d] Department of Mathematical Sciences, University of Bath, BA2 7AY Bath, UK

ARTICLE INFO

Article history:
Received 1 October 2013
Received in revised form 19 December 2014
Accepted 31 December 2014
Available online 7 January 2015
Communicated by O. Watanabe

Keywords:
Boolean functions
Probability distribution
Random Boolean formulas
Random trees
Asymptotic ratio
Analytic combinatorics

ABSTRACT

Since the 1990s, the probability distribution on Boolean functions, induced by some random formulas built upon the connectives *And* and *Or*, has been intensively studied. These formulas rely on plane binary trees. We extend all the results, in particular the relation between the probability and the complexity of a function, to more general formula structures: non-binary or non-plane trees. These formulas satisfy the natural properties of associativity and commutativity.

1. Introduction

Since the 1980s, several papers have focused on the probability distribution on Boolean functions induced by random Boolean formulas. We first mention the result of Valiant [25] who constructs a small formula that with high probability represents the Boolean function `Majority`. The method he developed, often called the probabilistic amplification, has then been adapted to build other Boolean functions [2,7,18,24]. The main goal of such studies was to build explicitly a small formula (of polynomial size in the number of variables) for important Boolean functions. All these results are based on very constrained Boolean formulas: the formulas, seen as trees, are balanced and the labelling of the internal nodes is very regular. Later, some results on larger classes of formulas have been obtained, still based on the approach of amplification: [3,10,5].

During the 90s, other authors [22,20] aimed at defining some "natural" probability distributions for Boolean functions based on large random Boolean formulas seen as trees. In these papers no structural constraints are imposed. The internal nodes of the trees are usually labelled by two connectives *And* and *Or* and the external nodes by symbols taken from a fixed

☆ This research was partially supported by the ANR project *BOOLE*, 09BLAN0011, and by the P.H.C. Amadeus projects F03/2010 and F03/2013. The Austrian authors' work has been supported by *FWF (Austrian Science Foundation), National Research Area S9600, grant S9604* and *ÖAD, grants F03/2010 and F03/2013*.

* Corresponding author.
E-mail addresses: antoine.genitrini@lip6.fr (A. Genitrini), gittenberger@dmg.tuwien.ac.at (B. Gittenberger), veronika.kraus@gmx.at (V. Kraus), c.mailler@bath.ac.uk (C. Mailler).

http://dx.doi.org/10.1016/j.tcs.2014.12.025
0304-3975/

set of literals. The support of the resulting probability distribution on Boolean functions is the whole set of functions and no more a distribution concentrated on a small subset of functions, like the one of Valiant.

Other papers appeared during the last 15 years: their central goal was to obtain quantitative results from a logic point of view. The first result in this direction has been obtained in 2000 by Moczurad et al. [21]. It is based on formulas built with the single connective *Implication* and is dedicated to the study of the quantitative ratio of intuitionistic logic within classical logic. The paper presents exact results for the logics induced by a very small number of variables and states a conjecture on the asymptotic behaviour of the ratios of both logics, when the number of variables tends to infinity. The conjecture has then been proved in [13].

This model, based on a single connective, has then been studied in detail in order to understand the behaviour of the whole probability distribution on Boolean functions. The first results on tautologies [13] have proven to be crucial for the study of the whole distribution. The complete study by Fournier et al. [11,12] has linked together the complexity and the probability of a function.

In parallel, models based on Boolean formulas built with two connectives, *And* and *Or*, have been studied. First, Lefmann and Savický [20] established some bounds for the probability of a function, bounds that are linked to the complexity of the functions. These bounds have been improved by Chauvin et al. [4] where other models based on Galton–Watson branching processes have been studied as well. Then Kozik [19] has developed a powerful tool based on pattern languages that allows to classify and count large trees according to some structural constraints. Using this tool he managed to compute the asymptotic order of the probability of a function. Both implicational and *And*/*Or* models exhibit the same relation between complexity and probability and, though the way to prove it is not at all the same, the same paradigm is underlying. Namely, almost all trees computing a fixed function can be constructed in a particular way: Start with a minimal tree and attach a large tree such that the function computed is not changed.

As pointed out by Gardy [14] the results discussed above have a fundamental weakness. All models use plane binary trees as their underlying tree model. This implies that formulas which should be considered the same are counted separately in the models: Indeed, since *And* and *Or* are commutative and associative operations, the underlying trees should neither be plane nor binary. Similarly, plane trees are not appropriate for the implication model since the premises of an implicational formula can be interchanged without changing the function. This issue was addressed in [16] where a model of *Implication* which is insensitive to the commutation of premises has been studied.

This paper aims at a thorough analysis of the relation between complexity and probability of a Boolean function given by a large random *And*/*Or*-formula as well as at the study of the influence of associativity and commutativity on the behaviour of the model. Thus we will present results for four models: Formulas with or without associativity and with or without commutativity of the connectives. We will derive precise asymptotic results (including numerical constants) for the probability of functions of smallest complexity (literals and constants) as well as the asymptotic behaviour for functions of higher complexity. The paradigm mentioned above (a typical tree is a minimal tree expanded once) still holds for all our models. In this paper we also analyse where such expansions can take place which enables us to derive bounds for the multiplicative constants of the asymptotic expressions. Our method would allow also the precise computation of the constants in this case, though the derivation would be much more involved. The analysis will utilize and extend Kozik's theory of pattern languages [19]. This method was designed for and successfully applied to the binary plane case. However, the non-binary cases require a modification of the method and in the non-plane case there are no exact formulas available anymore, but only approximate ones. We have to utilize Pólya's enumeration theory which makes the analysis of the models technically more difficult. Moreover, we have to work with more general pattern languages and introduce semi-planar structures (which we call mobiles) in Sections 3.3 and 3.4. Unfortunately, these pattern languages are not subcritical anymore which was a crucial property in the analysis of the binary plane case. For a global reference on non-plane tree-structures and the techniques that are necessary the reader can refer to Drmota's book [6].

The results for the first of these models (neither associative nor commutative) are partially known [26,19]. However, for comparison and in order to put all the models under a common roof, we will include this model as well.

The paper is organized as follows. Section 2 is dedicated to introduce the whole context of Boolean formulas seen as trees and presents the models and probability distributions we will study. Then the complete study of the distributions is presented. It is decomposed in three sections: Sections 3, 4 and 5. Each section is presented is the same way. First we present an overview of the corresponding result of Kozik in the case of binary and plane formulas in order to point out precisely the technical arguments that must be adapted to address our context of non-binary or non-plane formulas. And then we prove the generalised versions of the key tools we need. We will prove that the probability of a given Boolean function is asymptotically proportional to a power of the number of allowed variables with exponent related to the complexity. The results are stated in Section 5, Theorems 5.3, 5.8 and 5.9. Moreover, we derive narrow bounds for proportionality factors for the probability of any Boolean function and the proofs of these theorems exhibit what most of the formulas for a fixed function look like.

2. Associative and commutative trees: definitions, generating functions

Kozik [19] has shown that in binary plane trees the order of magnitude of the limiting probability of a given Boolean function is related to its complexity. We generalise this result and therefore define the complexity of a function by the following: