THE RELATION BETWEEN TREE SIZE COMPLEXITY AND PROBABILITY FOR BOOLEAN FUNCTIONS GENERATED BY UNIFORM RANDOM TREES*

VERONIKA DAXNER, ANTOINE GENITRINI†, BERNHARD GITTENBERGER‡, AND CÉCILE MAILLER$^{\mathcal{P}}$

Abstract. An associative Boolean tree is a plane rooted tree whose internal nodes are labelled by and or or and whose leaves are labelled by literals taken from a fixed set of variables and their negations. We study the distribution induced on the set of Boolean functions by the uniform distribution on the set of associative trees of a large fixed size, where the size of a tree is defined as the *number of its nodes*. Using analytic combinatorics, we prove a relation between the probability of a given function and its *tree size complexity*.

1. INTRODUCTION

A Boolean function of n variables is a mapping from $\{\texttt{True}, \texttt{False}\}^n$ onto $\{\texttt{True}, \texttt{False}\}$. The interest in these objects dates back to the 1940s when Riordan and Shannon [26, 29] discovered the so-called Shannon effect: the uniformly random n variables Boolean function has asymptotically almost surely exponential *complexity* when n goes to infinity. Since then numerous papers have been devoted to developing a better understanding of various aspects of Boolean functions. Concerning random Boolean functions and the Shannon effect, further investigations were carried out by Lupanov [23, 24] and a proof based on simple combinatorial counting arguments is presented in Flajolet and Sedgewick's book [9]. All these results concern the uniform probability distribution on the set of Boolean functions in n variables.

In the last two decades people became interested in non-uniform distributions. A natural way to pick a Boolean function at random is to pick a Boolean formula at random and look at the function it represents. It is convenient to see Boolean formulas as rooted plane trees whose internal nodes are labelled by logical connectives like $\wedge$, $\vee$ or $\implies$ (being respectively the conjunction, disjunction and implication operators) and whose leaves are labelled by (possibly negated) variables $\{x_1, \ldots, x_n\}$ (where n is a fixed integer).

The first efforts in generating a random Boolean function via a random Boolean formula/tree go back to Paris *et al.* [25] and Lefmann and Savický [22] where the authors pick a tree uniformly at random among all binary and/or (meaning that only the binary connectives $\wedge$ and $\vee$ are allowed) trees having m leaves. Lefmann and Savický [22] showed the existence of a limiting distribution when the size m of the tree (i.e. its number of leaves) tends to infinity and also bounded the probability of a Boolean function in terms of its complexity, the complexity of a Boolean function being the size of the smallest trees representing it. The existence of a limit distribution when m goes to infinity was shown independently for nonbinary trees by Woods [31]. Woods' method is basically a special case of a more general result, the Drmota-Lalley-Woods theorem originating in the works [5, 21, 31]. See Flajolet and Sedgewick [9] for an easily accessible formulation,

2010 *Mathematics Subject Classification.* 05A15, 06E30, 60C05.

Key words and phrases. Boolean functions; Probability distribution; Random Boolean formulas; Tree size complexity; Formula size complexity; Analytic combinatorics.

*) This research was partially supported by the P.H.C. AMADEUS project 29281NE, the ÖAD project F03/2013 and the FWF grant SFB F50-03.

† Sorbonne Universités, UPMC Univ Paris 06, CNRS, LIP6 UMR 7606, 4 place Jussieu 75005 Paris. Antoine.Genitrini@lip6.fr.

‡ Technische Universität Wien, Wiedner Hauptstrasse 8-10/104, A-1040 Wien, Austria. gittenberger@dmg.tuwien.ac.at.

$^{\mathcal{P}}$ Department of Mathematical Sciences, University of Bath, BA2 7AY Bath, UK. c.mailler@bath.ac.uk.

Drmota [6] for a detailed discussion, and Fournier *et al.* [12] for an application in the context of Boolean formulas. The bounds of Lefmann and Savický were later refined in Chauvin *et al.* [4] by methods from analytic combinatorics. A survey of this topic was written by Gardy [13].

Different but somehow related problems in the framework of balanced trees have been pursued by Valiant [30] who wanted to generate a particular Boolean function with high probability. His results were extended in Boppana [1] and Gupta and Mahajan [19]. The existence of a limit distribution for balanced and/or trees can be found in Fournier, Gardy and Genitrini [10]. The influence of different connectives was studied in Savický [27] and Brodsky and Pippenger [2], but under different distributions.

The study of the relation between the probability of a given Boolean function and its complexity under models similar to the Lefmann-Savický model was resumed recently by Kozik [20] (and by Fournier *et al.* [11, 12] for the implicational model) and later refined and extended in [16]. It was proved that if $p_{n,m}(f)$ is the probability that a uniform binary and/or tree of size m (and on n variables) calculates f, then

$$\lim_{m\to\infty} p_{n,m}(f) =: p_n(f) \sim \frac{\lambda_f}{n^{\ell(f)+1}}, \text{ when } n \to \infty, \tag{1}$$

where $\ell(f)$ is the complexity of the Boolean function f.

The drawback of the models based on plane binary trees is that basic algebraic properties like commutativity and associativity of the connectives are not incorporated into the model. The papers [16, 15] extend the binary plane models to more general tree classes. For instance, due to the associativity of the connectives $\wedge$ and $\vee$, it is very natural to consider nonbinary trees, but in that case there is no justification why the size of an and/or tree should be its number of leaves (as considered in [16]) and not its number of internal nodes or its total number of nodes. And since the complexity of a Boolean function is defined as the size of the smallest trees calculating this function, changing the notion of size changes the notion of complexity.

In computer science, the *formula size complexity* (often called formula complexity or only complexity) is an important quantity in the investigation of Boolean functions. When looking at the tree representation of a Boolean formula, the formula size is the number of leaves of the tree. It is the size notion and thus the complexity notion used in the and/or trees literature mentioned above [25, 22, 4, 20]. This notion of size is natural in the binary connectives context (i.e. with gates of fan-in 2) because, in that context, the number of connectives is the number of leaves shifted by 1. However, when one turns to Boolean circuits, and consequently to (computational) complexity theory questions (see e.g. Graham, Grötschel and Lovász [18, Chapter 40]), the natural notion of size in this context is the number of connectives (or gates).

In this article, we consider nonbinary and/or trees. The leaves are not shared like in circuits, and thus both their number and the number of connectives are important when defining the size of a Boolean function. Moreover, from a computer science point of view, the size of the storage is a function of the number of all vertices. We thus define the *tree size* of a Boolean formula as the total number of nodes of its tree representation. The complexity of Boolean functions associated to this new notion of size is called the *tree size complexity*, as opposed to the *formula size complexity* used in the literature.

In this article, we show that the typical uniform and/or tree of size m changes drastically when changing the size notion. This comes from the two successive limits taken in our set-up and in the literature: in Equation (1), the size of the trees first goes to infinity and then, we let the number n of variables labelling the trees go to infinity. The typical and/or tree of size m in the tree size model depends on the number of variables n: the larger n is, the more likely it is for the typical tree to have many leaves, whereas in the formula size model, the shape of the typical tree of size m does not depend on n. This difference between the two models is the reason why we need to develop a whole new approach in this article since this difference of typical shapes make the proofs developed in the formula size literature collapse.

However, it eventually turns out that the two distributions induced on the set of Boolean functions by the tree size model and the formula size model very well fall under the same paradigms. As in the formula size model (see Equation (1)), we can prove that there is a strong relation between

the (tree size) complexity and the probability of a function. More precisely, if we pick uniformly at random an and/or tree of tree size m (labelled on n variables) and denote by $\mathbb{P}_{m,n}(f)$ the probability that this tree represents the Boolean function f, then we prove in this article that

$$\lim_{m\to\infty} \mathbb{P}_{m,n}(f) =: \mathbb{P}_n(f) \sim \frac{\lambda_f}{n^{L(f)}}, \tag{2}$$

where $L(f)$ is the tree size complexity. This result is very close to Equation (1), but the difference in the exponent of n actually makes a clear difference, namely that in our model, the asymptotic probability of the two constant functions `True` and `False` (which have complexity 0) is of constant order while it was tending to zero with n in the formula size model. This immediately implies that the tree size model does not exhibit the Shannon effect, whereas proving that the formula size model does not exhibit the Shannon effect is proved but far from trivial (see [14, 15, 17]).

The present paper is organized as follows: In the next section we introduce the model and present our main result. In Section 3, we introduce the combinatorial setting (generating functions for the basic tree classes) and briefly discusses the existence of the limit distribution $\mathbb{P}_n$ (see Equation (2)). Section 4 is devoted to proving key properties of our model: In particular, we investigate in this part the typical shape of the random uniform and/or tree of size m, when m and then n tend to infinity. This section shows how the typical tree in our model and the typical tree in the formula size model are different. In Section 5 we study a particular subfamily of trees which will be used as an auxiliary structure in Sections 6 and 9. In Section 6, we prove that the probability of the two constant functions is of constant order when n tends to infinity, meaning that we prove Equation (2) in the special cases $f = \texttt{True}$ and $f = \texttt{False}$ (note that $L(\texttt{True}) = L(\texttt{False}) = 0$). A fixed Boolean function divides the Boolean lattice into larger and smaller functions (as well as non-comparable functions). This subdivision is quantified in Section 7 and the result will help us to estimate the limiting probability of literal functions, the simplest non-constant functions, in Section 8. We finally have all ingredients needed to prove our main result (i.e. Equation (2)), which we do in Section 9.

Notations: Throughout the article, we define many different classes of trees and thus many different notations. To help the reader, most notations are summarised in Table 1 at the end of the article.

2. MODEL AND RESULTS

Definition 2.1. *An* **associative tree** *is a rooted plane tree whose nodes have arity in* $\mathbb{N} \setminus \{1\}$*, and such that each internal node is labelled by a connector AND (denoted by* $\wedge$ *in the following) or by a connector OR (denoted by* $\vee$ *in the following) such that two identical connectives cannot be neighbours (trees are* **stratified***), and where each leaf is labelled by a literal taken from* $\{x_1, \bar{x}_1, \dots, x_n, \bar{x}_n\}$ *(see Figure 1 for an example).*

Remark: In our context, we take the connectives with an unbounded arity. That corresponds to the property of associativity of the connectives $\wedge$ and $\vee$. Furthermore, the fact that the trees are stratified corresponds to a canonical way of representing a formula as an associative tree. E.g. the formula $A_1 \vee A_2 \vee \cdots \vee A_k$ is represented as a tree consisting of an $\vee$-root and k subtrees representing the formulas $A_1, \dots, A_k$.

By interpreting $\bar{x}_i$ as the negation of x_i, every such tree represents a Boolean formula (or expression) and therefore calculates a Boolean function of n variables. We denote by $\mathcal{F}_n$ the set of such Boolean functions.

Definition 2.2. *The* **tree size** *of a tree is the number of its nodes (internal nodes and leaves). The* **formula size** *of a tree is the number of its leaves. In both cases, we denote by* $|t|$ *the size of a tree* t*.*

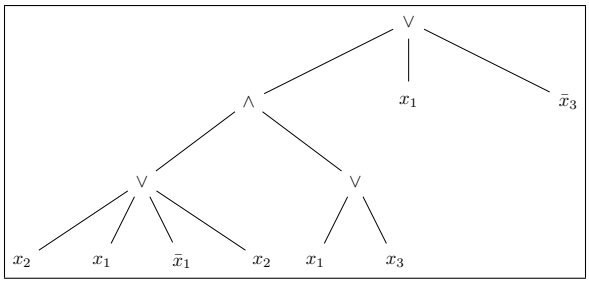

FIGURE 1. An associative tree which represents the constant function `True`.

The aim of this article is to focus on the tree size notion; therefore, in the following, *size* used without specification needs to be understood as the tree size. We denote by $\mathcal{A}_{m,n}$ the set of associative trees of tree size m (labelled on n variables), and by $A_{m,n}$ the cardinality of this set.

Given a notion of size, one can define the complexity as follows: the **complexity** of a Boolean function f is the size of the smallest trees representing it. The following definitions include the special cases of the constant and literal functions for which we need *ad hoc* definitions.

Definition 2.3. *Let $f \in \mathcal{F}_n$ be a Boolean function, its* **tree size complexity***, denoted by $L(f)$, is*

- *0 if f is the function* $\texttt{True} : (x_1, \dots, x_n) \mapsto \texttt{True}$ *or* $\texttt{False} : (x_1, \dots, x_n) \mapsto \texttt{False}$*;*
- *2 if f is a literal function,* i.e. *there exists $x_i \in \{x_1, \dots, x_n\}$ such that $f : (x_1, \dots, x_n) \mapsto x_i$ or $f : (x_1, \dots, x_n) \mapsto \bar{x}_i$;*
- *the* tree *size of the smallest trees computing f if f is neither a literal function nor a constant function. These trees of size $L(f)$ computing f are called* **minimal trees** *of f and their set is denoted by $\mathcal{M}_f$.*

Remark: The literal functions will be treated separately in the whole paper. But they could in fact be treated as any other non-constant function by considering that its two minimal trees are the tree rooted by an $\wedge$ with a single child labelled by x_i and the tree rooted by an $\vee$ with a single child labelled by x_i, even if those two trees are not associative trees according to Definition 2.1!

Although we do not use the formula size complexity notion in this article, our aim is to compare our results to those obtained under the formula size notion; we therefore provide its definition below.

Definition 2.4. *Let $f \in \mathcal{F}_n$ be a Boolean function, its* **formula size complexity** *is*

- *0 if f is the function* $\texttt{True} : (x_1, \dots, x_n) \mapsto \texttt{True}$ *or* $\texttt{False} : (x_1, \dots, x_n) \mapsto \texttt{False}$*;*
- *the* **formula** *size of the smallest trees computing f if f is not a constant function.*

Example: The tree pictured in Figure 1 has tree size 11 (and formula size 8). The function $(x_1, \dots, x_n) \mapsto x_1 \text{ XOR } x_2 = (\bar{x}_1 \wedge x_2) \vee (x_1 \wedge \bar{x}_2)$ has tree size complexity 7 (and formula size complexity 4).

The main contribution of this article is to understand how the change of size notion impacts the distribution induced on the set of Boolean functions by the uniform distribution on the set of and/or trees of size m and labelled on n variables.

Definition 2.5. *For any Boolean function f, we denote by $\mathbb{P}_{m,n}(f)$ the proportion of trees calculating f among all trees in $\mathcal{A}_{m,n}$ (being the set of all associative trees of size m over $\{x_1, \bar{x}_1, \dots, x_n, \bar{x}_n\}$).*

The aim of the present paper is to prove the following result, which sums up the asymptotic behaviour of this distribution $\mathbb{P}_{m,n}$ when the size m of the considered tree tends to infinity.

Theorem 2.6. *For any Boolean function f, the following limit exists and is positive*

$$\mathbb{P}_n(f) = \lim_{m\to\infty} \mathbb{P}_{m,n}(f),$$

and the measure $\mathbb{P}_n(f)$ is a probability distribution on the set $\mathcal{F}_n$ of all Boolean functions on n variables. Moreover,

(i) Asymptotically when n goes to infinity, $\mathbb{P}_n(\texttt{True}) = \mathbb{P}_n(\texttt{False}) = \alpha + o(1)$, where

$$\alpha = \sqrt{2}\int_0^\infty \mathtt{e}^{-v^2-\sqrt{2}v}\,\mathrm{d}v - \frac{1}{2}$$

(ii) If f is a literal function, then $\mathbb{P}_n(f) = \dfrac{(\alpha + o(1))}{n^2}$.

(iii) For all Boolean functions f such that $L(f) \geq 3$,

$$\mathbb{P}_n(f) \sim \frac{\lambda_f}{n^{L(f)}} \text{ when } n \to \infty,$$

where λ_f is a well defined constant satisfying

$$\frac{\alpha L(f)|\mathcal{M}_f|}{2^{L(f)}} \leq \lambda_f \leq \frac{3\alpha L(f)|\mathcal{M}_f|}{2^{L(f)+1}}. \tag{3}$$

Remark: Thanks to the *ad hoc* definition of the complexity (*cf.* Definition 2.3), the above theorem can be expressed as: For all Boolean functions f,

$$\mathbb{P}_n(f) \sim \frac{\lambda_f}{n^{L(f)}}, \text{ as } n \text{ tends to infinity.}$$

As already mentioned in the introduction, our main result shows that changing the notion of size/complexity does not affect the behaviour of the induced distribution on Boolean functions: the main result above is very similar to Equation (1), valid in the formula size complexity context.

We believe that more important than the main result itself are the proofs developed in this article. While the theorem above is a copycat of Kozik's result in the formula size context, the proofs leading to both results are completely different; Kozik's approach collapses in our context. The main reason for this discrepancy lies in the fact that in our model, a typical tree under the uniform distribution on $\mathcal{A}_{m,n}$ has about $\sqrt{n}$ leaves attached to the root (see Section 4), while this number does not depend on n in Kozik's model.

Our main contribution is thus the development of a whole new approach to prove our main result. Having to develop new methods for this slightly different model shows that the proofs developed for the formula size are not robust to this change of metric while the results are. Defining and studying this new model as we do in this article raises the question of the universality of this result, and opens the door to a more general approach to these questions.

3. EXISTENCE OF THE LIMITING DISTRIBUTION $\mathbb{P}_n$

In this section, we prove the existence of the limiting distribution $\mathbb{P}_n$ on the set $\mathcal{F}_n$, which corresponds to the first assertion of Theorem 2.6. In the whole paper, we deeply use generating functions, singularity analysis and the symbolic method. We refer the reader to [9] for a comprehensive introduction to this domain.

An associative tree can be formally described by a grammar. Let us denote by $\mathcal{L}$ the set of literals, $\hat{\mathcal{C}} = \{\wedge\}$ and $\check{\mathcal{C}} = \{\vee\}$. Let $\hat{\mathcal{A}}$ be the set of all associative trees of size 1 or rooted by an $\wedge$-connective, $\check{\mathcal{A}}$ the set of all associative trees of size 1 or rooted by a $\vee$-connective, and denote by $\hat{A}(z)$ and $\check{A}(z)$ their respective generating functions. Then, we have

$$\hat{\mathcal{A}} = \mathcal{L} + \hat{\mathcal{C}} \times \mathtt{seq}_{\geq 2}(\check{\mathcal{A}}) \quad \text{and} \quad \check{\mathcal{A}} = \mathcal{L} + \check{\mathcal{C}} \times \mathtt{seq}_{\geq 2}(\hat{\mathcal{A}});$$

and, by using the symbolic method,

$$\hat{A}(z) = 2nz + z \cdot \frac{\check{A}(z)^2}{1-\check{A}(z)} \quad \text{and} \quad \hat{A}(z) = 2nz + z \cdot \frac{\hat{A}(z)^2}{1-\check{A}(z)}.$$

Since, by symmetry, $\check{A}(z) = \hat{A}(z)$, we get

$$\hat{A}(z) = \frac{2nz + 1 - \sqrt{(4n^2-8n)z^2 - 4nz + 1}}{2(z+1)}. \tag{4}$$

If we denote by $A(z) = \sum_{m\geq 0} A_{m,n} z^m$ the generating function of all associative trees, then

$$A(z) = \hat{A}(z) + \check{A}(z) - 2nz = 2\hat{A}(z) - 2nz.$$

Moreover, let $\hat{A}_f(z)$ (resp. $\check{A}_f(z)$) denote the generating function of associative trees rooted by an $\wedge$ (resp. an $\vee$) connective and computing the Boolean function f. Using the symbolic method, we get

$$\hat{A}_f(z) = z \cdot \mathbb{1}_{f\text{lit}} + z \cdot \sum_{\ell\geq 2} \sum_{g_1\wedge\cdots\wedge g_\ell = f} \check{A}_{g_1}(z)\dots\check{A}_{g_\ell}(z)$$
$$\check{A}_f(z) = z \cdot \mathbb{1}_{f\text{lit}} + z \cdot \sum_{\ell\geq 2} \sum_{g_1\vee\cdots\vee g_\ell = f} \hat{A}_{g_1}(z)\dots\hat{A}_{g_\ell}(z),$$

where $\mathbb{1}_{f\text{lit}}$ is one when f is a literal function and zero otherwise.

We thus have a system of functional equations to which we will apply the Drmota-Lalley-Woods theorem. Let us check the hypothesis of this theorem: We refer the reader to [8, page 489] for a precise statement of the Drmota-Lalley-Woods theorem for polynomial systems. A more general form can be found in the original work [5] and in [6, Section 2.2.5]. The above system is nonlinear, it has nonnegative coefficients and it satisfies a Lipschitz condition. It is irreducible, because for every function f we have $f \wedge \texttt{True} \equiv f$ and $f \vee \texttt{True} \equiv \texttt{True}$. Finally, note that a tree having an $\vee$-root and two subtrees, one being only one leaf labelled by x_1 and the other one a tree of arbitrary size but with all leaves labelled by $\bar{x}_1$, is a tautology. Thus, for all $m \geq 3$, there exists a tree of size m calculating the constant function $\texttt{True}$ and hence $\check{A}_{\texttt{True}}(z)$ is aperiodic. This is enough to imply that all the functions $\hat{A}_f(z)$ and $\check{A}_f(z)$ are also aperiodic. Thus we have shown that all hypotheses of the Drmota-Lalley-Woods are true.

We can therefore infer that all the generating functions $\hat{A}_f(z)$, $\check{A}_f(z)$ and $A(z)$ have the same unique singularity $\rho > 0$ on their (common) circle of convergence, it is a square-root singularity and $\hat{A}_f(z)$, $\check{A}_f(z)$ as well as $A(z)$ admit a singular expansion of the same type (Puiseux expansion in terms of powers of $\sqrt{\rho - z}$ at ρ). Applying a transfer lemma (see [8]) to $A(z)$, $\hat{A}_f(z)$, and $\check{A}_f(z)$ we conclude that the limit $\mathbb{P}_n(f) = \lim_{m\to\infty} \frac{[z^m](\hat{A}_f(z)+\check{A}_f(z))}{[z^m]A(z)}$ exists and is positive for all f. Finally, the fact that $\mathbb{P}_{m,n}$ is a probability distribution on the finite set $\mathcal{F}_n$ (of Boolean functions on n variables) for all $m \geq 1$ implies that $\mathbb{P}_n$ is also a probability distribution, proving the first statement of Theorem 2.6.

Knowing the generating function of $\wedge$–rooted associative trees (given in Equation (4)), we can state the following proposition that will be widely used later on.

Proposition 3.1. *The singularity ρ of $\hat{A}(z)$ satisfies*

$$\rho = \frac{1}{2} \cdot \frac{1}{n+\sqrt{2n}} = \frac{1}{2n} - \frac{1}{n\sqrt{2n}} + \mathcal{O}\left(\frac{1}{n^2}\right), \text{ as } n \to \infty. \tag{5}$$

In particular, for all large enough n, we have $\frac{1}{2n} - \frac{1}{2n\sqrt{n}} \leq \rho < \frac{1}{2n}$. Moreover,

$$A(\rho) = 1 - \frac{1}{n} + \mathcal{O}\left(\frac{1}{n\sqrt{n}}\right) \quad \text{and} \quad \hat{A}(\rho) = 1 - \frac{1}{\sqrt{2n}} + \mathcal{O}\left(\frac{1}{n}\right).$$

Finally, if $\hat{B}(z)$ denotes the generating function of $\wedge$–rooted associative trees, then $\hat{B}(z) = \hat{A}(z) - 2nz$. For all integer n,

$$\hat{B}(\rho) < \frac{1}{\sqrt{2n}} \quad \text{and} \quad \hat{B}(\rho) = \frac{1}{\sqrt{2n}} + \mathcal{O}\left(\frac{1}{n}\right), \text{ when } n \to \infty.$$

4. MISCELLANEOUS PROPERTIES OF THE MODEL

In this section, we prove several propositions that lead to a better understanding of the model and that will be useful throughout the paper. We prove among other results that the expected number of leaves on the first level of a large associative tree behaves like $\sqrt{n}$, and that there are few trees with no leaves on the first level. All these results are really surprising in view of the formula size models of the literature where the number of leaves attached to the root of a typical tree of large size does not depend on the number n of variables. Therefore, in the present model, it is quite likely to have a literal and its negation appearing as labels of two leaves attached to the root, making the whole tree calculate one of the two constant functions, whereas this configuration is very unlikely in the formula size model (see [16]).

First of all, let us define the limiting ratio of a family:

Definition 4.1. *Let $\mathcal{T}$ be a family of associative trees and T_m be the number of trees of size m in this family. The limiting ratio of $\mathcal{T}$, if it exists, is defined (and denoted) by*

$$\mu_n(\mathcal{T}) = \lim_{m\to\infty} \frac{T_m}{A_{m,n}}.$$

The following standard result will be extensively used in the following:

Lemma 4.2. *Let $T(z)$ be the generating function of a family $\mathcal{T}$ of associative trees. Assume that ρ (*cf. *Proposition 3.1) is the unique singularity of $T(z)$ on its circle of convergence and that this singularity is of square-root type,* i.e. *$T(z)$ admits a Puiseux expansion into powers of $\sqrt{\rho - z}$ at ρ. Then, one has*

$$\mu_n(\mathcal{T}) = \lim_{z\to\rho} \frac{T'(z)}{A'(z)}$$

where z must move towards ρ in such a way that $\arg(z-\rho) \neq 0$ and inside the domain of analyticity of $T(z)$ and $A(z)$.

Proof. This is an immediate consequence of the fact that the asymptotic expansion of the derivative of an analytic function is the derivative of the asymptotic expansion of the original function. □

Proposition 4.3. *The limiting ratio of trees with no leaf on the first level is given by*

$$\mu_n(\mathcal{A}^{(0)}) = \frac{1}{n\sqrt{2n}} + \mathcal{O}\left(\frac{1}{n^2}\right), \text{ as } n \to \infty.$$

Proof. The generating function of trees with no leaf on the first level is given by

$$A^{(0)}(z) = 2z\frac{\hat{B}(z)^2}{1-\hat{B}(z)}.$$

The limiting ratio of such trees is thus given by

$$\begin{aligned}\mu_n(\mathcal{A}^{(0)}) = \lim_{m\to\infty} \frac{[z^m]A^{(0)}(z)}{[z^m]A(z)} &= \lim_{z\to\rho} \frac{A'^{(0)}(z)}{A'(z)} \\ =& 2\rho\frac{2\hat{B}(\rho)}{1-\hat{B}(\rho)} \lim_{z\to\rho} \frac{\hat{B}'(z)}{A'(z)} + 2\rho\frac{\hat{B}(\rho)^2}{(1-\hat{B}(\rho))^2} \lim_{z\to\rho} \frac{\hat{B}'(z)}{A'(z)} \\ &+ 2\frac{\hat{B}(\rho)^2}{1-\hat{B}(\rho)} \lim_{z\to\rho} \frac{1}{A'(z)}.\end{aligned}$$

Observe that the third term of the sum is equal to zero since $A'(z)$ tends to infinity when z tends to ρ. This observation will be used in the whole paper and in the following, such terms will be omitted without mentioning. Note also that $\lim_{z\to\rho} {\hat{B}'(z)}/{A'(z)} = {1}/{2}$. Finally, we use the asymptotics given in Proposition 3.1 and get

$$\mu_n(\mathcal{A}^{(0)}) = \frac{1}{n\sqrt{2n}} + \mathcal{O}\left(\frac{1}{n^2}\right) \text{ as } n \to \infty. \qquad \square$$

Proposition 4.4. *Let Γ be a subset of γ literals of $\{x_1, \bar{x}_1, \ldots, x_n, \bar{x}_n\}$. The limiting ratio of the set of all associative trees with at least one leaf on the first level labelled by a literal from Γ is given by*

$$\mu_n(\mathcal{A}_\Gamma) = \gamma\sqrt{\frac{2}{n}} + \mathcal{O}\left(\frac{1}{n}\right), \textit{ when } n \textit{ tends to infinity.}$$

Proof. Let $A_\Gamma(z)$ be the generating function of associative trees with at least one leaf one the first level labelled by a literal from Γ:

$$A_\Gamma(z) = \frac{2\gamma z^2}{(1-(\hat{A}(z)-\gamma z))(1-\hat{A}(z))} - 2\gamma z^2$$

because a tree in $\mathcal{A}_\Gamma$ has a root labelled by $\wedge$ or $\vee$ (which gives a factor $2z$), a first sequence of subtrees which are not a leaf labelled by a literal from Γ (which gives the factor $(1-(\hat{A}(z)-\gamma z))^{-1}$), then a leaf labelled by a literal from Γ (factor γz) and a sequence of arbitrary trees. Since sequences may be empty, this construction also generates trees consisting of only two nodes. These have to be subtracted due to the vertex degree constraints in associative trees. In view of Lemma 4.2, we know that

$$\mu_n(\mathcal{A}_\Gamma) = \lim_{m\to\infty} \frac{[z^m]A_\Gamma(z)}{[z^m]A(z)} = \lim_{z\to\rho} \frac{A'_\Gamma(z)}{A'(z)} = \gamma\sqrt{\frac{2}{n}} + \mathcal{O}\left(\frac{1}{n}\right). \qquad \square$$

Proposition 4.5. *Let $X_{m,n}$ be the random number of leaves in the first level of the uniformly-distributed n variable associative tree of size m.*
We have $\lim_{m\to\infty} \mathbb{E}(X_{m,n}) \sim 2\sqrt{2n}$, as $n \to \infty$.

Proof. Let us consider the bivariate generating function where z marks the nodes and u the leaves on the first level. We have the following equation:

$$A(z,u) = 2z\frac{(\hat{B}(z)+2nzu)^2}{1-(\hat{B}(z)+2nzu)}.$$

Therefore,

$$\frac{\partial}{\partial u}A(z,u)_{|u=1} = 8nz^2\frac{\hat{A}(z)}{1-\hat{A}(z)} + 4nz^2\frac{\hat{A}(z)^2}{(1-\hat{A}(z))^2}.$$

In view of Lemma 4.2, the expected number of nodes in the first level is given by

$$\begin{aligned}\lim_{m\to\infty} \frac{[z^m]\frac{\partial}{\partial u}A(z,u)_{|u=1}}{[z^m]A(z)} &= \lim_{z\to\rho} \frac{\frac{d}{dz}\left(\frac{\partial}{\partial u}A(z,u)\,|_{u=1}\right)}{\frac{d}{dz}A(z)} \\ &= 2\sqrt{2n} + \mathcal{O}(1), \text{ as } n \to \infty \qquad \square\end{aligned}$$

Remark: Using a similar calculation but plugging $u = e^{it/2\sqrt{2n}}$ instead of $u = 1$, we can obtain even the limiting distribution. Indeed, one easily shows that the distribution of the random variable ${X_n}/{2\sqrt{2n}}$ converges to $\Gamma(2, {1}/{2})$, as $n \to \infty$, *i.e.*, the distribution having density $4xe^{-2x}\mathbb{1}_{x\geq 0}$.

5. A USEFUL FAMILY OF TREES

In the following, we will need some information about a specific class of trees which will serve as an auxiliary construction for our further investigations. We will prove that some families of trees are included contained in some of the families defined in this section. This will enable us to give upper bounds for their limiting ratio. The family we define here depends on three parameters, two of them being the number of different labels appearing on the first level leaves and the number of non-leaf subtrees of the root. We prove in particular in this section that

- The limiting ratio of trees with fewer than $n^{1/4}$ different labels appearing on the first-level leaves is $\mathcal{O}(1/\sqrt{n})$ (see Lemma 5.5);
- The limiting ratio of trees whose root has more than $n^{1/8}$ non-leaf subtrees is $\Theta(n^{1/8}2^{-n^{1/8}})$ (see Lemma 5.6).

The results from this section help us to better understand what is the typical shape of an associative tree under the tree-size notion: The two results mentioned above tell us, in a very precise way, that a typical tree *has many different labels appearing on its first level and its root has few non-leaf subtrees.*

Definition 5.1. *Let $k, r, \ell \geq 0$ be three integers, let $\Gamma = \{\gamma_1, \dots, \gamma_p\}$ be a subset of literals (with no occurrence of both a variable and its negation). Let $\mathcal{M}^{\Gamma}_{k,\ell,r}$ be the family of all $\vee$-rooted trees such that there exists $k+r$ pairwise different literals $\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_r \notin \Gamma$ with no occurrence of both a variable and its negation such that*

- *the leaves attached to the root are labelled by literals chosen from $\{\alpha_1, \dots, \alpha_k\}$,*
- *the root has exactly ℓ non-leaf subtrees,*
- *at least one non-leaf subtree attached to the root is in the set $\mathcal{J}^{\Gamma}_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ comprising all $\wedge$-rooted trees such that*
  - *the root has r leaf-children labelled by $\beta_1, \dots, \beta_r$ in that order;*
  - *all the other leaves on the first level have labels from the set $\{\alpha_1, \dots, \alpha_k, \beta_1 \dots, \beta_r, \gamma_1, \dots, \gamma_p\}$ or their negations, but the first appearance of each β_i is in non-negated form.*

Lemma 5.2. *Let $J(z)$ and $M(z)$ denote the generating functions associated with $\mathcal{J}^{\Gamma}_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ and $\mathcal{M}^{\Gamma}_{k,\ell,r}$, respectively. Then*

$$J(z) = \frac{z^{r+1}}{(1-(\hat{B}(z)+2(k+p)z))\dots(1-(\hat{B}(z)+2(k+p+r)z))} \tag{6}$$

and

$$\begin{aligned}[z^m]M(z) \leq & \binom{n-p}{k+r}\binom{k+r}{k}\frac{2^{k+r}k!r!}{(\ell-1)!} \\ & \cdot[z^{m-1}]\left(z^{k+\ell}\prod_{\nu=1}^{k}\frac{1}{1-\nu z}\right)^{(\ell)}\frac{z^{r+1}}{\prod_{\nu=0}^{r}(1-(\hat{B}(z)+2(k+p+\nu)z))}\hat{B}(z)^{\ell-1},\end{aligned} \tag{7}$$

for all $m \geq 0$, where $f^{(\ell)}$ denotes the ℓth derivative of f.

Proof. The first equation is obvious from the definition of $\mathcal{J}^{\Gamma}_{k,r}$. Since Equation (7) is an inequality and not an equality we can do the following reasoning without worrying about double-counting some trees. Let us follow the definition of $\mathcal{M}^{\Gamma}_{k,\ell,r}$: First note that there are $\binom{n-p}{k+r}\binom{k+r}{k}2^{k+r}r!k!$ ways to choose $\alpha_1, \dots, \alpha_k$ and $\beta_1, \dots, \beta_r$ (and their respective order) such that a variable and its negation cannot both be chosen. The leaves of the first level form a sequence of α_1 followed by the first occurrence of α_2, then a sequence of leaves with labels in $\{\alpha_1, \alpha_2\}$ followed by the first occurrence of α_3, and so on. This corresponds to the generating functions $z^k\prod_{\nu=1}^{k}\frac{1}{1-\nu z}$ and if we choose the places of the ℓ non-leaf subtrees of the root between the leaves of the first level, then we have to apply the operator $\frac{1}{\ell!}\frac{\partial^\ell}{\partial z^\ell}$, because $[z^m]\frac{(z^\ell f(z))^{(\ell)}}{\ell!} = \binom{m+\ell}{\ell}f_m$. The generating function of the non-leaf subtree taken from $\mathcal{J}^{\Gamma}_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ and its place among the non-leaf subtrees is $\ell J(z)$ and that of the other non-leaf subtrees is $\hat{B}(z)^{\ell-1}$. When we collect all these terms and take into account

that multiple counting occurs in the cases where several non-leaf subtrees of the root are in $\mathcal{J}^{\Gamma}_{\boldsymbol{\alpha},\boldsymbol{\beta}}$, we get

$$[z^m]M(z) \le [z^m]z\binom{n-p}{k+r}\binom{k+r}{k}2^{k+r}k!\frac{1}{\ell!}\left(z^{k+\ell}\prod_{\nu=1}^{k}\frac{1}{1-\nu z}\right)^{(\ell)}\ell J(z)\hat{B}(z)^{\ell-1}, \tag{8}$$

where the factor z at the very beginning stands for the root. Finally, using Equation (6) to replace $J(z)$ in Equation (8) gives Equation (7) as desired. □

By means of the previous lemma we can express the limiting ratio of the family $\mathcal{M}^{\Gamma}_{k,\ell,r}$ in the form

$$\begin{aligned}\lim_{z\to\rho}\frac{M'(z)}{A'(z)} \le &\binom{n-p}{k+r}\binom{k+r}{k}\frac{2^{k+r}k!r!}{(\ell-1)!}\\ &\cdot\left(\rho^{k+\ell}\prod_{\nu=1}^{k}\frac{1}{1-\nu\rho}\right)^{(\ell)}\rho^{r+2}\lim_{z\to\rho}\frac{K'(z)}{A'(z)}\end{aligned} \tag{9}$$

where $K(z) = {\hat{B}(z)^{\ell-1}}/{\prod_{\nu=0}^{r}(1-(\hat{B}(z)+(k+p+\nu)z))}$. The only term depending on z is ${K'(z)}/{A'(z)}$. The next lemma gives the order of this term.

Lemma 5.3. *Let $K(z)$ be as above. Moreover, assume that $k+p+r<n$ and $r=\mathcal{O}(\sqrt{n})$. Then*

$$\lim_{z\to\rho}\frac{K'(z)}{A'(z)} = \mathcal{O}\left(r2^r\ell\left(\frac{1}{2n}\right)^{\frac{\ell-1}{2}}\right), \text{ as } n\to\infty.$$

Proof. We have, as $z\to\rho$,

$$\begin{aligned}\frac{K'(z)}{A'(z)} \sim &\frac{\hat{B}'(z)}{A'(z)}\frac{1}{\prod_{m=0}^{r}(1-\hat{B}(\rho)-2(k+p+m)\rho)}\\ &\cdot\left[(\ell-1)\hat{B}(\rho)^{\ell-2}+\sum_{m=0}^{r}\frac{\hat{B}(\rho)^{\ell-1}}{1-\hat{B}(\rho)-2(k+p+m)\rho}\right].\end{aligned} \tag{10}$$

Moreover, we know that $\lim_{z\to\rho}\frac{\hat{B}'(z)}{A'(z)}=\frac{1}{2}$ and that $\hat{B}(\rho)\le\frac{1}{\sqrt{2n}}$ (*cf.* Proposition 3.1). Hence

$$\frac{1}{\prod_{m=0}^{r}(1-\hat{B}(\rho)-2(k+p+m)\rho)} \le \left(\frac{1}{2}\left(1-\frac{\sqrt{2}}{\sqrt{n}}\right)\right)^{-r} = \mathcal{O}(2^r).$$

We get

$$\frac{K'(z)}{A'(z)} = \mathcal{O}\left(r2^r\ell\left(\frac{1}{2n}\right)^{\frac{\ell-1}{2}}\right).$$

The term in brackets in Equation (10) consists of $\mathcal{O}(r)$ terms, each bounded by $\ell\hat{B}(\rho)^{\ell-1}$ which immediately yields the assertion. □

Lemma 5.4. *If $k=\Omega(n^{1/4})$, $\ell=\mathcal{O}(n^{1/8})$ and $r\le\ell$, then the limiting ratio of the family $\mathcal{M}^{\Gamma}_{k,\ell,r}$ satisfies $\mu_n(\mathcal{M}^{\Gamma}_{k,\ell,r})=\mathcal{O}\left(\frac{1}{n^{3/2}}\right)$, as n tends to infinity.*

Proof. Turning back to (9) we see that the last factor is covered by Lemma 5.3. In order to estimate the penultimate factore, let $f(z)=z^{k+\ell}\prod_{\nu=1}^{k}{}^{1}/_{(1-\nu z)}$ and consider a circle of radius $n^{-3/2}$ centered at ρ. Since $f(z)$ is a power series with nonnegative coefficients, Cauchy's estimate gives

$$\left|\frac{f^{(\ell)}(\rho)}{\ell!}\right| \le f\left(\rho+\frac{1}{n^{3/2}}\right)n^{3\ell/2}.$$

Now, performing the substitution $z = \rho + \frac{1}{n^{3/2}}$ into $f(z)$ on the right-hand side, using the monotonicity of f and the inequality $\rho < 1/2n$ (*cf.* Proposition 3.1), we obtain

$$\begin{aligned}\left|\frac{f^{(\ell)}(\rho)}{\ell!}\right| &= \left(\rho + \frac{1}{n^{3/2}}\right)^{k+\ell} n^{3\ell/2} \prod_{\nu=1}^{k} \frac{1}{1-\nu\left(\rho+\frac{1}{n^{3/2}}\right)} \\ &\leq \left(\frac{1}{2n}\right)^{k+\ell}\left(1+\frac{2}{\sqrt{n}}\right)^{k+\ell} n^{3\ell/2} \prod_{\nu=1}^{k} \frac{1}{1-\frac{\nu}{2n}\left(1+\frac{2}{\sqrt{n}}\right)}.\end{aligned}$$

Thanks to Equation (7), we get

$$\begin{aligned}\lim_{z\to\rho}\frac{M'(z)}{A'(z)} &\leq \rho^{r+2}\binom{n-p}{k+r}\binom{k+r}{k}2^{k+r}k!r!\frac{f^{(\ell)}(\rho)}{(\ell-1)!}\lim_{z\to\rho}\frac{K'(z)}{A'(z)} \\ &\leq \ell\left(\frac{1}{2n}\right)^{k+\ell+r+2} 2^{k+r}(n)_k n^r\left(1+\frac{2}{\sqrt{n}}\right)^{k+\ell} n^{3\ell/2} \\ &\quad\cdot\prod_{\nu=1}^{k}\frac{1}{1-\frac{\nu}{2n}\left(1+\frac{2}{\sqrt{n}}\right)}\lim_{z\to\rho}\frac{K'(z)}{A'(z)} \\ &\leq \ell e^2 2^{-\ell-2}\cdot n^{\ell/2-2}\left(1+\frac{2}{\sqrt{n}}\right)^k \prod_{\nu=1}^{k}\frac{1-\frac{\nu}{n}}{1-\frac{\nu}{2n}\left(1+\frac{2}{\sqrt{n}}\right)}\lim_{z\to\rho}\frac{K'(z)}{A'(z)}.\end{aligned}$$

The last product is bounded by 1 and if $k = \mathcal{O}(\sqrt{n})$, then $\left(1+\frac{2}{\sqrt{n}}\right)^k$ is bounded as well. In this case we get

$$\lim_{z\to\rho}\frac{M'(z)}{A'(z)} = \mathcal{O}\left(\ell 2^{-\ell} n^{\ell/2-2}\lim_{z\to\rho}\frac{K'(z)}{A'(z)}\right). \tag{11}$$

If $k > 4\sqrt{n}$ then it suffices to show that $\left(1+\frac{2}{\sqrt{n}}\right)^k n^{3\ell/2}\prod_{\nu=1}^{k}\frac{1}{1-\frac{\nu}{2n}\left(1+\frac{2}{\sqrt{n}}\right)}$ is bounded. We proceed as follows: First observe that

$$\frac{1-\frac{\nu}{n}}{1-\frac{\nu}{2n}\left(1+\frac{2}{\sqrt{n}}\right)} < 1-\frac{\nu}{2n}\left(1-\frac{2}{\sqrt{n}}\right)$$

and thus we can write

$$\begin{aligned}&\left(1+\frac{2}{\sqrt{n}}\right)^k\prod_{\nu=1}^{k}\frac{1-\frac{\nu}{n}}{1-\frac{\nu}{2n}\left(1+\frac{2}{\sqrt{n}}\right)} \\ &= \left(1+\frac{2}{\sqrt{n}}\right)^{4\lfloor\sqrt{n}\rfloor}\prod_{\nu=1}^{4\lfloor\sqrt{n}\rfloor}\frac{1-\frac{\nu}{n}}{1-\frac{\nu}{2n}\left(1+\frac{2}{\sqrt{n}}\right)} \\ &\qquad\cdot\left(1+\frac{2}{\sqrt{n}}\right)^{k-4\lfloor\sqrt{n}\rfloor}\prod_{\nu=4\lfloor\sqrt{n}\rfloor+1}^{k}\frac{1-\frac{\nu}{n}}{1-\frac{\nu}{2n}\left(1+\frac{2}{\sqrt{n}}\right)} \\ &\leq e^8\left(1+\frac{2}{\sqrt{n}}\right)^{k-4\lfloor\sqrt{n}\rfloor}\cdot\left(1-\frac{2}{\sqrt{n}}\left(1-\frac{2}{\sqrt{n}}\right)\right)^{k-4\lfloor\sqrt{n}\rfloor} \\ &= e^8\left(1+\frac{8}{n^{3/2}}\right)^{k-4\lfloor\sqrt{n}\rfloor} \\ &= \mathcal{O}(1).\end{aligned}$$

Finally, if we let n tend to infinity and apply Lemma 5.3, we get

$$\mu_n(\mathcal{M}^{\Gamma}_{k,\ell,r}) = \mathcal{O}\left(r\ell^2 2^{r-\frac{3\ell-1}{2}} n^{-3/2}\right) = \mathcal{O}\left(\frac{1}{n^{3/2}}\right). \qquad \square$$

Lemma 5.5. *The limiting ratio of trees with fewer than $n^{1/4}$ different labels appearing on the first level leaves is $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$.*

Proof. The generating function of trees with exactly k different labels appearing on the first level (with no occurrence of a variable and its negation) is given by

$$G_k(z) = \binom{n}{k} 2^k k! z \prod_{m=0}^{k} \frac{z}{1 - mz - \hat{B}(z)},$$

and therefore, their limiting ratio is given by

$$\begin{aligned}
\lim_{z\to\rho} \frac{G_k'(z)}{A'(z)} &= \binom{n}{k} 2^k k! \rho \prod_{m=0}^{k} \frac{\rho}{1 - m\rho - \hat{B}(\rho)} \sum_{m=0}^{k} \frac{1}{1 - m\rho - \hat{B}(\rho)} \cdot \lim_{z\to\rho} \frac{\hat{B}'(z)}{A'(z)} \\
&\le \frac{1}{4n} \prod_{m=0}^{k-1} \left(1 - \frac{m}{n}\right) \prod_{m=0}^{k} \frac{1}{1 - \frac{m}{2n} - \frac{1}{\sqrt{2n}}} \left(\sum_{m=0}^{k} \frac{1}{1 - \frac{m}{2n} - \frac{1}{\sqrt{2n}}}\right) \\
&\le \frac{1}{4n} \left(\frac{1}{1 - \frac{k}{2n} - \frac{1}{\sqrt{2n}}}\right)^{k+1} (k+1) \frac{1}{1 - \frac{k}{2n} - \frac{1}{\sqrt{2n}}} = \mathcal{O}\left(\frac{k}{n}\right).
\end{aligned}$$

Summing up over all k from 1 to $\lfloor n^{1/4} \rfloor$ yields the result. $\square$

Lemma 5.6. *The limiting ratio of the family of all trees where the root has more than $n^{1/8}$ non-leaf subtrees is $\Theta\left(\frac{n^{1/8}}{2^{n^{1/8}}}\right)$.*

Proof. The generating function of the trees with exactly ℓ non-leaf subtrees is

$$H_\ell(z) = 2z \frac{\hat{B}^\ell(z)}{(1 - 2nz)^{\ell+1}}.$$

Therefore, the limiting ratio of trees with exactly ℓ non-leaf subtrees is given by

$$\begin{aligned}
\lim_{z\to\rho} \frac{H_\ell'(z)}{A'(z)} &= \frac{1}{n} \frac{\ell \hat{B}^{\ell-1}(\rho)}{(1 - 2n\rho)^{\ell+1}} \cdot \lim_{z\to\rho} \frac{\hat{B}'(z)}{A'(z)} \\
&\sim \frac{1}{2n} \frac{\ell \left(\frac{1}{\sqrt{2n}}\right)^{\ell-1}}{\left(\sqrt{\frac{2}{n}}\right)^{\ell+1}} \\
&\sim \frac{1}{2n} \frac{\ell \left(\frac{1}{\sqrt{2n}}\right)^{\ell-1}}{\left(\sqrt{\frac{2}{n}}\right)^{\ell+1}} \sim \frac{\ell}{2^{\ell+1}}, \text{ as } n \to \infty.
\end{aligned}$$

This implies that the limiting ratio of trees with more than $n^{1/8}$ non-leaf subtrees is given by

$$\sum_{\ell \ge n^{1/8}} \frac{\ell}{2^{\ell+1}} = \Theta\left(\frac{n^{1/8}}{2^{n^{1/8}}}\right). \qquad \square$$

6. TAUTOLOGIES

In this section, we calculate the probability of the two constant functions `True` and `False`, and prove Theorem 2.6(i), i.e. that, asymptotically when n tends to infinity,

$$\mathbb{P}_n(\texttt{True}) = \mathbb{P}_n(\texttt{False}) = (\sqrt{2} + o(1)) \int_0^\infty \mathtt{e}^{-v^2-\sqrt{2}v}\,\mathrm{d}v - \frac{1}{2} \approx 0.15567954.$$

Note that the proofs of Theorem 2.6(ii) and (iii) rely on this first result on the probability of constant functions.

The proof of this result has two steps, corresponding to the two following lemmas: first calculate the limiting ratio of a family of tautologies called *simple tautologies*, then prove that this limiting ratio is asymptotically equal to the limiting ratio of all tautologies, said differently, *asymptotically almost all tautologies are simple.*

Definition 6.1. *A* **simple tautology** *(resp. a simple contradiction) realized by the variable x is an $\vee$-rooted (resp. $\wedge$-rooted) tree such that the labels x and $\bar{x}$ both appear on the first level of the tree. The set of simple tautologies will be denoted by $\mathcal{S}_n$, its complement by $\overline{\mathcal{S}_n}$.*

Lemma 6.2. *The limiting ratio of simple tautologies satisfies*

$$\mu_n(\mathcal{S}_n) = \sqrt{2}\int_0^\infty \mathtt{e}^{-v^2-\sqrt{2}v}\,\mathrm{d}v - \frac{1}{2} + o(1), \text{ as } n \to \infty.$$

Lemma 6.3. *Asymptotically when n tends to infinity, almost every tautology is a simple tautology, i.e.*

$$\mathbb{P}_n(\texttt{True}) = \mu_n(\mathcal{S}_n) + o(1), \text{ as } n \to \infty.$$

The following two sections contain the proofs of those two lemmas.

6.1. Limiting ratio of simple tautologies (proof of Lemma 6.2).

Let us consider the family of trees rooted by $\vee$ that are not simple tautologies, meaning that their first-level leaves do not contain both a variable and its negation. First note that, the number of ways to label ℓ leaves with literals from the set $\{x_1, \bar{x}_1, \ldots, x_n, \bar{x}_n\}$ without both a variable and its negation appearing among those labels is equal to the number of ways to choose ℓ variables and then which of them are negated. Written as labelled combinatorial structure, this is $\prod_{i=1}^n \left((\texttt{set}\{x_i\} \cup \texttt{set}\{\bar{x}_i\}) \setminus \{\emptyset\}\right)$ and, using the symbolic method, the number of structures of size ℓ is thus $\ell![x^\ell]\big(2\mathtt{e}^x - 1\big)^n$. The result was also shown in a more elementary way by Selivanov [28]. The class of $\vee$–rooted trees that are not simple tautologies is either $\vee$–rooted trees without leaves on the first level, and thus at least two non-leaf subtrees (on the first level) or $\vee$–rooted trees with a single leaf on the first level, and thus at least one non-leaf subtree (on the first level) or $\vee$–rooted trees with $\ell \geq 2$ leaves on the first level and $\ell + 1$ (possibly empty) sequences of non-leaf subtrees. Thus its generating function is given by

$$\begin{aligned}\Psi(z) &= \frac{z\hat{B}^2(z)}{1-\hat{B}(z)} + \frac{2nz^2}{(1-\hat{B}(z))^2} - 2nz^2 + \sum_{\ell\geq 2} \ell![x^\ell]\big(2\mathtt{e}^x - 1\big)^n \frac{z^{\ell+1}}{(1-\hat{B}(z))^{\ell+1}},\\ &= \left(\sum_{\ell\geq 0} \ell![x^\ell]\big(2\mathtt{e}^x - 1\big)^n \frac{z^{\ell+1}}{(1-\hat{B}(z))^{\ell+1}}\right) + \frac{z(\hat{B}^2(z)-1)}{1-\hat{B}(z)} - 2nz^2.\end{aligned}$$

By Lemma 4.2, we have

$$\mu_n(\mathcal{S}_n) = \frac{1}{2} - \lim_{z\to\rho} \frac{\Psi'(z)}{A'(z)} = \frac{1}{2} - \frac{1}{2}\lim_{z\to\rho}\frac{\Psi'(z)}{\check{A}'(z)}.$$

First, recall that $\lim_{z\to\rho} \check{A}'(z) = \infty$, $\hat{B}(\rho) < \frac{1}{\sqrt{2n}}$ and $\lim_{z\to\rho} \dfrac{\hat{B}'(z)}{\check{A}'(z)} = 1$. Thus

$$\lim_{z\to\rho}\left(\frac{z(\hat{B}^2(z)-1)}{1-\hat{B}(z)} - 2nz^2\right)' \frac{1}{\check{A}'(z)} = o(1), \text{ as } n \to \infty.$$

Second, let us denote

$$\bar{\Psi}(z) = \sum_{\ell \geq 0} \ell![x^\ell]\big(2\mathsf{e}^x - 1\big)^n \frac{z^{\ell+1}}{(1-\hat{B}(z))^{\ell+1}}.$$

Thus,

$$\begin{aligned}\lim_{z\to\rho} \frac{\bar{\Psi}'(z)}{\check{A}'(z)} &= \sum_{\ell\geq 0}(\ell+1)!\,[x^\ell](2\mathsf{e}^x-1)^n \frac{\rho^{\ell+1}}{(1-\hat{B}(\rho))^{\ell+2}} \\ &= \frac{\rho}{(1-\hat{B}(\rho))^2}\int_0^\infty t\mathsf{e}^{-t}\left(2\exp\left(\frac{t\rho}{1-\hat{B}(\rho)}\right)\right)^n \mathrm{d}t,\end{aligned}$$

where we have used the identity $(\ell+1)! = \int_0^\infty t^{\ell+1}\mathsf{e}^{-t}\,\mathrm{d}t$. By the change of variables $x = \exp\left(\frac{t\rho}{1-\hat{B}(\rho)}\right)$, we get

$$\lim_{z\to\rho} \frac{\bar{\Psi}'(z)}{\check{A}'(z)} = \frac{1}{\rho}\int_1^\infty \frac{\log x}{x}\,\frac{(2x-1)^n}{x^{f(n)}}\,\mathrm{d}x, \tag{12}$$

where

$$f(n) := \frac{1-\hat{B}(\rho)}{\rho} = 2n + \sqrt{2n} + \mathcal{O}(1),$$

using the exact expressions of ρ and $\hat{B}(z)$ given in Proposition 3.1.

We now use Laplace's method (see e.g. [3]) to estimate the above integral when n tends to infinity. Note that, for all $\varepsilon > 0$, for n sufficiently large,

$$\begin{aligned}\int_{1+\varepsilon}^\infty \frac{\log x}{x}\,\frac{(2x-1)^n}{x^{f(n)}}\,\mathrm{d}x &\leq \frac{(2\varepsilon+1)^n}{(1+\varepsilon)^{f(n)-1}}\int_{1+\varepsilon}^\infty \frac{\log x}{x^2}\,\mathrm{d}x \\ &\leq \frac{(2\varepsilon+1)^n}{(1+\varepsilon)^{2n}}\,\frac{\log(1+\varepsilon)-1}{1+\varepsilon}.\end{aligned}$$

Since ε can be chosen arbitrarily small,

$$\begin{aligned}\int_{1+\varepsilon}^\infty \frac{\log x}{x}\,\frac{(2x-1)^n}{x^{f(n)}}\,\mathrm{d}x &\leq 3\exp\left(n\log\left(1-\frac{\varepsilon^2}{(1+\varepsilon)^2}\right)\right) \leq 3\exp\left(\frac{-n\,\varepsilon^2}{(1+\varepsilon)^2}\right) \\ &= o(1), \text{ as } n\to\infty.\end{aligned}$$

Furthermore, for n sufficiently large and ε sufficiently small, one can apply a Taylor expansion near $x = 1$ and get

$$\begin{aligned}\int_1^{1+\varepsilon} \frac{\log x}{x}\,\frac{(2x-1)^n}{x^{f(n)}}\,\mathrm{d}x &= (1+o(1))\int_1^{1+\varepsilon}(x-1)\mathsf{e}^{-n(x-1)^2-(f(n)-2n)(x-1)}\,\mathrm{d}x \\ &= (1+o(1))\int_1^{1+\varepsilon}(x-1)\mathsf{e}^{-n(x-1)^2-\sqrt{2n}(x-1)}\,\mathrm{d}x.\end{aligned}$$

One can check that the remainder terms in the Taylor expansions are indeed negligible when n tends to infinity and also that replacing $f(n)-2n$ by its first order approximation $\sqrt{2n}$ introduces only a negligible error term; it follows from standard arguments involved when applying the Laplace method. Thus, setting $v = \sqrt{n}(x-1)$, we get

$$\begin{aligned}\int_1^{1+\varepsilon} \frac{\log x}{x}\,\frac{(2x-1)^n}{x^{f(n)}}\,\mathrm{d}x &= \frac{1+o(1)}{2n}\int_0^{\varepsilon\sqrt{n}} 2v\mathsf{e}^{-v^2-\sqrt{2}v}\,\mathrm{d}v \\ &= \frac{1+o(1)}{2n}\left(1-\sqrt{2}\int_0^\infty \mathsf{e}^{-v^2-\sqrt{2}v}\,\mathrm{d}v\right).\end{aligned}$$

Therefore, since $\rho \sim 1/2n$, Equation (12) gives

$$\lim_{z\to\rho}\frac{\Psi'(z)}{\check{A}'(z)} = (1+o(1))\left(1-\sqrt{2}\int_0^\infty \mathsf{e}^{-v^2-\sqrt{2}v}\,\mathrm{d}v\right),$$

which concludes the proof.

6.2. **Almost every tautology is simple (proof of Lemma 6.3).** To prove Lemma 6.3, we consider the family of non-simple tautologies, study the structure of its elements and show that its limiting ratio tends to zero when n tends to infinity.

Note that a tautology rooted by the connective $\wedge$ cannot have leaves on the first level. Thus the family of tautologies rooted by $\wedge$ is contained into the family $\mathcal{A}^{(0)}$ of trees with no leaf on the first level, which, in view of Proposition 4.3, has a limiting ratio of order $n^{-3/2}$. The set of $\wedge$-rooted tautologies is thus negligible compared to the set of all tautologies and we can focus on $\vee$-rooted tautologies.

Let us consider $\mathcal{N}_n$ the set of tautologies that are not simple. Let $t \in \mathcal{N}_n$, with ℓ non-leaf subtrees $A_1, \ldots, A_\ell$ and k different labels $\alpha_1, \ldots, \alpha_k$ appearing in the first level. Since t is not a simple tautology, the set $L = \{\alpha_1, \ldots, \alpha_k\}$ cannot contain both a variable and its negation. For each $i \in \{1, \ldots, \ell\}$, the tree A_i has some of its leaves in its first level labelled by labels in L and others carrying new labels. What will be called a "new variable" in the following is the label of a leaf in the *first level* of one of the A_i if neither the label itself nor its negation belongs to L.

Let us show that there exists at least one $i \in \{1, \ldots, \ell\}$ such that A_i has at most $\ell - 1$ new variables.

Let us assume that for all $i \in \{1, \ldots, \ell\}$, A_i has at least ℓ new variables in the first level. We assign one of those new variables, say ν_1 and belonging to A_1, to `False`. Then A_1 computes `False`. But, there must exist $A_{(2)} \in \{A_2, \ldots, A_\ell\}$ which is not a contradiction for this assignment, because if $A_1 = A_2 = \ldots = A_\ell \equiv$ `False` for $\nu_1 =$ `False`, then the whole tree t would compute $\alpha_1 \vee \ldots \vee \alpha_k$ and would thus not be a tautology.

Let us iterate this algorithm: After step $s - 1 \leq \ell$, we have assigned $\nu_1 = \nu_2 = \ldots = \nu_{s-1} =$ `False` and at least $s - 1$ trees of $\{A_1, \ldots, A_\ell\}$ compute `False`. At step s note that at least one of the remaining subtrees must still not be a contradiction: Let us call this tree $A_{(s)}$. It has at least ℓ new variables and we have assigned $s - 1 \leq \ell$ variables to `False` so far. Therefore we still have free new variables among the new variables of $A_{(s)}$, and we can assign one, denoted by ν_s, to `False`.

After the ℓth step, we have found an assignment of variables different from those in L such that all trees of $\{A_1, \ldots, A_\ell\}$ compute `False`. Thus t computes $\alpha_1 \vee \ldots \vee \alpha_k$, which is impossible since t is a tautology.

Therefore, there exist at least one $i \in \{1, \ldots, \ell\}$ such that A_i has at most $\ell - 1$ new variables in the first level.

It means that $\mathcal{N}_n \subseteq \bigcup_{k=0}^{n} \bigcup_{\ell=0}^{\infty} \bigcup_{r=0}^{\ell-1} \mathcal{M}^{\emptyset}_{k,\ell,r}$ (*cf.* Section 5). Let us decompose this union into three distinct unions:

$$\bigcup_{k=0}^{n} \bigcup_{\ell=0}^{\infty} \bigcup_{r=0}^{\ell-1} \mathcal{M}^{\emptyset}_{k,\ell,r} = \left(\bigcup_{k=0}^{\lfloor n^{1/4} \rfloor} \bigcup_{\ell=0}^{\infty} \bigcup_{r=0}^{\ell-1} \mathcal{M}^{\emptyset}_{k,\ell,r} \right) \cup \left(\bigcup_{k=\lfloor n^{1/4} \rfloor}^{n} \bigcup_{\ell=\lfloor n^{1/8} \rfloor}^{\infty} \bigcup_{r=0}^{\ell-1} \mathcal{M}^{\emptyset}_{k,\ell,r} \right)$$
$$\cup \left(\bigcup_{k=\lfloor n^{1/4} \rfloor}^{n} \bigcup_{\ell=0}^{\lfloor n^{1/8} \rfloor} \bigcup_{r=0}^{\ell-1} \mathcal{M}^{\emptyset}_{k,\ell,r} \right).$$

Thanks to Lemma 5.5, the first term has a limiting ratio tending to zero as n tends to infinity; Lemma 5.6 guarantees that the second term has also a limiting ratio tending to zero, and by Lemma 5.4, the third term satisfies

$$\mu_n \left(\left(\bigcup_{k=\lfloor n^{1/4} \rfloor}^{n} \bigcup_{\ell=0}^{\lfloor n^{1/8} \rfloor} \bigcup_{r=0}^{\ell-1} \mathcal{M}^{\emptyset}_{k,\ell,r} \right) \right) = \mathcal{O}\left((n - n^{1/4})(n^{1/8})^2 \frac{1}{n^{3/2}} \right) = \mathcal{O}\left(\frac{1}{n^{1/4}} \right)$$

and is thus also tending to zero when n tends to infinity. Thus, $\mu_n(\mathcal{N}_n \setminus \mathcal{S}_n) = o(1)$, which concludes the proof.

7. PROBABILITY OF FUNCTIONS LARGER THAN A FIXED f_0

In this section we investigate the probability of the set of all functions that are "larger" than a given Boolean function. The result will enable us to prove the second statement of Theorem 2.6 about literal functions, proved in Section 8.

Definition 7.1. *Let f and g be two Boolean functions of n variables, we say that $g \geq f$ if, and only if, $g(x_1, \ldots, x_n) \geq f(x_1, \ldots, x_n)$ for all $(x_1, \ldots, x_n) \in \{0,1\}^n$.*

Proposition 7.2. *Choose n_0 and a non-constant function $f_0 \in \mathcal{F}_{n_0}$. Moreover, denote the natural extension of f_0 to $\mathcal{F}_n$ for $n > n_0$,* i.e. *the function $(x_1, \ldots, x_{n_0}, x_{n_0+1}, \ldots, x_n) \mapsto f_0(x_1, \ldots, x_{n_0})$, by f_0 as well. Denote by F a random Boolean function of law $\mathbb{P}_n$,* i.e. *$\mathbb{P}(F = f) = \mathbb{P}_n(f)$ for any Boolean function f. Then*

$$\mathbb{P}(F \geq f_0) = \mathbb{P}_n(\texttt{True})\ (1 + o(1)), \textit{ as } n \textit{ tends to infinity.}$$

The remaining part of this section is devoted to the proof of Proposition 7.2. We prove this proposition in two steps: First, assume that f_0 is a conjunction of literals, and then we extend the proof to obtain the general result.

7.1. Case I: f_0 is a conjunction of literals.

Let $f_0 = \gamma_1 \wedge \ldots \wedge \gamma_p$, where the γ_i's are literals. First, let us remark that $\mathbb{P}(F \geq f_0) \geq \mathbb{P}_n(\texttt{True}) \geq \alpha > 0$. Let us consider an associative tree t computing a Boolean function which is larger than f_0 and not a tautology.

The family of trees with no leaf on the first level has a limiting ratio which is asymptotically equal to $1/n\sqrt{2n}$ when n tends to infinity (*cf.* Proposition 4.3). It is thus negligible compared to $\mathbb{P}(\texttt{True})$. Thus, we can assume that t has at least one leaf on the first level.

Consider first the case where t is rooted by an $\vee$:

- Let us first assume that there exists one leaf on the first level of t, labelled by one of the γ_i. The family of trees with at least one leaf on the first level and with a label from the set $\Gamma = \{\gamma_1, \ldots, \gamma_p\}$ has a limiting ratio equivalent to $p\sqrt{2/n}$, in view of Proposition 4.4. The limiting ratio of such trees is thus negligible compared to the limiting ratio of the set of tautologies. We can thus neglect this family.
- Let us assume that t has no leaf on the first level labelled by a literal chosen in $\Gamma = \{\gamma_1, \ldots, \gamma_p\}$. Let us denote by k the number of different labels, denoted by $\alpha_1, \ldots, \alpha_k$, appearing on the first level of t and by ℓ the number of its non-leaf subtrees, denoted by $A_1, \ldots, A_\ell$. Observe that since t is not a tautology, the labels appearing on the first level of t cannot contain a variable and its negation. The subtrees $A_1, \ldots, A_\ell$ have themselves leaves on their first level (*i.e.* on the second level of t), and those leaves are labelled either by "old variables", *i.e.* by literals chosen from $\mathcal{O} = \{\alpha_1, \ldots, \alpha_k, \gamma_1, \ldots, \gamma_p\}$ and their negations, or by "new variables", i.e by other literals. Assume that for all $i \in \{1, \ldots, \ell\}$, A_i has at least ℓ different new variables appearing on its first level. Thus, since each A_i is rooted by an $\wedge$, we can find an assignment of the variables $\{x_1, \ldots, x_n\} \setminus \mathcal{O}$ such that all A_i compute `False` for this assignment. Assign then all the leaves on the first level of t to `False` and $\gamma_1, \ldots, \gamma_p$ to `True`. Then t computes `False` while $\gamma_1 \wedge \ldots \wedge \gamma_p$ takes the value `True` for this assignment. But this is impossible!

  Thus, there exists at least one A_i which has fewer than ℓ new variables on its first level. This means that t belongs to the set $\bigcup_{r=0}^{\ell-1} \mathcal{M}^{\Gamma}_{k,\ell,r}$.

  The limiting ratio of such trees is thus less than the limiting ratio of the union $\bigcup_{k,\ell \geq 0} \bigcup_{r=0}^{\ell-1} \mathcal{M}^{\Gamma}_{k,\ell,r}$, and thanks to the results proved in Section 5, we have

$$\mu_n\left(\bigcup_{k,\ell\geq 0}\bigcup_{r=0}^{\ell-1} \mathcal{M}^{\Gamma}_{k,\ell,r}\right) = \mathcal{O}\left((n - n^{1/4})(n^{1/8})^2 \frac{1}{n^{3/2}}\right) = \mathcal{O}\left(\frac{1}{n^{1/4}}\right)$$

  and this family is also negligible compared to the set of tautologies.

If t is rooted by an $\wedge$, then its first level leaves have labels chosen from the set $\Gamma = \{\gamma_1, \ldots, \gamma_p\}$. The family of trees with first level leaves labelled in $\{\gamma_1, \ldots, \gamma_p\}$ has generating function

$$H_\Gamma(z) = 2nz + 2z\frac{(\hat{B}(z) + pz)^2}{1 - \hat{B}(z) - pz}$$

and its limiting ratio is asymptotically equal to $1/n\sqrt{2n}$. It is thus negligible compared to the set of tautologies. We have proved that the limiting ratio of the set of n-variable associative trees that are not tautologies and that represent a Boolean function larger than $\gamma_1 \wedge \ldots \wedge \gamma_p$ is negligible compared to the limiting ratio of the set of all n-variable tautologies when n goes to infinity. Thus

$$\mathbb{P}(F \geq \gamma_1 \wedge \ldots \wedge \gamma_p) \sim \mathbb{P}_n(\texttt{True}),$$

as n tends to infinity.

7.2. **Case II: f_0 is any non-constant Boolean function.** All Boolean functions f_0 can be written as $f_0 = (\gamma_1 \wedge \ldots \wedge \gamma_p) \vee g_0$ for some integer $p \geq 1$, some literals $\{\gamma_1, \ldots, \gamma_p\}$, and some Boolean function g_0. Thus,

$$\mathbb{P}_n(\texttt{True}) \leq \mathbb{P}(F \geq f_0) \leq \mathbb{P}(F \geq \gamma_1 \wedge \ldots \wedge \gamma_p) \sim \mathbb{P}_n(\texttt{True})$$

and Proposition 7.2 is proved. □

8. LITERALS

The aim of this section is to estimate the probability of a literal Boolean function *i.e.* a function of the shape $((x_1, \ldots, x_n) \mapsto x)$ where x is a literal among $\{x_1, \bar{x}_1, \ldots, x_n, \bar{x}_n\}$, therefore proving the second statement of Theorem 2.6. As for tautologies, we will prove that a typical tree computing this function has a very simple shape.

Definition 8.1. *A tree t is a* **simple x tree** *if it is rooted by an $\wedge$ (resp. $\vee$), with one single leaf on the first level, labelled by x and with one non-leaf subtree which is a tautology (resp. a contradiction). We denote by $\mathcal{X}$ the family of such trees and by X_m the number of simple x trees of size m.*

Lemma 8.2. *For n tending to infinity, the limiting ratio of the set of simple x trees satisfies*

$$\mu_n(\mathcal{X}) \sim \frac{\mathbb{P}_n(\texttt{True})}{n^2}.$$

Proof. The generating function associated with the set of simple x trees is given by the following generating function:

$$X(z) = 4z^2 T(z),$$

where the 4 factor contains the choice of the label of the root and the order of its first and second child. Therefore,

$$\mu_n(\mathcal{X}) = \lim_{z \to \rho} \frac{X'(z)}{A'(z)} = 4\rho^2 \lim_{z \to \rho} \frac{T'(z)}{A'(z)}.$$

Observing that $\frac{T'(z)}{A'(z)}$ tends to $\mathbb{P}_n(\texttt{True})$ when z tends to ρ and that $\rho \sim 1/2n$ when n tends to infinity (*cf.* Proposition 3.1) permits to complete the proof. □

Theorem 8.3. *For n tending to infinity, we have $\mathbb{P}_n(x) \sim \mu_n(\mathcal{X})$.*

Proof. Thanks to Lemma 8.2 and Theorem 2.6(i), we know that $\mathbb{P}_n(x) \geq \frac{\mathbb{P}_n(\texttt{True})}{2n^2}$ is of order at least $1/n^2$ when n tends to infinity. Let t be a tree computing x. Let us assume that it is rooted by an $\wedge$ (the case of an $\vee$-rooted tree would be treated in the very same way). The family of trees computing x with no leaf on the first level has the same limiting ratio for all $x \in \{x_1, \bar{x}_1, \ldots, x_n, \bar{x}_n\}$. Therefore, if we denote this family by $\mathcal{A}_x^{(0)}$, its limiting ratio satisfies (*cf.* Proposition 4.3)

$$2n\mu_n(\mathcal{A}_x^{(0)}) \leq \mu_n(\mathcal{A}^{(0)}) \sim \frac{1}{n\sqrt{2n}}.$$

Thus, the limiting ratio of trees with no leaf on the first level, computing x has order $\mathcal{O}(n^{-5/2})$. This family is negligible compared to the family of simple x trees. Thus we can focus on the family $\mathcal{T}_x$ of trees having at least one leaf on the first level.

The leaves on the first level of t have to be labelled by x since t computes x. And the non-leaf subtrees calculate functions that are larger than x (in the sense of Definition 7.1). Thus, a tree computing x is almost surely an $\wedge$-rooted (resp. $\vee$-rooted) tree with leaves on the first level labelled by x and with non-leaf subtrees larger than x (resp. smaller than x).

Let us denote by $L_x(z)$ the generating function of trees larger than x. In view of Proposition 7.2, we have

$$\lim_{z\to\rho} \frac{L_x'(z)}{A'(z)} \sim \mathbb{P}_n(\texttt{True}), \text{ when } n \to \infty.$$

Note also that, by symmetry, the family of trees smaller than x has the same generating function. Thus, the above described family of trees has the same limiting ratio as that of all trees computing x and its generating function is given by

$$T_x(z) = \frac{2z^2}{(1-(L_x(z)-z))(1-L_x(z))} - 2z^2.$$

The limiting ratio is then

$$\begin{aligned}\mu_n(\mathcal{T}_x) &\sim 2\rho^2 \left(\frac{1}{(1-(L_x(\rho)-\rho))^2(1-L_x(\rho))} + \frac{1}{(1-(L_x(\rho)-\rho))(1-L_x(\rho))^2} \right) \\ &\quad \cdot \lim_{z\to\rho} \frac{L_x'(z)}{A'(z)} \\ &\sim 4\rho^2\, \mathbb{P}_n(\texttt{True}), \text{ as } n \to \infty.\end{aligned}$$

Thus,

$$\mathbb{P}_n(x) \sim \mu_n(\mathcal{X}) \sim \frac{\mathbb{P}_n(\texttt{True})}{n^2}, \text{ as } n \to \infty. \qquad \square$$

9. GENERAL CASE: MINIMAL TREES AND EXPANSIONS

In this last section we prove the last statement of Theorem 2.6. We use different *expansions* of trees, as it was done in other random Boolean tree models (*cf.* [12] for implication random trees and [16] for and/or trees). The first subsection defines the expansions, the second subsection states an asymptotic lower bound for $\mathbb{P}_n(f)$, and the third subsection states an asymptotic upper bound and thus completes the proof of Theorem 2.6.

Convention: *Throughout the whole section we denote by f a non-constant Boolean function.*

9.1. Expansions.

Definition 9.1. *Let t be an associative tree. The tree given by adding a new subtree τ to an internal node ν of t is called an* **expansion** *of t. An expansion is* **valid** *if the expanded tree computes the same function as t.*

- *The expansion is called a* **tautology expansion** *(resp. a contradiction expansion) if the added tree τ is a tautology (resp. a contradiction) and if ν is labelled by a $\wedge$ (resp. $\vee$). Obviously, such an expansion is valid.*
- *It is called a* **$\hat{B}$-expansion** *if the added tree τ is not a single leaf.*

Given a family of trees $\mathcal{T}$, we denote by $E(\mathcal{T})$ the set of trees obtained by a single tautology expansion of a tree in $\mathcal{T}$, by $E^k(\mathcal{T})$ the set of trees obtained by k successive tautology expansions done at (not necessarily distinct) vertices of a tree in $\mathcal{T}$, and by $E^{\geq k}(\mathcal{T})$ the set of all trees obtained by at least k successive tautology expansions done at (not necessarily distinct) vertices of a tree in $\mathcal{T}$. Finally, we set $\overline{E}(\mathcal{T}) := \bigcup_{k\geq 1} E^k(\mathcal{T})$.

If the considered expansions are $\hat{B}-$expansions, we change the above notation by replacing E by $E_{\hat{B}}$.

Remark: Whatever type of expansion (tautology or $\hat{B}$) we consider, note that nesting expansions (adding τ to t, then expanding τ, and so on) does not generate new structures, since this can always be realized by a single expansion. Therefore, requiring that the expansions are done at the vertices of the original tree is no restriction.

Remark: For every family $\mathcal{T}$ of trees the inclusion $\overline{E}(\mathcal{T}) \subseteq \overline{E}_{\hat{B}}(\mathcal{T})$ holds.

9.2. **Tautology expansions.** The following proposition states that the limiting ratio of the set of all trees obtained by expanding minimal trees of a given Boolean function f is of order $1/n^{L(f)}$. Recall that $\mathcal{M}_f$ stands for the set of all minimal trees of f (see Definition 2.3).

Proposition 9.2. *For n tending to infinity, the limiting ratio of $\overline{E}(\mathcal{M}_f)$ is asymptotically equal to the limiting ratio of $E(\mathcal{M}_f)$. Furthermore, if i_t denotes the number of internal nodes of a tree t and $|t|$ its size, then*

$$\mu_n(E(\mathcal{M}_f)) \sim \frac{\lambda_f}{n^{L(f)}}, \text{ as } n \to \infty, \tag{13}$$

where

$$\lambda_f = \frac{\alpha}{2^{L(f)}} \sum_{t \in \mathcal{M}_f} (i_t + |t| - 1)$$

with α defined by $\mathbb{P}_n(\texttt{True}) = \alpha + o(1)$. Moreover, the constant λ_f can be estimated by (3). Since every tree in $\overline{E}(\mathcal{M}_f)$ computes f, this implies

$$\mathbb{P}_n(f) \geq \mu_n(\overline{E}(\mathcal{M}_f)) \sim \frac{\lambda_f}{n^{L(f)}}.$$

Proof. Let $\Phi_k(z)$ be the generating function of $E^k(\mathcal{M}_f)$. Given a tautology τ, the number of places where τ can be added to a given minimal tree t is

$$P_t = \sum_{i \text{ internal node of } t} (d(i) + 1) = i_t + |t| - 1, \tag{14}$$

where $d(i)$ is the number of children of the internal node i. Since t is minimal (*i.e.* $|t| = L(f)$) and since $1 \leq i_t \leq \lfloor \frac{L(f)}{2} \rfloor$, we have $L(f) \leq P_t \leq \frac{3L(f)}{2}$ which yields

$$m_f z^{L(f)} L(f) T(z) \prec \Phi_1(z) = \sum_{t \in \mathcal{M}_f} z^{L(f)} P_t T(z) \prec m_f z^{L(f)} \frac{3L(f)}{2} T(z), \tag{15}$$

where $m_f = |\mathcal{M}_f|$. Thus

$$m_f L(f) \rho^{L(f)} \lim_{z \to \rho} \frac{T'(z)}{A'(z)} \leq \mu_n(E(\mathcal{M}_f)) \leq m_f \frac{3L(f)}{2} \rho^{L(f)} \lim_{z \to \rho} \frac{T'(z)}{A'(z)}. \tag{16}$$

In view of Theorem 2.6(i), we have $\lim_{z \to \rho} \frac{T'(z)}{A'(z)} = \mathbb{P}_n(\texttt{True}) = \alpha + o(1)$ and since $\rho \sim \frac{1}{2n}$ when n tends to infinity. When using the bounds given in (15) we get $\mu_n(E(\mathcal{M}_f)) = \Theta\left(n^{-L(f)}\right)$, whereas plugging these results into the exact expression for $\Phi_1(z)$, given also in (15), yields (13).

If we do k successive expansions in a minimal tree, we have at most $\lfloor 3L(f)/2 \rfloor$ different places for the first one, $\lfloor 3L(f)/2 \rfloor + 1$ for the second one, and so on. We thus have the following inequality:

$$\Phi_k(z) \prec m_f z^{L(f)} \binom{\lfloor 3L(f)/2 \rfloor + k - 1}{k} T(z)^k$$

and thus

$$\mu_n(E^k(\mathcal{M}_f)) = \lim_{z\to\rho} \frac{\Phi_k'(z)}{A'(z)} \le m_f \rho^{L(f)} \binom{\lfloor 3L(f)/2 \rfloor + k - 1}{k} kT(\rho)^{k-1} \lim_{z\to\rho} \frac{T'(z)}{A'(z)}$$
$$\le \beta m_f \rho^{L(f)} \binom{\lfloor 3L(f)/2 \rfloor + k - 1}{k} kT(\rho)^{k-1},$$

for all $n \ge 0$. Hence

$$\mu_n(E^{k\ge 2}(\mathcal{M}_f)) \le \beta m_f \rho^{L(f)} \lfloor 3L(f)/2 \rfloor \left(\frac{1}{(1-T(\rho))^{\lfloor 3L(f)/2 \rfloor + 1}} - 1 \right)$$

where we used $\sum_{k\ge 2} \binom{C+k-1}{k} kz^{k-1} = \frac{C}{(1-z)^{C+1}} - C$ which is an immediate consequence of

$$\sum_{k\ge 0} \binom{C+k-1}{k} z^k = \frac{1}{(1-z)^C}. \tag{17}$$

Since $T(\rho) \le \hat{B}(\rho) \le 1/\sqrt{2n}$ (a tautology cannot be a single leaf) and $\frac{C}{(1-z)^{C+1}} - C = O(z)$, we obtain

$$\mu_n(E^{\ge 2}(\mathcal{M}_f)) = \mathcal{O}\left(\frac{1}{n^{L(f)+1/2}} \right).$$

By (16) the same calculations yield the lower bound. □

9.3. **Irreducible trees.**

Definition 9.3. *Let t be a tree computing f. If t cannot be obtained by a tautology expansion of a smaller tree computing f, then t is called an* **irreducible** *tree of f. We denote by $\mathcal{I}_f$ the set of irreducible trees of f which are not minimal trees of f.*

Take a tree computing f and simplify it according to tautology expansions until it is irreducible. The simplified tree is either in $\mathcal{M}_f$ or in $\mathcal{I}_f$. Thus

$$\mathbb{P}_n(f) \le \mu_n(\overline{E}(\mathcal{M}_f)) + \mu_n(\overline{E}(\mathcal{I}_f)). \tag{18}$$

Proposition 9.4. *We have the following asymptotic result:*

$$\mu_n\left(\overline{E}(\mathcal{I}_f)\right) = o\left(\frac{1}{n^{L(f)}} \right).$$

To prove this proposition, we have to better understand the shape of an irreducible tree of f. For that reason we define a procedure for how to "zip" a (not necessarily irreducible) tree as follows:

- Assign all inessential variables of f[1] to `True`, and substitute each occurrence of an inessential variable in t by its evaluation (either `True`, or `False` when it occurs as a negation of the variable), and then
- simplify the tree as follows: As soon as a leaf is assigned to `True` (resp. `False`) and its parent is $\wedge$ (resp. $\vee$), we cut the leaf. If its parent is $\vee$ (resp. $\wedge$), we cut the subtrees attached to this $\vee$ (resp. $\wedge$), *i.e.* at the parent.

Remark: The order in which the leaves are processed does not influence the final result of the zipping. Only the number of cuttings and their type (leaf or larger subtree is cut) is affected, since cutting a large tree first may result in many leaves not being processed any more, because they disappeared with the large cut tree. However, we will need to impose a certain order (explained where appropriate) in some arguments later on.

Since f is not constant, zipping cannot make the tree vanish. The obtained tree, denoted by $t^\star$ contains no inessential variable and still computes f. It possibly has internal nodes with a single child (called unary nodes), and connectives labelling some leaves (where $\vee$-leaves compute `False` and $\wedge$-leaves compute `True`), the other leaves being labelled by essential variables of f. But, since

[1] A variable x is an inessential variable of f if the restriction of f on the subset $\{x = \texttt{True}\}$ is equal to its restriction on the subset $\{x = \texttt{False}\}$

this tree computes f, it can be seen that its size cannot be smaller than $L(f)$. Indeed, a tree with unary nodes and leaves labelled by connectives can be simplified such that we obtain a proper and/or tree that still computes f, *i.e.* with at least $L(f)$ nodes, and this simplification process reduces the number of nodes. The tree $t^\star$ belongs to the family of *zipped trees* defined hereafter.

Definition 9.5. *Let us denote by $\mathcal{Z}$ the set of trees with internal nodes labelled by $\wedge$ and $\vee$ in a stratified way and with leaves labelled by essential variables of f (or their negations) or $\wedge$ and $\vee$ (again in compliance with the stratification), in which internal nodes can have one child or more. We call these trees* **zipped** *trees and denote by Z_ℓ the number of such trees of size ℓ (the size being the total number of nodes of the zipped tree).*

Remark: By symmetry under relabelling of the variables, one can see that the number Z_ℓ only depends on ℓ and γ (the number of essential variables of f), but not on the set Γ_f of essential variables of f.

During the simplification process, we have cut leaves or non-leaf subtrees; if two or more simplifications are nested, we only consider the largest one. We will prove in the following lemmas that the family of trees such that the simplification process cuts no non-leaf tree and the family of trees in which we have cut "many" single leaves are negligible compared to the family of trees computing f (the limiting ratio of which is at least of order $1/n^{L(f)}$).

Lemma 9.6. *Let $\Gamma \subset \{x_1, \dots, x_n\}$ be of cardinality γ and set $Y = \{x_1, \dots, x_n\} \setminus \Gamma$. Moreover, let $\mathcal{N}_\Gamma$ be the family of trees such that no node labelled by $\vee$ (resp. $\wedge$) has a leaf labelled by a positive (resp. negated) variable from Y as a child. Then $\mu_n(\mathcal{N}_\Gamma) = 0$, for all large enough n.*

Moreover, the following result for expansions of trees in $\mathcal{N}_\Gamma$ holds: Let $\mathcal{N}_\Gamma(f)$ denote the set of all trees in $\mathcal{N}_\Gamma$ which compute f, but are not elements of $\mathcal{M}_f$. Then $\mu_n(\bar{E}(\mathcal{N}_\Gamma)) = o\left(1/n^{L(f)+1}\right)$.

Remark: Note that $\mu_n(\mathcal{N}_\Gamma)$ is a limit for m tending to infinity. So, the assertion can be rephrased as follows: If n is sufficiently large, then, among all trees of size m, the proportion of trees which lie in $\mathcal{N}_\Gamma$ tends to zero, as m tends to infinity.

Remark: The family $\mathcal{N}_\Gamma$ contains the family of trees computing a function f having Γ as its set of essential variables and such that the simplification process only cuts leaves and no non-leaf tree.

Proof. The family $\mathcal{N}_\Gamma$ has the same limiting ratio as associative trees in which leaves are labelled by literals from a set of cardinality $2n - (n - \gamma) = n + \gamma$. Therefore, the singularity ν_n of the generating function of this family satisfies $\nu_n \sim 1/n+\gamma$ and is thus strictly larger than $\rho \sim 1/2n$ for all large enough n. This implies the assertion.

To prove the second assertion we use the idea of the proof of Proposition 9.2. Let $\Phi_k(z)$ and $N(z)$ denote the generating function of $E^k(\mathcal{N}_\Gamma)$ and $\mathcal{N}_\Gamma(f)$, respectively. Then,

$$\Phi_1(z) = \sum_{t \in \mathcal{N}_\Gamma(f)} z^{|t|} P_t T(z)$$

where P_t, given by (14), is the number of places in the tree t where a tautology expansion is possible. It satisfies $P_t \leq 3|t|/2$. This implies

$$\Phi_1(z) \prec \frac{3}{2} \sum_{n \geq L(f)+1} n z^n T(z)[z^n]N(z) = \frac{3}{2} T(z) z N'(z)$$

and consequently

$$\begin{aligned}\mu_n(E(\mathcal{N}_\Gamma(f)) &= \lim_{z \to \rho} \frac{\Phi_1'(z)}{A'(z)} \\ &= \frac{3}{2}\rho N'(\rho) \lim_{z \to \rho} \frac{T'(z)}{A'(z)} + \frac{3}{2} T(\rho) \lim_{z \to \rho} \frac{(zN'(z))'}{A'(z)}\end{aligned}$$

Since $N(z)$ is a power series divisible by $z^{L(f)+1}$ the order of magnitude of the first term is $\mathcal{O}\left(1/(2n)^{L(f)+1}\right)$. The dominant singularity of $N(z)$ (and so the one of $zN'(z)$) is ν_n, which is strictly smaller than ρ. Hence the second term is equal to zero.

If we do $k > 1$ expansions, we get, as in the proof of Proposition 9.2,

$$\Phi_k(z) = T(z)^k \sum_{n \geq L(f)+1} \binom{\frac{3n}{2} + k - 1}{k} z^n [z^n] N(z).$$

This expression can be written as a sum of terms of the form $T(z)\left(\frac{\mathrm{d}}{\mathrm{d}z}\right)^j N(z)$, where $j \leq k$, and the argument from above apply again and yield $\mu_n(E(\mathcal{N}_\Gamma(f)) = \mathcal{O}\left(1/(2n)^{L(f)+1}\right)$. □

The following lemma ensures that we have cut only a few "single" leaves:

Lemma 9.7. *Let ℓ be an integer and $\Gamma \subset \{x_1, \ldots, x_n\}$ be of cardinality γ. Set $Y = \{x_1, \ldots, x_n\} \backslash \Gamma$ and let us denote by $\mathcal{M}_\ell^\Gamma$ the family of trees with at least ℓ leaves labelled by variables from Y (or their negations) such that none of these leaves has an ancestor labelled by $\vee$ (resp. $\wedge$) which has a child being a leaf labelled by a positive (resp. negated) variable from Y. Then $\mu_n(\mathcal{M}_\ell^\Gamma) = o\left(1/n^{\ell+1}\right)$.*

Remark: Note that the family $\mathcal{M}_\ell^\Gamma$ contains the family of irreducible trees computing a function f having Γ as its set of essential variables and in which the simplification process cuts at least ℓ single leaves. It contains all expansions of such irreducible trees as well since $\bar{E}(\mathcal{M}_\ell^\Gamma) \subseteq \mathcal{M}_\ell^\Gamma$.

Proof. We will prove the assertion by constructing a superset of $\mathcal{M}_\ell^\Gamma$ which will be shown to satisfy the assertion as well. This will be done in two steps. First, we construct a set of trees which contains all trees in $\mathcal{M}_\ell^\Gamma$ where the ℓ distinguished leaves have pairwise distinct parent nodes. The second case, where at least two of the distinguished leaves have the same parent node, is then treated in a similar fashion.

Let us consider the family of trees obtained as follows (see Figure 2):

(1) Take a rooted unlabelled tree t_0 having ℓ leaves and no nodes of arity 1. Then $\ell + 1 \leq |t_0| \leq 2\ell - 1$;
(2) add to each internal node some subtrees (with internal nodes unlabelled and leaves labelled by literals) which are not single leaves labelled by a variable from Y or its negation;
(3) replace each edge by a sequence of internal (unlabelled) nodes with subtrees (with internal nodes unlabelled and leaves labelled by literals) attached to them which are not single leaves labelled by a positive (resp. negated) variable from Y if their parent is $\vee$ (resp. $\wedge$);
(4) choose a label ($\wedge$ or $\vee$) for the root and deduce the labels of all internal nodes, knowing that a node and its child cannot have the same label (stratification property);
(5) finally, replace each leaf of t_0 by a tree rooted by $\wedge$ (resp. $\vee$ according to the stratification) with at least one literal from Y (resp. from the negations of Y) and no literal from the negations of Y (resp. from Y) on the first level.

The obtained family contains all trees from $\mathcal{M}_\ell^\Gamma$ where the ℓ distinguished leaves have altogether ℓ parent nodes. Its generating function $F_1(z)$ is given by

$$F_1(z) = 2C_\ell \sum_{r=\ell}^{2\ell-2} z^{r+1-\ell} (zX(z))^\ell V(z)^r W(z)^{2r+1-\ell},$$

where

- the index r in the summation represents the number of edges of the tree t_0 chosen in the construction;
- the factor C_ℓ is the number of choices for this t_0 and the factor 2 for its root label;
- the factor $z^{r+1-\ell}$ marks the internal nodes of t_0;
- the function

$$zX(z) = \frac{(n-\gamma)z^2}{(1 - (\hat{A}(z) - 2(n-\gamma)z))(1 - (\hat{A}(z) - (n-\gamma)z))} - (n-\gamma)z^2$$

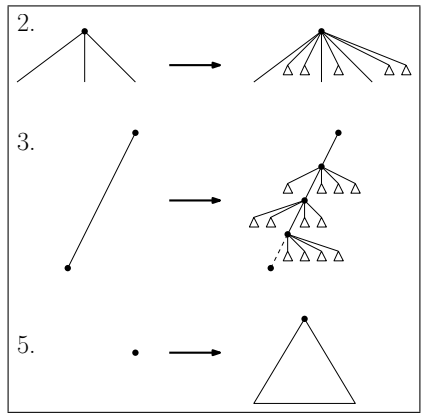

FIGURE 2. Proof of Lemma 9.7.

is the generating function of the set of trees rooted by $\wedge$ with at least one literal from Y and no literal from the negations of Y on the first level;

- the function

$$V(z) = \frac{1}{1 - z\left(\left(\frac{1}{1-(\hat{A}(z)-(n-\gamma)z)}\right)^2 - 1\right)}$$

is the generating function of the sequences of internal nodes that replace the edges of t_0: It is the generating function of (possibly empty) sequences of one internal node marked by z and two sequences (one on the left and one on the right of the existing edge - that cannot be both empty) of subtrees being different from a leaf labelled by a positive variable of Y; and

- the function

$$W(z) = \frac{1}{1 - (\hat{A}(z) - (n-\gamma)z)}$$

is the generating function of the sequences of trees which are different from a single leaf with a label from Y. Note that such sequences are attached in the 3rd step and they can be placed left from every edge in t_0 (r choices) or to the right of the rightmost child of any internal node of t_0 ($r + 1 - \ell$ choices).

To estimate the limiting ratio of this family, observe that the singularity of F_1 is ρ and that it is a squareroot singularity. Therefore, the limiting ratio of this family can be computed by Lemma 4.2, *i.e.* by $\lim_{z\to\rho} \frac{F_1'(z)}{A'(z)}$. To compute this limiting ratio, let us note that $\lim_{z\to\rho} \frac{1}{A'(z)} = 0$ such that many terms in $F_1'(z)/A'(z)$ can be neglected (*cf.* last paragraph in the proof of Proposition 4.3). We obtain:

$$\begin{aligned}\lim_{z\to\rho} \frac{F_1'(z)}{A'(z)} = 2C_\ell \sum_{r=\ell}^{2\ell-2} \rho^{r+1} \Bigg(& \ell X(\rho)^{\ell-1} V(\rho)^r W(\rho)^{2r+1-\ell} \lim_{z\to\rho} \frac{X'(z)}{A'(z)} \\ & + rX(\rho)^\ell V(\rho)^{r-1} W(\rho)^{2r+1-\ell} \lim_{z\to\rho} \frac{V'(z)}{A'(z)} \\ & + (2r+1-\ell) X(\rho)^\ell V(\rho)^r W(\rho)^{2r-\ell} \lim_{z\to\rho} \frac{W'(z)}{A'(z)} \Bigg).\end{aligned}$$

Recall that $\rho \sim 1/2n$ and $\hat{A}(\rho) \sim 1 - 1/\sqrt{2n}$, as $n \to \infty$, and hence $X(\rho) \sim 1/3$, $V(\rho) \sim 1$ and $W(\rho) \sim 2$, as $n \to \infty$. Moreover,

$$\begin{aligned}\lim_{z\to\rho} \frac{W'(z)}{A'(z)} &= \lim_{z\to\rho} \frac{\hat{A}'(z) - (n-\gamma)}{A'(z)} \frac{1}{(1-(\hat{A}(\rho)-(n-\gamma)\rho))^2} \\ &= \frac{1}{2(1-(\hat{A}(\rho)-(n-\gamma)\rho))^2},\end{aligned}$$

because $\lim_{z\to\rho} \frac{\hat{A}'(z)}{A'(z)} = \frac{1}{2}$ and $\lim_{z\to\rho} \frac{n-\gamma}{A'(z)} = 0$. Thus, $\lim_{n\to\infty} \lim_{z\to\rho} \frac{W'(z)}{A'(z)} = 2$. Using similar arguments, we can prove that

$$\lim_{n\to\infty} \lim_{z\to\rho} \frac{V'(z)}{A'(z)} = 8 \quad \text{and} \quad \lim_{z\to\rho} \frac{X'(z)}{A'(z)} \sim \frac{8}{9n}, \text{ as } n \to \infty.$$

All these relations imply

$$\lim_{z\to\rho} \frac{F_1'(z)}{A'(z)} \sim \frac{\kappa_1}{n^{\ell+1}}, \text{ as } n \to \infty,$$

where κ_1 is a positive constant.

Now let us turn to the second case. We construct a tree family containing all trees in $\mathcal{M}_\ell^\Gamma$ where the ℓ distinguished leaves have altogether ℓ_0 parent nodes, with $1 \le \ell_0 < \ell$. This can be done by two obvious modifications: In the first step we start with a tree having ℓ_0 nodes and the last step is replaced by

5'. Replace each leaf of t_0 by a tree rooted by $\wedge$ (resp. $\vee$ according to the stratification) with at least one literal from Y (resp. from the negations of Y) and no literal from the negations of Y (resp. from Y) on the first level, such that the total number of literals from Y and from the negations of Y that are generated is at least ℓ.

Then the generating function $F_2(z)$ of this family is given by

$$F_2(z) = \begin{cases} f_2(z) & \text{if } \ell_0 > 1, \\ z^\ell X_\ell(z) & \text{if } \ell_0 = 1, \end{cases}$$

where

$$z^j X_j(z) = \frac{(n-\gamma)z^{j+1}}{(1-(\hat{A}(z)-2(n-\gamma)z))^j(1-(\hat{A}(z)-(n-\gamma)z)))} - (n-\gamma)z^2$$

is the generating function of the set of trees rooted by $\wedge$ with at least j literals from Y and no literal from the negations of Y on the first level and

$$f_2(z) = 2C_{\ell_0} \sum_{r=\ell_0}^{2\ell_0-2} z^{r+1-\ell_0} \sum_{\substack{m_1,\dots,m_{\ell_0}\ge 1 \\ m_1+\cdots+m_{\ell_0}=\ell}} \left(\prod_{j=m_1}^{m_{\ell_0}} z^j X_j(z) \right) V(z)^r W(z)^{2r+1-\ell}.$$

Similar calculations as above show that

$$\lim_{z\to\rho} \frac{F_2'(z)}{A'(z)} \sim \frac{\kappa_2}{n^{\ell+1}}, \text{ as } n \to \infty,$$

with a suitable positive constant κ_2, which completes the proof. □

We are now ready to prove Proposition 9.4. The two previous lemmas allow us to consider only expansions of irreducible trees in which the zipping process cuts at least one non-leaf tree and fewer than $L(f)$ single leaves. Note that we have to be more precise here: It may happen that we cut many single leaves which form a whole subtree, which has eventually been cut. But we could have started with cutting the large subtree and then all its leaves to be cut are already gone. To avoid ambiguity we require that, whenever cuttings form a set of nested simplifications, we cut the largest first, i.e. cutting all of them is accomplished by one single cut. Let us denote by $\mathcal{I}_1$ the set of irreducible trees of f for which the simplified tree $t^\star$ has size $L(f)$, and by $\mathcal{I}_2$ the set of irreducible trees t of f such that $t^\star$ has size at least $L(f)+1$.

Lemma 9.8. *We have the following asymptotic result:* $\mu_n\big(\overline{E}(\mathcal{I}_1)\big) = o\left(1/n^{L(f)}\right)$.

Proof. Let $t \in \mathcal{I}_1$. To obtain $t^\star$, we have cut subtrees of t and we can assume that we have cut at least one large subtree and at most $L(f)$ single leaves. Otherwise t and all its expansions would be in one of the classes treated in Lemmas 9.6 and 9.7 and these lemmas already tell us that those classes are small.

Assume first that we have cut only one non-leaf subtree rooted at a node ν during the algorithm. Then, either this tree contains an essential variable on its first-level leaves, or it belongs to $\bigcup_{k,\ell\geq 0}\bigcup_{r=0}^{\ell-1}\mathcal{M}_{k,\ell,r}^{\Gamma_f}$, where Γ_f is the set of essential variables of f (see Definition 5.1 for the definition of $\mathcal{M}_{k,\ell,r}^{\Gamma_f}$). In this case the set inclusion $\bar{E}(\{t\}) \subseteq \bigcup_{k,\ell\geq 0}\bigcup_{r=0}^{\ell-1}\mathcal{M}_{k,\ell,r}^{\Gamma_f}$ holds. Otherwise, we could find an assignment of inessential variables such that we can cut the parent of ν without changing the function computed by the tree t. This new assignment of inessential variables leads to a different simplification of the tree t that will cut at least the single leaves that were cut before, plus the larger subtree and its parent: We thus obtain a tree of size less than $L(f)$ that computes f, which is impossible.

Any tree of $\mathcal{I}_1$ is obtained by expanding a zipped tree s of $\mathcal{Z}_{L(f)}$ (see Definition 9.5) as follows:

- choose an integer $q \geq 1$ (q represents the number of large trees that were cut during the process described beforehand: $q \geq 1$ holds because of the remark before Lemma 9.8),
- if $q = 1$, plug a tree from $\left(\bigcup_{k,\ell\geq 0}\bigcup_{r=0}^{\ell-1}\mathcal{M}_{k,\ell,r}^{\Gamma_f}\right) \cup \mathcal{A}_{\Gamma_f}$ at a node of s and at most $L(f)$ inessential leaves at other nodes of s (where $\mathcal{A}_{\Gamma_f}$ is the set of trees containing at least one first-level leaf labelled by an essential variable of f, *cf.* Proposition 4.4),
- else plug $q \geq 2$ non-leaf subtrees and at most $L(f)$ inessential leaves at nodes of s.

We are interested in expansions of trees from $\mathcal{I}_1$. In fact, since we do not impose any restrictions on the trees we plug at nodes of s, we can consider only expansions in the nodes of s, since expansions in the plugged trees are then already counted.

The generating function of trees obtained by successive expansions of trees from $\mathcal{I}_1$, denoted by $I_1(z)$ thus satisfies:

$$I_1(z) \prec Z_{L(f)} z^{L(f)} \sum_{k\geq 0}\left(\binom{3L(f)+1+k}{L(f)+1,k,2L(f)} M(z)(2(n-\gamma)z)^{L(f)}\hat{B}(z)^k \right.$$
$$\left. + \sum_{q\geq 2}\binom{3L(f)+q+k}{L(f)+q,k,2L(f)}\hat{B}(z)^{q+k}(2(n-\gamma)z)^{L(f)}\right),$$

where k counts the number of successive expansions done into the irreducible tree, and where $M(z)$ is the generating function of $\left(\bigcup_{k,\ell\geq 0}\bigcup_{r=0}^{\ell-1}\mathcal{M}_{k,\ell,r}^{\Gamma_f}\right) \cup \mathcal{A}_\gamma$. The multinomial coefficient represents the number of choices for the places where we plug trees in s and where we then do the expansions (the orders of the "pluggings" and of the expansions do not matter, but expansions are done after the "pluggings").

The coefficient $\binom{3L(f)+1+k}{L(f)+1,k,2L(f)}$ counts the number of different ways to plug one tree from $\left(\bigcup_{k,\ell\geq 0}\bigcup_{r=0}^{\ell-1}\mathcal{M}_{k,\ell,r}^{\Gamma_f}\right) \cup \mathcal{A}_\gamma$, $L(f)$ inessential leaves and k $\hat{B}$-expansions into $s \in \mathcal{Z}_{L(f)}$. Let us start from s, the first plugging can be done at the left of each edge of s, or at the right of every rightmost edge, or at every leaf. Note that $s \in \mathcal{Z}_{L(f)}$ has at most $L(f)$ leaves and $L(f)$ edges. The number of different places for the first plugging is thus at most $3L(f)$. The second plugging can then be made in $3L(f)+1$ different places, and so on, and so forth. The fact that we first do the $L(f)+1$ plugging and then the k expansions gives the multinomial coefficient. The second binomial coefficient is given by the same reasoning with $L(f)+q$ pluggings and k expansions.

Thanks to Proposition 4.4 and Section 5, we know that the limiting ratio of the set $\left(\bigcup_{k,\ell\geq 0}\bigcup_{r=0}^{\ell-1}\mathcal{M}_{k,\ell,r}^{\Gamma_f}\right) \cup \mathcal{A}_{\Gamma_f}$ has order $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$ and that $M(\rho) \leq \hat{B}(\rho) \sim \frac{1}{\sqrt{2n}}$. Thus, when

n tends to infinity,

$$\mu_n(\overline{E}(\mathcal{I}_1)) \leq Z_{L(f)}\rho^{L(f)}\hat{B}(\rho)\sum_{k\geq 0}\left(\binom{3L(f)+1+k}{L(f)+1,k,2L(f)}\left(\hat{B}(\rho)^k+\frac{k}{2}\hat{B}(\rho)^{k-1}\right)\right.$$
$$\left.+\sum_{q\geq 2}\binom{3L(f)+q+k}{L(f)+q,k,2L(f)}\frac{q+k}{2}\hat{B}(\rho)^{q+k-2}\right). \tag{19}$$

The first factor $Z_{L(f)}\rho^{L(f)}\hat{B}(\rho)$ behaves as $\frac{1}{n^{L(f)+1/2}}$ when n tends to infinity. Let us prove that the second term of the sum behaves as $\mathcal{O}(1)$ when n tends to infinity. Let us first focus on

$$\sum_{k\geq 0}\binom{3L(f)+1+k}{L(f)+1,k,2L(f)}\hat{B}(\rho)^k = \frac{(3L(f)+1)!}{(L(f)+1)!(2L(f))!}\sum_{k\geq 0}\binom{3L(f)+1+k}{k}\hat{B}(\rho)^k$$
$$= \frac{(2L(f)+2)!}{(L(f)+1)!(2L(f))!}\frac{1}{(1-\hat{B}(\rho))^{3L(f)+2}}$$

in view of (17). Very similar calculations lead to

$$\sum_{k\geq 0}\binom{3L(f)+1+k}{L(f)+1,k}\frac{k}{2}\hat{B}(\rho)^{k-1} = \frac{(3L(f)+1)!}{2(3L(f)+1)!(2L(f)!)}\frac{3L(f)+2}{(1-\hat{B})^{3L(f)+3}}.$$

Moreover, using (17) again,

$$\sum_{k\geq 0}\sum_{q\geq 2}\binom{3L(f)+q+k}{L(f)+q,k,2L(f)}\frac{q}{2}\hat{B}(\rho)^{q+k-2}$$
$$= \sum_{q\geq 2}\frac{(3L(f)+q)!}{(L(f)+q)!(2L(f))!}\frac{q}{2}\hat{B}(\rho)^{q-2}\sum_{k\geq 0}\binom{3L(f)+q+k}{k}\hat{B}(\rho)^k$$
$$= \frac{1}{2(1-\hat{B}(\rho))^{3L(f)+3}}\sum_{q\geq 0}\binom{3L(f)+q+2}{L(f)+q+2}(q+2)\left(\frac{\hat{B}(\rho)}{(1-\hat{B}(\rho))}\right)^q$$

Simplifying the binomial coefficient gives

$$\sum_{k\geq 0}\sum_{q\geq 2}\binom{3L(f)+q+k}{L(f)+q,k,2L(f)}\frac{q}{2}\hat{B}(\rho)^{q+k-2}$$
$$= \frac{1}{2(1-\hat{B}(\rho))^{3L(f)+3}}\sum_{q\geq 0}\left[\prod_{j=2}^{L(f)+2}\left(\frac{2L(f)+q+j}{q+j}\right)\right]$$
$$\cdot\binom{2L(f)+q+1}{q+1}(q+2)\left(\frac{\hat{B}(\rho)}{1-\hat{B}(\rho)}\right)^q$$
$$\leq \frac{(1+2L(f))^{L(f)+1}}{2(1-\hat{B}(\rho))^{3L(f)+3}}\sum_{q\geq 0}\binom{2L(f)+q+1}{q+1}(q+2)\left(\frac{\hat{B}(\rho)}{1-\hat{B}(\rho)}\right)^q$$
$$= \frac{(1+2L(f))^{L(f)+1}}{2(1-\hat{B}(\rho))^{3L(f)+3}}\left(\frac{2L(f)+1}{\left(1-\frac{\hat{B}(\rho)}{1-\hat{B}(\rho)}\right)^{2L(f)+2}}+\frac{1}{\left(1-\frac{\hat{B}(\rho)}{1-\hat{B}(\rho)}\right)^{2L(f)+1}}\right),$$

since $\frac{\hat{B}(\rho)}{1-\hat{B}(\rho)}$ is smaller than 1 for all large enough n. Similar calculations can be done for the last term of the sum (19), and we eventually get

$$\mu_n\big(\overline{E}(\mathcal{I}_1)\big) = \mathcal{O}\left(\frac{1}{n^{L(f)+1/2}}\right). \qquad \square$$

Lemma 9.9. *We have the following asymptotic result:* $\mu_n\big(\overline{E}(\mathcal{I}_2)\big) = o\left(1/n^{L(f)}\right)$.

Proof. The generating function $I_2(z)$ of trees obtained by successive expansions of a tree from $\mathcal{I}_2$ satisfies

$$I_2(z) \leq \sum_{\ell \geq L(f)+1} Z_\ell z^\ell \sum_{k\geq 0}\sum_{q\geq 0} \binom{2\ell + L(f) + q + k}{L(f)+q, k, 2\ell} \hat{B}(z)^{k+q}(2(n-\gamma)z)^{L(f)}$$

where Z_ℓ is the number of zipped trees of size ℓ, where ℓ is the size of the seminal tree in $\mathcal{I}_2$, and where the multinomial coefficient counts the numbers of different ways to plug first $L(f)$ leaves and q non-leaf trees and then k expansions in this seminal tree. Thus, calculations in the same vein as those done in the proof of Lemma 9.8 yield

$$\mu_n\big(\overline{E}(\mathcal{I}_2)\big) \leq \sum_{\ell \geq L(f)+1} Z_\ell \rho^\ell \sum_{k\geq 0}\sum_{q\geq 0} \binom{2\ell + L(f) + q + k}{L(f)+q, k, 2\ell} \frac{k+q}{2} \frac{1}{\sqrt{n}^{k+q-1}}.$$

Thus, we conclude

$$\mu_n\big(\overline{E}(\mathcal{I}_2)\big) \leq \mathcal{O}\left(\frac{1}{n^{L(f)+1}}\right). \qquad \square$$

Lemmas 9.8 and 9.9 directly induce Proposition 9.4, and we are now able to complete the proof of Theorem 2.6.

Proof of Theorem 2.6. The probabilities of the set of tautologies and that of literals are treated in Sections 6 and 8, respectively. We now have to prove that for all Boolean functions f, $\mathbb{P}_n(f) \sim \lambda_f/n^{L(f)}$. In view of Proposition 9.2, we have a lower bound which is asymptotically equivalent to $\lambda_f/n^{L(f)}$, and (18) together with Proposition 9.4 gives an upper bound which differs from the lower bound by an additive term of magnitude $\mathcal{O}\left(1/n^{L(f)+1}\right)$. This completes the proof of Theorem 2.6. $\square$

10. CONCLUSION AND FURTHER WORK

This article presents the first attempt to discuss the size definition in quantitative logics: while the formula size is commonly used in the literature, we believe that in the case of nonbinary trees, the tree size considered in this article is at least as natural.

We proved that this change of size/complexity notion does not affect the first order behaviour of the distribution induced on the set of Boolean functions by the uniform distribution on and/or trees of a given (large) size and labelled on n variables. However, we have also exhibited how the large typical and/or tree gets a very different shape when changing the size notion. Trees having more leaves are more and more likely as the number n of variables increases. This change of typical shape has forced us to develop original, more intricate proofs for this new model.

Considering nonbinary trees is a way to take into account the associativity of the logical connectives $\wedge$ and $\vee$. Note that there is no reason - except technicality - justifying that we consider plane trees instead of nonplane trees while the connectives $\wedge$ and $\vee$ are commutative. Nonplane associative trees have been studied in the formula size model (see [16]), and we believe that the same could be done for the tree size model, although once again, the technical level of the computations would considerably increase since in nonplane models the generating functions are not known explicitly.

Acknowledgements: We are indebted to two anonymous referees for pointing out several imprecisions and for comments which improved the presentation. In particular, we express our gratitude to one of these referees who outlined the proof of Lemma 6.2 which has led to an improvement of our results. Last but not least, we thank Jakub Kozik for the careful reading of the manuscript, many suggestions to improve the presentation and to remove minor errors, and especially for pointing out a subtle mistake concerning the proofs of Lemmas 9.6 and 9.8.

TABLE 1. Notations used in the article for different families of trees.

notation	object	cardinal	generating function	page of definition
$\mathcal{A}_{m,n}$	set of all the n-variable associative trees of size m	$A_{m,n}$	$A(z)$	4
$\mathcal{M}_f$	set of all the n-variable associative trees representing the Boolean function f (see Definition 2.3)	m_f	$m_f z^{L(f)}$	4
$\hat{\mathcal{A}}_{m,n}$	set of all the n-variable $\wedge$-rooted (if $m \geq 2$) associative trees of size m	$\hat{A}_{m,n}$	$\hat{A}(z)$	5
$\check{\mathcal{A}}_{m,n}$	set of the n-variable $\vee$-rooted (if $m \geq 2$) associative trees of size m	$\check{A}_{m,n}$	$\check{A}(z)$	5
$\hat{B}(z)$	generating function of the n-variable $\wedge$-rooted associative trees of size at least 2 (see Proposition 3.1)		$\hat{B}(z)$	7
$\mathcal{A}^{(0)}$	set of the n-variable associative trees with no leaf on the first level		$A^{(0)}(z)$	7
$\mathcal{A}_\Gamma$	set of all n-variable associative trees such that at least one first-level leaf is labelled by a literal from the set Γ		$A_\Gamma(z)$	8
$\mathcal{M}^\Gamma_{k,\ell,r}$	see Definition 5.1		$M(z)$	9
$\mathcal{J}^\Gamma_{\boldsymbol{\alpha},\boldsymbol{\beta}}$	see Definition 5.1		$J(z)$	8
$\mathcal{T}$	set of all n-variable tautologies		$T(z)$	13
$\mathcal{S}$	set of all n-variable simple tautologies (see Definition 6.1)		$S(z)$	13
$\mathcal{N}$	set of all n-variable tautologies that are not simple tautologies ($\mathcal{N} = \mathcal{T} \setminus \mathcal{S}$)			15
$\mathcal{X}$	set of all n-variable simple x (see Definition 8.1)		$X(z)$	13
$\mathcal{I}_f$	set of all n-variable irreducible but not minimal trees of the Boolean function f (see Definition 9.3)			20
$\mathcal{Z}_\ell$	set of zipped trees of size ℓ (see Definition 9.5)	Z_ℓ		21
$\mathcal{N}_\Gamma$	see Lemma 9.6			21
$\mathcal{M}^\Gamma_\ell$	see Lemma 9.7			22

REFERENCES

[1] R. B. Boppana, *Amplification of probabilistic Boolean formulas,* Proceedings of the 26th Annual IEEE Symposium on Foundations of Computer Science (1985), 20–29.

[2] A. Brodsky and N. Pippenger, *The Boolean functions computed by random Boolean formulas or how to grow the right function,* Random Structures and Algorithms, **27**(2005), 490–519.

[3] N. G. de Bruijn, *Asymptotic Methods in Analysis*, Bibliotheca Mathematica, Dover Publications, 1970.

[4] B. Chauvin, P. Flajolet, D. Gardy, and B. Gittenberger, *And/Or trees revisited,* Combinatorics, Probability and Computing, **13(4-5)**(2004), 475–497.

[5] M. Drmota, *Systems of functional equations,* Random Structures and Algorithms, **10(1-2)**(1997),103–124.

[6] M. Drmota, *Random trees*, Springer, Vienna-New York, 2009.

[7] P. Flajolet, D. Gardy, and L. Thimonier, *Birthday Paradox, Coupon Collectors, Caching Algorithms and Self-Organizing Search,* D. A. Math., **39(3)**(1992),207–229.

[8] P. Flajolet and A. M. Odlyzko, *Singularity analysis of generating functions,* SIAM J. Discrete Math., **3**(1990), 216–240.

[9] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, Cambridge, 2009.

[10] H. Fournier, D. Gardy, and A. Genitrini, *Balanced And/Or trees and linear threshold functions,* Proceedings of the 5th SIAM Workshop on Analytic and Combinatorics (2009), 51–57.

[11] H. Fournier, D. Gardy, A. Genitrini, and B. Gittenberger, *Complexity and limiting ratio of Boolean functions over implication,* MFCS (2008), 347–362.

[12] H. Fournier, D. Gardy, A. Genitrini, and B. Gittenberger, *The fraction of large random trees representing a given Boolean function in implicational logic,* Random Structures and Algorithms, **40(3)**(2012), 317–349.

[13] D. Gardy, *Random Boolean expressions,* Colloquium on Computational Logic and Applications, DMTCS, **AF**(2006), 1–36.
[14] A. Genitrini and B. Gittenberger, *No Shannon effect on probability distributions on Boolean functions induced by Boolean expressions,* Proc. of Analysis of Algorithms, DMTCS **AM**(2010), 303–316.
[15] A. Genitrini, B. Gittenberger, V. Kraus, and C. Mailler, *Probabilities of Boolean functions given by random implicational formulas,* Electronic Journal of Combinatorics, **19(2):P37**(2012), 20 pages.
[16] A. Genitrini, B. Gittenberger, V. Kraus, and C. Mailler, *Associative and commutative tree representations for Boolean functions,* Theoretical Computer Science, **570**(2015), 70–101.
[17] A. Genitrini, B. Gittenberger, and C. Mailler, *No Shannon effect induced by And/Or trees,* 25th International Meeting on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms(2014), 109–120.
[18] R. L. Graham, M. Grötschel, and L. Lovász, *Handbook of Combinatorics (Vol. 2)*, MIT Press, Cambridge, MA, USA, 1995.
[19] A. Gupta and S. Mahajan, *Using amplification to compute majority with small majority gates,* Computational Complexity, **6(1)**(1997), 46–63.
[20] J. Kozik. *Subcritical pattern languages for and/or trees,* Fifth Colloquium on Mathematics and Computer Science, DMTCS, **AI**(2008), 437–448.
[21] S. P. Lalley, *Finite range random walk on free groups and homogeneous trees,* The Annals of Probability, **21**(1993).
[22] H. Lefmann and P. Savický, *Some typical properties of large And/Or Boolean formulas,* Random Structures and Algorithms, **10**(1997), 337–351.
[23] O. B. Lupanov, *Complexity of formula realization of functions of logical algebra,* Problems of Cybernetics, Pergamon Press, **3**(1962), 782–811.
[24] O. B. Lupanov, *On the realization of functions of logical algebra by formulae of finite classes (formulae of limited depth) in the basis* $\wedge$, $\vee$, $\neg$, Problems of Cybernetics, Pergamon Press, **6**(1965), 1–14.
[25] J. B. Paris, A. Vencovská, and G. M. Wilmers, *A natural prior probability distribution derived from the propositional calculus,* Ann. Pure Appl. Logic, **70(3)**(1994), 243–285.
[26] J. Riordan and C. E. Shannon, *The number of two terminal series-parallel networks,* Journal of Mathematics and Physics, **21**(1942), 83–93.
[27] P. Savický, *Random Boolean formulas representing any Boolean function with asymptotically equal probability,* Discrete Mathematics, **83**(1990), 95–103.
[28] B. I. Selivanov, *On waiting time in the scheme of random allocation of coloured particles,* Discrete Mathematics and Applications, **5(1)**(1995), 73–82.
[29] C. E. Shannon, *Communication theory of secrecy systems,* Bell System Tech. J., **28**(1949), 656–715.
[30] L. Valiant, *Short monotone formulae for the majority function,* Journal of Algorithms, **5**(1984), 363–366.
[31] A. R. Woods, *Coloring rules for finite trees, and probabilities of monadic second order sentences,* Random Structures and Algorithms, **10(4)**(1997), 453–485.

Veronika Daxner,

Antoine Genitrini, Sorbonne Universités, UPMC Univ Paris 06, CNRS, LIP6 UMR 7606, 4 place Jussieu 75005 Paris. `Antoine.Genitrini@lip6.fr`,

Bernhard Gittenberger, Technische Universität Wien, Wiedner Hauptstraße 8-10/104, 1040 Wien, Austria. `gittenberger@dmg.tuwien.ac.at` and

Cécile Mailler, Department of Mathematical Sciences, University of Bath, BA2 7AY Bath, UK. `c.mailler@bath.ac.uk`