The relation between tree size complexity and probability for Boolean functions generated by uniform random trees.\textsuperscript{*}

Antoine Genitrini\textsuperscript{1}, Bernhard Gittenberger\textsuperscript{2}, Veronika Kraus\textsuperscript{2}, Cécile Mailler\textsuperscript{3}

Abstract

We consider a probability distribution on the set of Boolean functions in \( n \) variables which is induced by random Boolean expressions. Such an expression is a random rooted plane tree where the internal vertices are labelled with connectives And and OR and the leaves are labelled with variables or negated variables. We study limiting distribution when the tree size tends to infinity and derive a relation between the tree size complexity and the probability of a function. This is done by first expressing trees representing a particular function as expansions of minimal trees representing this function and then computing the probabilities by means of combinatorial counting arguments relying on generating functions and singularity analysis.

Keywords: Boolean functions; Probability distribution; Random Boolean formulas; Tree size complexity; Formula size complexity; Analytic combinatorics.

1. Introduction

The interest in random Boolean function dates back to the 1940ies when Riordan and Shannon \cite{23, 25} discovered the so-called Shannon effect: A Boolean function which is drawn uniformly at random from all functions in \( n \) variables has, asymptotically with high probability, exponential complexity. Since then numerous papers are 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 \cite{20, 21} and a proof based on simple combinatorial counting arguments is presented in Flajolet and Sedgewick’s book \cite{7}. 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. The most natural one is of course as follows: Express a Boolean function by a Boolean expression built of variables (with or without negation) and logical connectives like \( \wedge, \vee \) or \( \Rightarrow \). The first efforts were done in \cite{22} and \cite{19} where Boolean expressions were represented by binary plane rooted trees. Lefmann and Savický \cite{19} showed for instance the existence of a limiting distribution when the tree size tends to infinity as well as certain bounds relating the probability and the formula size complexity. The existence of the limiting distribution was shown independently for general formulas by Woods \cite{27}. Woods’ method is basically a special case of a more general result, the Drmota-Lalley-Woods theorem originating in the works \cite{14, 18, 27}. See \cite{4} for an easy accessible formulation, \cite{5} for a detailed discussion, and \cite{10} for an application in the context of Boolean formulas. The bounds of Lefmann and Savický were later refined and extended to further classes of plane binary trees in \cite{3}. A survey of this topic was written by Gardy \cite{11}.

Different but somehow related problems in the framework of balanced trees have been pursued by Valiant \cite{26} who wanted to generate a particular Boolean function with high probability. His results were extended in \cite{11, 10}. The existence of a limiting distribution for balanced And/Or trees can be found in \cite{8} The influence of different connectives was studied in \cite{24, 2}, however, under different distributions.

\textsuperscript{*}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.

\textsuperscript{1}Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, F-75005, Paris, France.
\textsuperscript{2}CNRS, UMR 7606, LIP6, F-75005, Paris, France.
\textsuperscript{3}Antoine.Genitrini@lip6.fr

\textsuperscript{2}Technische Universität Wien, Wieder Hauptstrasse 8-10/104, A-1040 Wien, Austria.
\textsuperscript{2}gittenberger@dmg.tuwien.ac.at

\textsuperscript{3}Department of Mathematical Sciences, University of Bath, BA2 7AY Bath, UK.
\textsuperscript{3}c.mailler@bath.ac.uk
The study of the relation between the probability and the formula size complexity under models similar to the Lefmann-Savický model was resumed recently. It was possible to improve the bounds and to obtain asymptotic relations: See [17] for And/Or trees and [9] for implicational trees. The latter result was improved and extended in [10].

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 [13] and [14] extend the binary plane models to more general tree classes.

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 Boolean expressions, the formula size complexity is exactly the number of leaves of the tree. This is, however, unsatisfactory in a purely mathematical context (i.e. in the context of graph theory) where the size of a tree is usually defined as number of all vertices (except for binary trees where both notions of size are inextricably entangled). Moreover, from a computer science point of view, the size of the storage needed to represent an expression in a computer may also depend on the number of all vertices. This triggered our interest in studying the relation between the tree size complexity and the probability of a function. We did not expect much difference, but first experiments indicated a rather different behaviour. Indeed, when we tried the method which worked for And/Or trees to prove the results we expected, we failed. The reason is that the typical shape as well as some technical properties of the trees in the “tree size” model are rather different from those we encountered in the “formula size” models. Though it will eventually turn out that the trees in the tree size model very well fall under the same paradigms we encountered in many formula size models, it requires a technically very different treatment to obtain those results. Summarizing:

As in the formula size models, there is a strong relation between the (tree size) complexity and the probability of a function which is essentially a power law. And as in the formula size models, the vast majority of trees representing a given function is obtained by expanding a tree of minimal size exactly once. Expanding means attaching a tree of particular type to the given tree. However, there is also a clear difference: Whereas the limiting probabilities occurring in the formula size models differ among themselves by constant factor, the tree size model gives a strong bias to the constant functions True and False. This immediately implies that the tree size model does not exhibit the Shannon effect. Though the formula size models do not exhibit the Shannon effect as well, this is much less obvious (see [12], [11], [15]).

The present paper is organized as follows: In the next section we introduce the model and present our main result. Section 3 introduces the combinatorial setting (generating functions for the basic tree classes) and briefly discusses the existence of the limiting distribution to be studied in the subsequent sections. Some particular properties of our model are studied in Section 3. These properties were surprising since they did not show up in all the formula size models we studied so far. In Section 5 we study a particular subfamily of trees which will be used as an auxiliary structure in Sections 6 and 9. The structure of the set of tautologies is the topic of Section 6. 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. Finally, Section 9 describes the expansion procedure. Moreover, we show in this section that the set of minimal trees of a Boolean function forms indeed a kind of basis of the set of all trees (up to asymptotically negligible tree sets). This is done by carefully estimating the contribution of iterated expansions as well as that of expanded non-minimal but irreducible trees.

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 \( \land \) in the following) or by a connector OR (denoted by \( \lor \) in the following) such that two identical connectives cannot be neighbours (trees are stratified), and where each leaf is labelled by a literal taken in \( \{x_1, \overline{x_1}, \ldots, x_n, \overline{x_n}\} \) (see Figure 1 for an example).

Every such tree represents a Boolean expression and therefore computes a Boolean function of \( n \) variables. We denote by \( \mathcal{F}_n \) the set of such Boolean functions.

**Definition 2.2.** The size \(|t|\) of an associative tree \( t \) is the number of its nodes (internal nodes and leaves). We denote by \( \mathcal{A}_{m,n} \) the set of associative trees of size \( m \), and by \( \mathcal{A}_{m,n} \) the cardinality of this set. Let \( f \in \mathcal{F}_n \) be a Boolean function, its complexity is

- 0 if \( f \) is the function \( \text{True} : (x_1, \ldots, x_n) \mapsto 1 \) or \( \text{False} : (x_1, \ldots, x_n) \mapsto 0 \);
Figure 1: An associative tree which represents the constant function True.

- 2 if \( f \) is a literal function, i.e. there exist \( x_i \in \{x_1, \ldots, x_n\} \) such that \( f : (x_1, \ldots, x_n) \mapsto x_i \) or \( f : (x_1, \ldots, x_n) \mapsto \bar{x}_i \);
- the 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 \( M_f \).

**Remark:** Note that our definition is the tree size complexity. We use the graph theoretical size as opposed the formula size complexity predominantly used in computer science. The formula size computer is the number of leaves of the tree.

**Example:** As an example, the function \( x_1 \oplus x_2 = (\bar{x}_1 \land x_2) \lor (x_1 \land \bar{x}_2) \) has complexity 7.

**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 a \( \land \) with a single child labelled by \( x_i \) and the tree rooted by a \( \lor \) with a single child labelled by \( x_i \), even if those two trees are not associative trees according to Definition 2.1!

**Definition 2.3.** Let \( f \in \mathcal{F}_n \). We denote by \( P_{m,n}(f) \) the probability of \( f \), being the proportion of trees of size \( m \) computing \( f \) among all trees of size \( m \).

The aim of the present paper is to prove the following result, which sums up the behaviour of this distribution \( P_{m,n} \) asymptotically when the size \( m \) of the considered tree tends to infinity.

**Theorem 2.4.** The asymptotic probability \( P_n(f) = \lim_{m \to \infty} P_{m,n}(f) \) exists for all Boolean function \( f \), and it satisfies:

- The probability of the constant functions True and False is such that:
  \[
  0 < \alpha \leq P_n(\text{True}) = P_n(\text{False}) \leq \beta < \frac{1}{2},
  \]
  for all \( n \geq 0 \), where \( \alpha \) and \( \beta \) are two constants, independent of \( n \).
- The probability of a literal function \( f \) satisfies
  \[
  P_n(f) = \Theta \left( \frac{1}{n^2} \right).
  \]
- For \( f \), a fixed Boolean function in \( \mathcal{F}_n \), such that \( L(f) \geq 3 \),
  \[
  P_n(f) = \Theta \left( \frac{1}{n^{L(f)}} \right).
  \]
Remark: Thanks to the ad hoc definition of the complexity (cf. Definition 2.2), the above theorem can be written: for all Boolean function $f$,

$$\mathbb{P}_n(f) = \Theta\left(\frac{1}{n^{L(f)}}\right),$$

asymptotically when $n$ tends to infinity.

Once we have seen such results we immediately note that they are really analogous to the one obtained for the other complexity measure. However, as it will be shown in Sections 4 and 5, we need to develop entirely new strategies in order to prove them.

3. Existence of the asymptotic probability $\mathbb{P}_n$

In this section, we prove the existence of the asymptotic distribution $\mathbb{P}_n$ on the set $\mathcal{F}_n$, which correspond to the first assertion of Theorem 2.4. In the whole paper, we deeply use generating functions, singularity analysis and the symbolic method. We refer the reader to [7] for a comprehensive introduction to this domain.

An associative tree can be formally described by a grammar: if $\mathcal{A}$ (resp. $\hat{\mathcal{A}}$) denotes the set of all associative trees rooted by an $\land$-connective (resp. an $\lor$-connective), $\mathcal{L}$ the set of literals and $\mathcal{C} = \{\land\}$, then $\hat{\mathcal{A}} = \mathcal{L} + \mathcal{C} \times \text{seq}_{\geq 2}(\hat{\mathcal{A}})$. Let us denote by $\hat{\mathcal{A}}(z)$ (resp. $\hat{A}(z)$) the generating function of associative trees rooted by an $\land$-connective (resp. an $\lor$-connective). The grammar can be translated into the functional equation

$$\hat{\mathcal{A}}(z) = 2nz + z \cdot \hat{\mathcal{A}}(z)^2 \cdot \frac{1}{1 - \hat{\mathcal{A}}(z)}.$$

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

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

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

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

Moreover, let $\hat{A}_f(z)$ (resp. $\hat{A}_f(z)$) denote the generating function of associative trees rooted by an $\land$ (resp. an $\lor$) connective and computing the Boolean function $f$. By symbolic arguments, we get that

$$\hat{A}_f(z) = z \cdot \mathbb{1}_{\text{flit}} + z \cdot \sum_{\ell \geq 2} \sum_{g_1 \land \cdots \land g_\ell = f} \hat{A}_{g_1}(z) \cdots \hat{A}_{g_\ell}(z),$$

$$\hat{A}_f(z) = z \cdot \mathbb{1}_{\text{flit}} + z \cdot \sum_{\ell \geq 2} \sum_{g_1 \lor \cdots \lor g_\ell = f} \hat{A}_{g_1}(z) \cdots \hat{A}_{g_\ell}(z).$$

We thus have a system of functional equations to which we can apply the Drmota-Lalley-Woods theorem and infer that all the generating functions $A_f(z)$ have the same dominant singularity $\rho$, it is a square-root singularity and $A(z) = \sum_{f \in \mathcal{F}_n} A_f(z)$ admits a singular expansion of the same type and at the same singularity $\rho$. Applying a transfer lemma (see [6]) to $A(z)$ and $A_f(z)$, we can conclude that the asymptotic distribution $\mathbb{P}_n(f) = \lim_{n \to \infty} \frac{\lfloor zm \rfloor A_f(z)}{\lfloor zm \rfloor A(z)}$ exists.

Moreover, thanks to [6], we can easily show the following proposition:

**Proposition 3.1.** The singularity $\rho$ satisfies, as $n \to \infty$,

$$\rho = \frac{1}{2} \cdot \frac{1}{n + \sqrt{2n}} = \frac{1}{2n} - \frac{1}{n\sqrt{2n}} + \mathcal{O}\left(\frac{1}{n^2}\right),$$

$$A(\rho) = 1 - \frac{1}{n} + \mathcal{O}\left(\frac{1}{n\sqrt{n}}\right),$$

$$\hat{A}(\rho) = 1 - \frac{1}{\sqrt{2n}} + \mathcal{O}\left(\frac{1}{n}\right).$$
In particular, \( \rho \) can be bounded by \( \frac{1}{\sqrt{n}} - \frac{1}{2n\sqrt{n}} \leq \rho < \frac{1}{\sqrt{n}} \).

If \( \hat{B}(z) \) denotes the generating function of associative trees with root label \( \bigwedge \), then we have
\[
\hat{B}(z) = \hat{A}(z) - 2nz
\]
and
\[
\hat{B}(\rho) < \frac{1}{\sqrt{2n}} \quad \text{as well as} \quad \hat{B}(\rho) = \frac{1}{\sqrt{2n}} + \mathcal{O}\left(\frac{1}{n}\right).
\]

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 “classical models” (formula size complexity equal to the number of leaves) where repetitions in the labels of the variables are not likely in the first few levels (cf. [10, 13, 14]).

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

**Definition 4.1.** Let \( T \) be a family of associative trees. Let \( T_m \) be the number of trees of size \( m \) in this family. The limiting ratio of \( T \), if it exists, is defined (and denoted) by
\[
\mu_n(T) = \lim_{m \to \infty} \frac{T_m}{A_{m,n}}.
\]

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

**Lemma 4.2.** Let \( T(z) \) be the generating function of a family \( T \) of associative trees. Assume that \( \rho \) (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 \( \sqrt{z} - \rho \) at \( \rho \). Then
\[
\mu_n(T) = \lim_{z \to \rho} \frac{T'(z)}{A'(z)}
\]
where \( z \) must move towards \( \rho \) in such a way that \( \text{arg}(z - \rho) \neq 0 \) and inside the domain of analyticity of \( T(z) \) and \( A(z) \).

**Proposition 4.3.** The limiting ratio of trees with no leaf on the first level is given by
\[
\mu_n(A^{(0)}) = \frac{1}{n\sqrt{2n}} + \mathcal{O}\left(\frac{1}{n^2}\right), \quad n \to \infty.
\]

**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
\[
\mu_n(A^{(0)}) = \lim_{m \to \infty} \frac{[z^m] A^{(0)}(z)}{[z^m] A(z)} = \lim_{z \to \rho} \frac{A^{(0)}(z)}{A'(z)} = \frac{2\rho - 2\hat{B}(\rho)}{1 - \hat{B}(\rho)} \lim_{z \to \rho} \frac{\hat{B}'(z)}{A'(z)} + 2\rho \frac{\hat{B}(\rho)^2}{(1 - \hat{B}(\rho))^2} \lim_{z \to \rho} \frac{\hat{B}'(z)}{A'(z)} + 2 \frac{\hat{B}(\rho)^2}{1 - \hat{B}(\rho)} \lim_{z \to \rho} \frac{1}{A'(z)}.
\]
Observe that the third term of the sum is equal to zero since \( A'(z) \) tends to infinity when \( z \) tends to \( \rho \). This observation will be used in the whole paper and in the following, such terms will be omitted without mentioning. Note also that \( \lim_{z \to \rho} \frac{\hat{B}(z)}{A'(z)} = 1/2 \). Finally, use the asymptotics given in Proposition 3.1 and get:
\[
\mu_n(A^{(0)}) = \frac{1}{n\sqrt{2n}} + \mathcal{O}\left(\frac{1}{n^2}\right). \quad \square
\]
Remark: Be careful that in the whole paper, the size \( m \) of the trees tends to infinity, and then, the number of variables for the labelling tends to infinity. The order of these two limits must be kept in mind.

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

\[
\mu_n(A_\Gamma) = \sqrt{\frac{2}{n}} + O\left(\frac{1}{n}\right), \text{ when } n \text{ tends to infinity.}
\]

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

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

because a tree in \( A_\Gamma \) has a root labelled by \( \land \) or \( \lor \) (which gives a factor \( 2z \)), a first sequence of subtrees which are not a leaf labelled by a literal from \( \Gamma \) (which gives the factor \( \frac{1}{1-(\hat{A}(z)-\gamma z)} \)), then a leaf labelled by literal from \( \Gamma \) (factor \( \gamma z \)) 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

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

Proposition 4.5. Let \( X_n \) be the number of leaves in the first level of an associative tree of size \( n \). Then \( \mathbb{E}(X_n) \sim 2\sqrt{2n} \), as \( n \to \infty \), and the random variable \( X_{n/2\sqrt{n}} \) converges to a \( \Gamma(2, 1/2) \) distribution.

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

\[
A(z, u) = 2z \frac{(\hat{B}(z) + 2nz u)^2}{1 - (\hat{B}(z) + 2nz u)}.
\]

Therefore,

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

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

\[
\lim_{m \to \infty} \frac{[z^m] \frac{\partial}{\partial u} A(z, u) \bigg|_{u=1}}{[z^m] A(z)} = \lim_{z \to \rho} \frac{\frac{d}{dz} A(z, u) \bigg|_{u=1}}{\frac{d}{dz} A(z)} = 2\sqrt{2n} + O(1),
\]

as \( n \to \infty \). A similar calculation yields

\[
\lim_{m \to \infty} \frac{[z^m] A(z, u)}{[z^m] A(z)} = \lim_{z \to \rho} \frac{\frac{d}{dz} A(z, u) \bigg|_{u=1}}{\frac{d}{dz} A(z)} \sim \frac{1}{(1 - (u - 1)\sqrt{2n})^2}.
\]

If we set \( u = e^{i/2\sqrt{n}} \) and let \( X_n \) denote the number of leaves in a random tree of size \( n \), then we obtain

\[
\mathbb{E}e^{itX_n/2\sqrt{n}} = \lim_{m \to \infty} \frac{[z^m] A(z, e^{it/2\sqrt{n}})}{[z^m] A(z)} = \frac{1}{(1 - \frac{it}{2})^2}
\]

which is the characteristic function of the \( \Gamma(2, 1/2) \) distribution. \( \square \)

---

\( ^4 \) The density of a \( \Gamma(2, 1/2) \) distribution is \( 4xe^{-2x} \mathbb{1}_{x \geq 0} \).
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 define thus a particular family of trees in this section and study its limiting ratio.

Definition 5.1. Let \( k, r, \ell, p \geq 0 \) be four integers, let \( \{ \gamma_1, \ldots, \gamma_p \} \) be a subset of literals (with no occurrence of both a variable and its negation). Let \( M_{k, \ell, r}^p \) be the family of \( \ell \)-rooted trees

- with exactly \( k \) different literals \( \alpha_1, \ldots, \alpha_k \) appearing as labels of leaves on the first level, such that both a variable and its negation cannot appear, and for all \( i = 1, \ldots, k \) and \( j = 1, \ldots, p \) we have \( \alpha_i \neq \gamma_j \),
- with the root having exactly \( \ell \) non-leaf subtrees,
- and with at least one non-leaf subtree chosen from the family \( J_{k,r}^p \):.

The family \( J_{k,r}^p \) contains all \( \wedge \)-rooted trees such that

- there are \( r \) leaves in the first level carrying pairwise different labels \( \beta_1, \ldots, \beta_r \) (with no occurrence of both a variable and its negation) which are different from \( \alpha_1, \ldots, \alpha_k, \gamma_1, \ldots, \gamma_p \) as well as their negations;
- all the other leaves on the first level have labels from \( \{ \alpha_1, \ldots, \alpha_k, \gamma_1, \ldots, \gamma_p, \beta_1, \ldots, \beta_r \} \) or their negations,

Lemma 5.2. If \( k = \Omega(n^{1/4}) \), \( \ell = O(n^{1/8}) \) and \( r \leq \ell \), then the limiting ratio of the family \( M_{k,\ell, r}^p \) satisfies \( \mu_n(M_{k,\ell, r}^p) = O \left( \frac{1}{n^{\gamma_2}} \right) \) when \( n \) tends to infinity.

Proof. The generating function of \( J_{k,r}^p \) is given by

\[
J(z) = \frac{z^{r+1}}{(1 - (\hat{B}(z) + 2(k+p)z)) \cdots (1 - (\hat{B}(z) + 2(k+p+r)z))}.
\]

The generating function \( M(z) \) of \( M_{k,\ell, r}^p \) is satisfies, for all \( m \geq 0 \),

\[
[z^m]M(z) \leq [z^m]\frac{\left( \frac{n-p}{k+r} \right) \left( \frac{k+r}{k} \right) 2^{k+r} k! \left( \frac{1}{\ell!} \right) (\ell J(z) \hat{B}(z)^{\ell-1})}{1 - \nu z}
\]

where the right-hand side is composed as follows: \( z \) stands for the root, \( \left( \frac{n-p}{k+r} \right) \left( \frac{k+r}{k} \right) 2^{k+r} k! \) correspond to the choice of \( \alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_r \) such that a variable and its negation cannot be both chosen; \( \hat{B}(z) \) is \( (z^{k+\ell} \prod_{\nu=1}^{k+\ell} 1 - \nu z) \) correspond to the choice of the places of the \( \ell \) non-leaf subtrees of the root between the leaves of the first level; \( \ell J(z) \) stands for the non-leaf subtree taken in \( J_{k,r}^p \) and its place among the non-leaf subtrees, and \( \hat{B}(z)^{\ell-1} \) for the \( \ell - 1 \) other non-leaf subtrees of the root. We remark that we multi-count the cases when several non-leaf subtrees of the root are in \( J_{k,r}^p \). Thus, we only have an inequality in (2). It implies that, for all \( m \geq 0 \),

\[
[z^m]M(z) \leq \left( \frac{n-p}{k+r} \right) \left( \frac{k+r}{k} \right) 2^{k+r} k! \left( \frac{1}{\ell!} \right) (\ell J(z) \hat{B}(z)^{\ell-1}) \times \frac{\hat{B}(z)^{\ell-1}}{\prod_{\nu=0}^{\ell-1} \hat{B}(z) + (k+p+\nu)z}.
\]

Let \( K(z) = \frac{\hat{B}(z)^{\ell-1}}{\prod_{\nu=0}^{\ell-1} \hat{B}(z) + (k+p+\nu)z} \). The limiting ratio of the family \( M_{k,\ell, r}^p \) is obtained by the asymptotic leading term of

\[
\frac{M'(z)}{A'(z)} = \left( \frac{n-p}{k+r} \right) \left( \frac{k+r}{k} \right) 2^{k+r} k! \left( \frac{1}{\ell!} \right) \left( \frac{\hat{B}(z)^{\ell-1}}{\prod_{\nu=0}^{\ell-1} \hat{B}(z) + (k+p+\nu)z} \right) \rho^{\ell+2} K'(z) (1 + o(1))
\]

\[\text{Remember that } [z^m]\frac{(z^{f(z)})^\ell}{\ell!} = \left( \frac{m+\ell}{\ell} \right) f_m.\]
when \( z \) tends to \( \rho \).

Let \( f(z) = z^{k+\ell} \prod_{\nu=1}^{k} \frac{1}{(1 - \nu z)} \) and consider a circle of radius \( n^{-3/2} \) centered at \( \rho \). Since \( f(z) \) is a power series with non-negative coefficients, Cauchy’s estimate gives

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

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 \( \rho < \frac{1}{2n} \), we obtain

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

Thanks to (3), we obtain for \( z \to \rho \):

\[
\frac{M'(z)}{A'(z)} \leq \rho^{r+2} \binom{n - p}{k + r} \binom{k + r}{k} 2^{k+r+2} \ell! f^{(\ell)}(\rho) K'(z) A'(z) (1 + o(1))
\]

\[
\leq K'(z) A'(z) \cdot \ell \left( \frac{1}{2n} \right)^{k+\ell+r+2} 2^{k+r} (n)_k n^{\ell} \left( 1 + \frac{2}{\sqrt{n}} \right)^{k+\ell} n^{3\ell/2} \prod_{\nu=1}^{k} \frac{1}{1 - \frac{\nu}{2n} \left( 1 + \frac{2}{\sqrt{n}} \right)} (1 + o(1))
\]

\[
\leq K'(z) A'(z) \cdot e^{2(\ell - 2) - \ell} \cdot n^{\ell/2 - 2} \left( 1 + \frac{2}{\sqrt{n}} \right)^{k+\ell} \prod_{\nu=1}^{k} \frac{1}{1 - \frac{\nu}{2n} \left( 1 + \frac{2}{\sqrt{n}} \right)} (1 + o(1)).
\]

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

\[
\frac{M'(z)}{A'(z)} = \mathcal{O} \left( e^{2(\ell - 2) - \ell} K'(z) A'(z) \right) \tag{4}
\]

If \( k > 4\sqrt{n} \) then 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

\[
\left( 1 + \frac{2}{\sqrt{n}} \right)^{k} \prod_{\nu=1}^{k} \frac{1}{1 - \frac{\nu}{n} \left( 1 + \frac{2}{\sqrt{n}} \right)} = \left( 1 + \frac{2}{\sqrt{n}} \right)^{4\sqrt{n}} \prod_{\nu=1}^{k} \frac{1}{1 - \frac{\nu}{2n} \left( 1 + \frac{2}{\sqrt{n}} \right)}
\]

\[
\times \left( 1 + \frac{2}{\sqrt{n}} \right)^{k-4\sqrt{n}} \prod_{\nu=4\sqrt{n}+1}^{k} \frac{1 - \frac{\nu}{n}}{1 - \frac{\nu}{2n} \left( 1 + \frac{2}{\sqrt{n}} \right)}
\]

\[
\leq e^{8} \left[ \left( 1 + \frac{2}{\sqrt{n}} \right) \left( 1 - \frac{2}{\sqrt{n}} \left( 1 - \frac{2}{\sqrt{n}} \right) \right) \right]^{k-4\sqrt{n}}
\]

\[
= e^{8} \left( 1 + \frac{8}{n^{3/2}} \right)^{-4\sqrt{n}} = \mathcal{O}(1).
\]

Moreover,

\[
\frac{K'(z)}{A'(z)} \sim \frac{B'(z)}{A'(z)} \prod_{m=0}^{r} \frac{1}{1 - B(p) - (k + p + m)\rho} \left[ (\ell - 1) B(p)^{\ell - 2} + \sum_{m=0}^{r} B(p)^{\ell - 1} \right],
\]

8
and since $\frac{\hat{B}(z)}{A(z)} \sim 1/2$ when $z$ tends to $\rho$, and $\hat{B}(\rho) \leq \frac{1}{\sqrt{2n}}$ (cf. Proposition 3.1), we get

$$\frac{K'(z)}{A'(z)} = O\left(r^{2r} \rho \left(1 + \frac{1}{8n}\right)^r\right).$$

Finally, if we let $n$ tend to infinity, we get

$$\mu_n(M_k; B, \rho) = O\left(r^{2r} 2^{r-\frac{4r}{2n}} n^{-3/4}\right) = O\left(\frac{1}{n^{3/4}}\right).$$

\[\square\]

**Lemma 5.3.** The limiting ratio of trees with less than $n^{1/4}$ different labels appearing on the first level leaves is $O\left(\frac{1}{\sqrt{n}}\right)$.

**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

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

and therefore, their limiting ratio is given by

$$\lim_{\rho \to \rho} \frac{G_k'(z)}{A'(z)} = \binom{n}{k} 2^k k! \prod_{m=0}^{k} \frac{1}{1-mz-B(z)} \sum_{m=0}^{k} \frac{1}{1-mz-B(z)} \cdot \lim_{\rho \to \rho} \frac{B'(z)}{A'(z)}$$

$$\leq \frac{1}{4n} \prod_{m=0}^{k-1} \left(1 - m/n\right) \prod_{m=0}^{k} \frac{1}{1-m/2n - 1/\sqrt{2n}} \left(\sum_{m=0}^{k} \frac{1}{1-m/2n - 1/\sqrt{2n}}\right)$$

$$\leq \frac{1}{4n} \left(1 - \frac{k}{2n} - \frac{1}{\sqrt{2n}}\right)^{k+1} (k+1) \frac{1}{1-m/2n - 1/n} = O\left(\frac{k}{n}\right).$$

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

**Lemma 5.4.** The limiting ratio of the family of all trees where the root has more than $n^{1/s}$ non-leaf subtrees is $\Theta\left(\frac{n^{1/s}}{2^{n^{1/s}}}\right)$.

**Proof.** The generating function of the trees with exactly $\ell$ non-leaf subtrees is

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

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

$$\lim_{\rho \to \rho} \frac{H_\ell'(z)}{A'(z)} = \frac{1}{n} \frac{\ell \hat{B}^{\ell-1}(\rho)}{(1-2n\rho)^{\ell+1}} \cdot \lim_{\rho \to \rho} \frac{\hat{B}'(z)}{A'(z)} \sim \frac{1}{2n} \frac{\ell}{\left(\frac{1}{\sqrt{2n}}\right)^{\ell+1}} \frac{\hat{B}'(z)}{A'(z)}$$

$$\sim \frac{1}{2n} \frac{\ell}{\left(\frac{1}{\sqrt{2n}}\right)^{\ell+1}} \sim \frac{\ell}{2^{\ell+1}}, \text{ as } n \to \infty.$$ 

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

$$\sum_{\ell \geq n^{1/s}} \frac{\ell}{2^{\ell+1}} = \Theta\left(\frac{n^{1/s}}{2^{n^{1/s}}}\right).$$ 

\[\square\]
6. Tautologies.

This section deals with tautologic trees and states the first statement of Theorem 2.4 about the probability of constant functions. This is not only a particular case, but also the first step to prove the whole theorem. The first subsection is devoted to prove the lower bound, the second one to prove the upper bound and the last one to prove that almost all tautologies have a very simple shape, which is similar to the behaviour of the classical associative model.

6.1. A non-negligible family of tautologies.

In this section, we define a set of “simple” tautologies and find a lower bound for the limiting ratio of this family of tautologies.

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

Let \( \mathcal{E}_k \) be the set of trees rooted by \( \lor \), with exactly \( k \) leaves at depth 1 and at most 5 non-leaf subtrees. Let us take \( k \) in \( M := \{ \lfloor \sqrt{n} \rfloor, \ldots, 15 \lfloor \sqrt{n} \rfloor \} \). For two generating functions \( f, g \in \mathbb{R}[z] \), we write \( f < g \) (resp. \( f > g \)) if \( [z^k] f \leq [z^k] g \) (resp. \( [z^k] f \geq [z^k] g \)) for all \( r \in \mathbb{N} \). The generating function of \( \mathcal{E} = \bigcup_{k \in M} \mathcal{E}_k \) satisfies

\[
E(z) \sim \sum_{j=1}^{5} \sum_{k=\sqrt{n}}^{15 \sqrt{n}} z \cdot (2n z)^k \cdot \frac{(k+1)^j}{j!} B^j(z).
\]

The latter generating function is not exactly \( E(z) \) because of the approximation of \((k+1) \cdots (k+j) \sim (k+1)^j\). After some computations, we get: \( \mu_n(\mathcal{E}) \geq 0.36618 + o(1) \) for \( n \) tending to infinity.

Let us now introduce two subfamilies of \( \mathcal{E}_k \). The first family \( \mathcal{E}_1^k \) contains all trees of \( \mathcal{E}_k \) such that

(a) the set \( \mathcal{L} \) of all labels appearing among the \( \lfloor \sqrt{n} \rfloor \) leftmost leaves at depth 1 is of cardinality greater than \( \sqrt{n}/2 \),

(b) the \( k - \lfloor \sqrt{n} \rfloor \) rightmost leaves of depth 1 are labelled by literals whose negation does not belong to \( \mathcal{L} \).

The family \( \mathcal{E}_2^k \) contains all trees of \( \mathcal{E}_k \) such that

(a’) the set \( \mathcal{L}' \) of all labels appearing among the \( \lfloor \sqrt{n} \rfloor \) leftmost leaves at depth 1 is of cardinality smaller than or equal to \( \sqrt{n}/2 \),

(b) the \( k - \lfloor \sqrt{n} \rfloor \) rightmost leaves of depth 1 are labelled by literals whose negation does not belong to \( \mathcal{L}' \).

The following lemma comes from the fact that a tautology which is not simple fulfills either condition (a) or (a’) and always satisfies condition (b) since a variable and its negation cannot appear in the first level of the tree. Therefore, a non-simple tautology is either in \( \mathcal{E}_1^k \) or in \( \mathcal{E}_2^k \). The same applies to trees which do not even represent tautologies. Thus we obtain the following result:

Lemma 6.2. We have \( \mathcal{E} \cap \overline{S}_n \subseteq \bigcup_{k \in M} (\mathcal{E}_1^k \cup \mathcal{E}_2^k) \).

Let us now estimate the sizes of the sets \( \mathcal{E}_1^k \) and \( \mathcal{E}_2^k \). We get an upper bound by multiple counting: Let \( E_1^k(z) \) (resp. \( E_2^k(z) \)) be the generating functions of \( \mathcal{E}_1^k \) (resp. \( \mathcal{E}_2^k \)). Then

\[
E_1^k(z) \sim \sum_{j=1}^{5} z \cdot (2n z)^k \cdot \left( \frac{2n - \sqrt{n}}{2} \right)^k \cdot \frac{(k+1)^j}{j!} B^j(z).
\]

Some multiple counting is done for the leftmost \( \lfloor \sqrt{n} \rfloor \) leaves at depth 1, because we did not pay attention to the number of used labels. Further over-counting appears by forbidding only \( \lfloor \sqrt{n}/2 \rfloor \) labels for the rightmost \( k - \lfloor \sqrt{n} \rfloor \) leaves whereas \( \mathcal{L} \) has at least this cardinality. Given this functional equation we get \( 0.24457 \) as an asymptotic upperbound for the limiting ratio of the family \( \bigcup_{k \in M} \mathcal{E}_1^k \).
Furthermore,
\[ E^k_2(z) < \sum_{j=1}^{5} \binom{n}{\lfloor \sqrt{n}/2 \rfloor} 2^{\sqrt{n}/2} \cdot z \cdot \left( \frac{\sqrt{n}}{2} \right)^{\sqrt{n}} \cdot (2nz)^{k-\sqrt{n}} \cdot \frac{(k+5)^j \hat{B}^j(z)}{j!}. \]

Again we do some multiple counting in the same fashion as before. This functional equation permits us to conclude that the limiting ratio of the family \( \bigcup_{k \in \mathbb{M}} \mathcal{E}^k_1 \) is, for \( n \) tending to infinity, of order \( o(1) \).

**Proposition 6.3.** The limiting ratio of simple tautologies is greater than \( \alpha := 0.12161 \).

**Proof.** The previous lemma is equivalent to \( \mathcal{E} \cap \mathcal{S} \supseteq \mathcal{E} \setminus \bigcup_{k \in \mathbb{M}} (\mathcal{E}^k_1 \cup \mathcal{E}^k_2) \). Thus, for \( n \to \infty \) we have
\[
\mu_n(\mathcal{S}_n) \geq \mu_n(\mathcal{S}_n \cap \mathcal{E}) \geq \mu_n(\mathcal{E}) - \left( \sum_{k=\sqrt{n}}^{15\sqrt{n}} \mu_n(\mathcal{E}^k_1) + \mu_n(\mathcal{E}^k_2) \right) \geq 0.36618 - 0.24457 + o(1). \]

6.2. A non-negligible family of non-tautologies.

In this section, we derive an upper bound for the probability of tautologies.

**Definition 6.4.** Let \( \mathcal{G}_k \) be the family of \( \lor \)-rooted trees with exactly \( k \) leaves on the first level, labelled by \( k \) different literals \( \alpha_1, \ldots, \alpha_k \) such that each variable can only appear positive or negative and whose non-leaf subtrees are all contradictions.

Note that an \( \lor \)-rooted tree in \( \mathcal{G} \) computes the function \( \alpha_1 \lor \ldots \lor \alpha_k \), and is therefore neither a tautology nor a contradiction.

**Lemma 6.5.** There exists a constant \( \beta \) such that, for all \( n \geq 0 \),
\[ \mathbb{P}_n(\text{True}) \leq \beta < \frac{1}{2}. \]

**Proof.** The generating function of the tree family defined in Definition 6.4 is given by
\[ \hat{G}_k(z) = \binom{n}{k} 2^{k!} n^{k+1} \frac{1}{(1 - T(z))^{k+1}}, \]
and its limiting ratio is given by
\[ \frac{\hat{G}'(z)}{A'(z)} \sim \binom{n}{k} 2^{k!} n^{k+1} \frac{k + 1}{(1 - T(z))^{k+1}} \frac{T'(z)}{A'(z)}. \]

But Proposition 6.3 tells us that \( \frac{T'(z)}{A'(z)} \geq \frac{\alpha}{2} \) for \( z \) sufficiently close to \( \rho \). Moreover, we know that \( T(\rho) > 0 \) and therefore \( \frac{1}{(1 - T(\rho))^{k+1}} > 1 \). Therefore,
\[
\lim_{z \to \rho} \frac{G'(z)}{A'(z)} \geq \binom{n}{k} 2^{k!} (k+1) \left( \frac{1}{2n} \right)^{k+1} \frac{\alpha}{2} = \frac{n(n-1)\ldots(n-k+1)k+1}{n^k} \frac{k+1}{4n} \geq \alpha \cdot \left( 1 - \frac{k-1}{n} \right)^{k-1} k+1 \frac{1}{4n}. \]

If \( k = \Theta(\sqrt{n}) \), then the right-hand side of the above inequality is of order \( \Theta(\frac{1}{\sqrt{n}}) \), and the limiting ratio of \( \mathcal{G}_k \) is thus larger than \( \frac{c_k}{\sqrt{n}} \) where \( c_k \) is a positive constant. Thus the limiting ratio of the family \( \bigcup_{k=\lfloor \sqrt{n} \rfloor} \mathcal{G}_k \) is bounded from below by a positive constant \( c \). We therefore have proved Lemma 6.5 and \( \beta = \frac{1}{2} - c \). \( \square \)
6.3. Almost every tautology is simple.

In general, it is not easy to describe precisely the shape of a tautologic tree, but according to our distribution on trees, almost every tautology is “simple”.

**Theorem 6.6.** Asymptotically almost every tautology is a simple tautology, i.e.

\[ P_n(\text{True}) = \mu_n(\mathcal{S}_n) + o(1), \text{ as } n \to \infty. \]

**Proof.** To prove Theorem 6.6 we will 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.

Let us consider \( \mathcal{N}_n = \mathcal{T}_n \setminus \mathcal{S}_n \) the set of tautologies which are not simple. Let \( t \in \mathcal{N}_n \), with \( \ell \) 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 \( \alpha = \{\alpha_1, \ldots, \alpha_k\} \) 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 \( \alpha \) and others, called “new leaves” – with new labels. Let us show that there exists at least one \( i \in \{1, \ldots, \ell\} \) such that \( A_i \) has at most \( \ell - 1 \) new leaves.

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

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

After the \( \ell \)th step, we have found an assignment of variables different from the variables involved in \( \alpha \), such that all trees of \( \{A_1, \ldots, A_\ell\} \) compute \( \text{False} \). Thus \( t \) computes \( \alpha_1 \lor \ldots \lor \alpha_k \), which is impossible since \( t \) is a tautology.

Thus, there exist at least one \( i \in \{1, \ldots, \ell\} \) such that \( A_i \) has at most \( \ell - 1 \) new leaves in the first level.

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

\[
\bigcup_{k=0}^{n} \bigcup_{\ell=0}^{\infty} \bigcup_{r=0}^{\ell-1} \mathcal{M}_{k,\ell,r}^0 = \left( \bigcup_{k=0}^{n^{1/4}} \bigcup_{\ell=0}^{\infty} \bigcup_{r=0}^{\ell-1} \mathcal{M}_{k,\ell,r}^0 \right) \cup \left( \bigcup_{k=n^{1/4}}^{n} \bigcup_{\ell=0}^{\infty} \bigcup_{r=0}^{\ell-1} \mathcal{M}_{k,\ell,r}^0 \right) \cup \left( \bigcup_{k=n^{1/4}}^{n} \bigcup_{\ell=n^{1/4}}^{\infty} \bigcup_{r=0}^{\ell-1} \mathcal{M}_{k,\ell,r}^0 \right) .
\]

Thanks to Lemma 5.3, the first term has a limiting ratio tending to zero as \( n \) tends to infinity; Lemma 5.4 guarantees that the second term has also a limiting ratio tending to zero, and by Lemma 5.2, the third term satisfies

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

and is thus also tending to zero when \( n \) tends to infinity. Thus, \( \mu_n(\mathcal{N}_n \setminus \mathcal{S}_n) = o(1) \) and Theorem 6.6 is proved. \( \square \)

7. Probability of functions larger than a fixed \( f_0 \)

In this section we investigate the probability of all the functions that are “larger” than a fixed given function. The result is quite surprising and enables us to prove the second statement of Theorem 2.4 about literal functions, proved in Section 5.

**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 \( (x_1, \ldots, x_n) \in \{0, 1\}^n \).

In the following we fix a Boolean function \( f_0 \) and estimate \( P_n(f \geq f_0) \). We first consider the case \( f_0 = x_1 \land \ldots \land x_p \) and then generalize the result to any Boolean function.
Proposition 7.2. For all Boolean functions \( f_0 \in \mathcal{F}_n \),
\[
\mathbb{P}_n(f \geq f_0) \sim \mathbb{P}_n(\text{True})
\]
when \( n \) tends to infinity.

7.1. The case \( f_0 \) is a conjunction of literals

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

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

Consider first the case where \( t \) is rooted by an \( \lor \):

- Let us first assume that there exists one leaf on the first level of \( t \), labelled by one of the \( \gamma_i \). The family of trees with at least one leaf on the first level and with a label from the set \( \{\gamma_1, \ldots, \gamma_p\} \) 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 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 \( \gamma_1, \ldots, \gamma_p \). 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 \( \ell \) 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 literal 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 \( \ell \) different new variables appearing on its first level. Thus, since each \( A_i \) is rooted by an \( \land \), we can find an assignment of the variables \( \{x_1, \ldots, x_n\} \setminus \mathcal{O} \) such that \( A_i \) computes \( \text{False} \) for this assignment. Assign then all the leaves on the first level of \( t \) to \( \text{False} \) and \( \gamma_1, \ldots, \gamma_p \) to \( \text{True} \). Then \( t \) computes \( \text{False} \) while \( \gamma_1 \land \ldots \land \gamma_p \) attains the value \( \text{True} \) for this assignment. But this is impossible!

Thus, there exists at least one \( A_i \) which has less than \( \ell \) new variables on its first level. It means that \( t \) belongs to the set \( \bigcup_{r=0}^{k-1} \mathcal{M}^p_{k,\ell,r} \).

The limiting ratio of such trees is thus less than the limiting ratio of \( \bigcup_{k,\ell \geq 0} \bigcup_{r=0}^\ell \mathcal{M}^p_{k,\ell,r} \) thanks to the results proved Section [5].

\[
\mu_n \left( \bigcup_{k,\ell \geq 0} \bigcup_{r=0}^{\ell-1} \mathcal{M}^p_{k,\ell,r} \right) = \mathcal{O} \left( (n - n^{1/2}) \left( n^{1/4} \right)^2 \frac{1}{n^{1/2}} \right) = \mathcal{O} \left( \frac{1}{n^{1/2}} \right)
\]

and this family is also negligible in comparison to tautologies.

If \( t \) is rooted by an \( \land \), then its first level leaves have labels chosen from the set \( \{\gamma_1, \ldots, \gamma_p\} \). The family of trees with first level leaves labelled in \( \{\gamma_1, \ldots, \gamma_p\} \) has generating function
\[
H_p(z) = 2nz + 2z \frac{(\hat{B}(z) + pz)^2}{1 - \hat{B}(z) - pz}
\]
and its limiting ratio is asymptotically equal to \( \frac{1}{\sqrt{n}} \). It is thus negligible in front of tautologies. We have thus proved that
\[
\mathbb{P}_n(f \geq \gamma_1 \land \ldots \land \gamma_p) \sim \mathbb{P}_n(\text{True})
\]
as \( n \) tends to infinity.

7.2. Any Boolean function \( f_0 \)

Any Boolean function \( f_0 \) can be written as \( f_0 = (\gamma_1 \land \ldots \land \gamma_p) \lor g_0 \) for some integer \( p \geq 1 \), some literals \( \{\gamma_1, \ldots, \gamma_p\} \) and some Boolean function \( g_0 \). Thus,
\[
\mathbb{P}_n(\text{True}) \leq \mathbb{P}_n(f \geq f_0) \leq \mathbb{P}_n(f \geq \gamma_1 \land \ldots \land \gamma_p) \sim \mathbb{P}_n(\text{True})
\]
and Proposition 7.2 is proved.
8. Literals

The aim of this section is to compute the limiting ratio of Boolean function of the shape \(((x_1, \ldots, x_n) \mapsto x)\) where \(x\) is a literal among \(\{x_1, \bar{x}_1, \ldots, x_n, \bar{x}_n\}\) and to prove the second statement of Theorem \[8.4\] 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 \(x\) if it is rooted by an \(\lor\) (resp. \(\land\)), 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 and by \(X_m\) the number of simple \(x\) of size \(m\).

**Lemma 8.2.** For \(n\) tending to infinity, the limiting ratio of simple \(x\) satisfies
\[
\mu_n(\mathcal{X}) \sim \frac{P_n(\text{True})}{2n^2}.
\]

*Proof.* The generating function of simple \(x\) is given by the following generating function:
\[
X(z) = 2z^2T(z)
\]
along with
\[
\mu_n(\mathcal{X}) = \lim_{z \to \rho} \frac{X'(z)}{A'(z)} = 2\rho^2 \frac{T'(z)}{A'(z)}.
\]
Observing that \(\frac{T'(z)}{A'(z)}\) tends to \(P_n(\text{True})\) when \(z\) tends to \(\rho\) 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,
\[
P_n(x) \sim \mu_n(\mathcal{X}).
\]

*Proof.* Thanks to Lemma \[8.2\] and Proposition \[6.3\] we know that
\[
P_n(x) \geq \frac{P_n(\text{True})}{2n^2} \geq \frac{\alpha}{2n^2}
\]
when \(n\) tends to infinity. Let \(t\) be a tree computing \(x\). Let us assume that it is rooted by \(a \land\) (the case of a \(\lor\)-rooted tree would be treated in a 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 \(A_x^{(0)}\), its limiting ratio satisfies (cf. Proposition \[4.3\])
\[
2n\mu_n(A_x^{(0)}) \leq \mu_n(A^{(0)}) \sim \frac{1}{n\sqrt{2n}}.
\]
Thus, the limiting ratio of trees with no leaf on the first level, computing \(x\) has order \(O(n^{-5/2})\). This family is negligible in comparison to the family of simple \(x\) trees. Thus we can focus on trees with leaves 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 \[7.4\]). Thus, a tree computing \(x\) is almost surely an \(\land\)-rooted (resp. \(\lor\)-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 that
\[
\lim_{z \to \rho} \frac{L'_{x}(z)}{A'(z)} \sim \frac{P_n(\text{True})}{2n^2}, \text{ when } n \to \infty.
\]
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
\[
T_x(z) = 2z^2 \frac{z}{1-z} \frac{1}{1-L_x(z)} - 2z^2.
\]
The limiting ratio is then
\[
\mu_n(T_x) \sim \frac{2\rho}{1-\rho} \lim_{z \to \rho} \frac{L'_{x}(z)}{A'(z)} \sim \frac{P_n(\text{True})}{2n^2}, \text{ as } n \to \infty.
\]
Thus,
\[
P_n(x) \sim \mu_n(\mathcal{X}) \sim \frac{P_n(\text{True})}{2n^2}, \text{ as } n \to \infty.
\]
9. General case: minimal trees and expansions.

In this last section we prove the last statement of Theorem 2.4. We use different expansions of trees, as it was done in other random Boolean tree models (cf. [10] for implication random trees and [13] for And/Or trees). The first subsection defines the expansions, the second subsection states an asymptotic lower bound for $P_n(f)$, and the third subsection states an asymptotic upper bound and thus completes the proof of Theorem 2.4.

9.1. Expansions

**Definition 9.1.** Let $t$ be an associative tree. The tree given by adding a new subtree $t_e$ to an internal node $v$ 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 $t_e$ is a tautology (resp. a contradiction) and if $v$ is labelled by a $\land$ (resp. $\lor$). Obviously, such an expansion is valid.

- It is called a $\hat{B}$-expansion if the added tree $t_e$ 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 $E^*(\mathcal{T}) := \bigcup_{k \geq 1} E^k(\mathcal{T})$.

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 $t_e$ to $t$, then expanding $t_e$, 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 $E^*(\mathcal{T)) \subseteq E^*[\hat{B}(\mathcal{T})]$ holds.

9.2. Tautology expansions

**Proposition 9.2.** For $n$ tending to infinity, the limiting ratio of $E^*(\mathcal{M}_f)$ is asymptotically equal to the limiting ratio of $E(\mathcal{M}_f)$. Furthermore,

$$
\mu_n(E(\mathcal{M}_f)) = \Theta \left( \frac{1}{n^{L(f)}} \right), \text{ as } n \to \infty.
$$

Since every tree in $E^*(\mathcal{M}_f)$ computes $f$, this implies

$$
P_n(f) \geq \mu_n(E^*(\mathcal{M}_f)) = \Theta \left( \frac{1}{n^{L(f)}} \right).
$$

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

$$
P_t = \sum_{i \text{ internal node of } t} (d(i) + 1)
$$

where $d(i)$ is the number of children of the internal node $i$. Let $i_t$ denote the number of internal nodes of $t$ and $|t|$ the size of $t$. Then $P_t = i_t + |t| - 1$. 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

$$
m_f z^{L(f)} L(f) T(z) \prec \Phi_1(z) \prec m_f z^{L(f)} \frac{3L(f)}{2} T(z),
$$

and thus

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

(5)
From Section 6 we know that $0 < \alpha \leq \mathbb{P}_n(\text{True}) = \lim_{z \to \rho} \frac{T^e(z)}{A(z)} \leq \beta$ and since $\rho \sim \frac{1}{\sqrt{n}}$ when $n$ tends to infinity, we get that

$$\mu_n(E(\mathcal{M}_f)) = \Theta \left( \frac{1}{n^{k(f)}} \right).$$

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

$$\Phi_k(z) < m_f z^{L(f)} \left( \frac{\lfloor \frac{3L(f)}{2} \rfloor + k - 1}{k} \right) T(z)^k$$

and thus:

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

for all $n \geq 0$. Hence

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

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

$$\sum_{k \geq 0} \left( \frac{C+k-1}{k} \right) z^k = \frac{1}{(1-z)^C}. \quad (6)$$

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

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

By (5) the same calculations yield the lower bound. \[ \square \]

### 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 **irreducible**. 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

$$\mathbb{P}_n(f) \leq \mu_n(E^*(\mathcal{M}_f)) + \mu_n(E^*(\mathcal{I}_f)). \quad (7)$$

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

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

To prove this proposition, we have to better understand the shape of an irreducible tree of $f$. Let $t$ be such a tree. Let us “simplify” the tree as follows:

- assign all leaves of $t$ which are labelled by inessential variables\(^6\) of $f$ to $\text{True}$, and then

\(^6\)A variable $x$ is an inessential variable of $f$ if the function $f$ computed by $t$ does not depend on $x$.\[ \]
• simplify the tree as follows: as soon as a leaf is assigned to True (resp. False) and its parent is \& (resp. \lor), we cut the leaf. If its parent is \lor (resp. \&), we cut the subtree rooted at this \lor (resp. \&), i.e. at the parent.

The obtained tree, denoted by \( t^* \) 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 \lor-leaves compute False and \&-leaves compute True, the other leaves being labelled by essential variables of \( f \). But, since 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 \( t^* \) belongs to the following family of trees:

**Definition 9.5.** Let us denote by \( S \) the set of trees with internal nodes labelled by \& and \lor in a stratified way and with leaves labelled by essential variables of \( f \) or \& and \lor (again in compliance with the stratification), in which internal nodes can have one child or more. We denote by \( S_\ell \) the number of such trees of size \( \ell \) (the size being the total number of nodes of the tree).

**Remark:** The number \( S_\ell \) only depends on \( \gamma \), the number of essential variables of \( f \), and on \( \ell \).

Note that during the simplification process, we have cut either leaves or larger subtrees. We will prove in the following lemmas that the family of trees such that the simplification process cuts no large tree and the family of trees in which we have cut “many” single leaves are negligible in comparison 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 \( \Gamma \subseteq \{x_1, \ldots, x_n\} \) be of cardinality \( \gamma \) and set \( Y = \{x_1, \ldots, x_n\} \setminus \Gamma \). Moreover, let \( \mathcal{N}_\gamma \) be the family of trees such that no node labelled by \lor (resp. \&) has a leaf labelled by a positive (resp. negated) variable from \( Y \) as a child. Then \( \mu_n(\mathcal{N}_\gamma) = 0 \), when \( n \) is large enough.

Note that the family \( \mathcal{N}_\gamma \) contains the family of trees computing a function \( f \) having \( \Gamma \) as its set of essential variables and such that the simplification process only cuts leaves and no larger 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 \( \nu_n \) of the generating function of this family is of squareroot type and satisfies \( \nu_n \sim 1/n^{\gamma} \) and is thus strictly larger than \( \rho_n \sim 1/2n \) for large enough \( n \). This implies the assertion. \( \square \)

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

**Lemma 9.7.** Let \( \ell \) be an integer and \( \Gamma \subseteq \{x_1, \ldots, x_n\} \) be of cardinality \( \gamma \). Set \( Y = \{x_1, \ldots, x_n\} \setminus \Gamma \) and let us denote by \( \mathcal{N}_\ell \) the family of trees with at least \( \ell \) leaves labelled by variables from \( Y \) such that none of these leaves has an ancestor labelled by \lor (resp. \&) which has a child being a leaf labelled by a positive (resp. negated) variable from \( Y \). Then \( \mu_n(\mathcal{N}_\ell) = o(1/n^{\ell+1}) \).

Note that the family \( \mathcal{N}_\ell \) contains the family of trees computing a function \( f \) having \( \Gamma \) as its set of essential variables and in which the simplification process cuts at least \( \ell \) single leaves.

**Proof.** Let us consider the family of trees obtained as follows:

1. take a rooted tree \( t_0 \) having \( \ell \) leaves and no nodes of arity 1, with \( \ell \leq t \leq 2\ell - 2 \),
2. label the root by \& or \lor,
3. add to each internal node some subtrees which are not single leaves labelled by a variable from \( Y \) or its negation (respecting stratification),
4. replace each edge by a sequence of stratified internal nodes with subtrees attached to them which are not single leaves labelled by a positive (negated) variable from \( Y \) if their parent is \lor (resp. \&),
5. replace each leaf by a tree rooted by \& (resp. \lor according to the stratification) with at least one literal from \( Y \) (resp. from the negations of \( Y \)) on the first level and with no literal from the negations of \( Y \) (resp. from \( Y \)) in the first generation.

17
The obtained family contains $\mathcal{N}_\ell$ and its generating function is given by

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

where

- the index $r$ in the summation represents the number of edges of the tree $t_0$ chosen in the construction;
- the factor $C_\ell$ is the number of choices for this $t_0$ and the factor 2 for its root label;
- the function
  $$zX(z) = \frac{(n-\gamma)z^2}{1 - (A(z) - (n - \gamma + 1)z)(1 - (A(z) - (n - \gamma)z))} - (n - \gamma)z^2$$
  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
  $$zV(z) = \frac{z}{1 - z \left( \frac{1}{1 - (A(z) - (n - \gamma)z)} \right)^2} - 1$$
  is the generating function of the sequences of internal nodes that replace the edges of $t_0$; and
- the function
  $$W(z) = \frac{1}{1 - (A(z) - (n - \gamma)z)}$$
  is the generating function 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 $F$ is $\rho$ and that it is a squareroot singularity. Therefore, the limiting ratio of this family can be computed by Lemma 4.2, i.e. by $\lim_{z \to \rho} \frac{F'(z)}{A'(z)}$. To compute this limiting ratio, let us note that $\lim_{z \to \rho} \frac{1}{A'(z)} = 0$ such that many terms in $F'(z)/A'(z)$ can be neglected (cf. last paragraph in the proof of Proposition 4.3). We obtain:

$$\lim_{z \to \rho} \frac{F'(z)}{A'(z)} = \sum_{r=\ell}^{2\ell-2} C_\ell \rho^{\ell+r+1} \left( \ell X(\rho)^{\ell-1} V(\rho)^r W(\rho)^{2r+1-\ell} \lim_{z \to \rho} \frac{X'(z)}{A'(z)} \\
+ r X(\rho)^{\ell} V(\rho)^{r-1} W(\rho)^{2r+1-\ell} \lim_{z \to \rho} \frac{V'(z)}{A'(z)} \\
+ (2r + 1 - \ell) X(\rho)^{\ell} V(\rho)^{r} W(\rho)^{2r-\ell} \lim_{z \to \rho} \frac{W'(z)}{A'(z)} \right).$$

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

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

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

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

All these relations imply

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

where $\kappa$ is a positive constant. \qed
We are now ready to prove Proposition 9.4. The two previous lemmas allow us to consider only irreducible trees in which the simplification process cuts at least one large tree and less than $L(f)$ single leaves. Let us denote by $I_1$ the set of such irreducible trees which the simplified tree $t^*$ has size $L(f)$, and by $I_2$ the set of such irreducible trees of $f$ such that $t^*$ has size at least $L(f) + 1$.

Lemma 9.8. We have the following asymptotic result:

$$E^*(I_1) = o \left( \frac{1}{n^{L(f)}} \right).$$

Proof. Let $t \in I_1$. To obtain $t^*$, 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. Assume first that we have cut only one non-leaf subtree rooted at a node $\nu$ during the algorithm. Then, either this tree contains an essential variable on its first-level leaves, or it belongs to $\bigcup_{k, \ell \geq 0} \bigcup_{r=0}^{\ell-1} M^r_{k, \ell, r}$. Otherwise, we could find an assignment of inessential variables such that we can cut the father of $\nu$ without changing the function computed by the tree $t$. This new assignment of inessential variables leads to different simplification of the tree $t$ that will cut at least the single leaves that were cut before, plus the larger subtree and its father: we thus obtain a tree of size less that $L(f)$ that computes $f$, which is impossible.

Any tree of $I_1$ is obtained by expanding a tree $s$ of $S_{L(f)}$ 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 $\left( \bigcup_{k, \ell \geq 0} \bigcup_{r=0}^{\ell-1} M^r_{k, \ell, r} \right) \cup A_r$ at a node of $s$ and at most $L(f)$ inessential leaves at other nodes of $s$ (where $A_r$ is the set of trees containing at least one first-level leaf labelled by an essential variable, 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 $I_1$. 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 $I_1$, denoted by $I_1(z)$ thus satisfies:

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

where $k$ counts the number of successive expansions done into the irreducible tree, and where $M(z)$ is the generating function of $\left( \bigcup_{k, \ell \geq 0} \bigcup_{r=0}^{\ell-1} M^r_{k, \ell, r} \right) \cup A_\gamma$. 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”).

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

$$\mu_n \left( E^*(I_1) \right) \leq S_{L(f)} \rho^{L(f)} \hat{B}(\rho) \sum_{k \geq 0} \left( \frac{3L(f) + 1 + k}{L(f) + 1, k, 2L(f)} \left( \hat{B}(\rho)^k + \frac{k}{2} \hat{B}(\rho)^{k-1} \right) \right. + \left. \sum_{q \geq 2} \left( \frac{3L(f) + q + k}{L(f) + q, k, 2L(f)} \frac{q + k}{2} \hat{B}(\rho)^{q+k-2} \right) \right). \tag{8}$$
The generating function $S_{L(f)} \rho^{L(f)} \hat{B}(\rho)$ behaves as $\frac{1}{\sqrt{\ell}}$ when $\ell$ tends to infinity. Let us prove that the second term of the sum behaves as $O(1)$ when $n$ tends to infinity. Let us first focus on

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

$$
= \frac{(2L(f) + 1)!}{(L(f) + 1)!(2L(f))!} \frac{1}{(1 - \hat{B}(\rho))^{3L(f)+2}}
$$
in view of Eq. (6). Very similar calculations lead to

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

Moreover, using (6) again,

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

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

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

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

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

$$
\leq \frac{1 + 2L(f)^{L(f)+1}}{2(1 - \hat{B}(\rho))^{3L(f)+3}} \left( \frac{2L(f) + 1}{1 - \frac{\hat{B}(\rho)}{1 - \hat{B}(\rho)} \frac{2L(f)+2}{2L(f)+1}} + \frac{1}{1 - \frac{\hat{B}(\rho)}{1 - \hat{B}(\rho)} \frac{2L(f)+1}{2L(f)+1}} \right),
$$

since $\frac{\hat{B}(\rho)}{1 - \hat{B}(\rho)}$ is smaller than 1 for large enough $n$. Similar calculations can be done for the last term of the sum (8), and we eventually get:

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

\[\square\]

**Lemma 9.9.** We have the following asymptotic result:

$$
E^*(\mathcal{I}_2) = o \left( \frac{1}{n^{L(f)}} \right).
$$

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

$$
I_2(z) \leq \sum_{\ell \geq L(f)+1} S_\ell \sum_{k \geq 0} \sum_{q \geq 0} \left( \frac{2\ell + L(f) + q + k}{L(f) + q, k, 2\ell} \right) \hat{B}(\rho)^{k+q}(2(n - \gamma)z)^{L(f)}
$$

where $S_\ell$ is the number of trees and thus, calculations in the same vein as those done in the proof of Lemma 9.8

$$
\mu_n(E^*(\mathcal{I}_2)) \leq \sum_{\ell \geq L(f)+1} S_\ell \sum_{k \geq 0} \sum_{q \geq 0} \left( \frac{2\ell + L(f) + q + k}{L(f) + q, k, 2\ell} \right) \frac{k + q}{2} \frac{1}{\sqrt{n^{k+q-1}}}
$$

$$
= O \left( \frac{1}{n^{L(f)+1}} \right).
$$

\[\square\]
Lemmas 9.8 and 9.9 directly induce Proposition 9.4 and we are now able to complete the proof of Theorem 2.4.

**Proof of Theorem 2.4.** The probability of tautologies and of literals are respectively treated in Sections 6 and 8. We now have to prove that for all Boolean function \( f \), \( \mathbb{P}_n(f) = \Theta \left( \frac{1}{n^{L(f)}} \right) \). In view of Proposition 9.2 we have that \( \mathbb{P}_n(f) = \Omega \left( \frac{1}{n^{L(f)}} \right) \) and Eq. 7 together with Proposition 9.4 gives \( \mathbb{P}_n(f) = \mathcal{O} \left( \frac{1}{n^{L(f)}} \right) \) which completes the proof of Theorem 2.4.

10. Conclusion

We have studied the relation between tree size complexity and probability of Boolean function under the non-binary plane model. It turns out that the model follows in large parts the paradigms already discovered in various formula size models, but gives, however, a much stronger bias to the constant functions and exhibits some surprising features concerning some technical properties of the trees.

The model discussed here relies on plane trees and does therefore not cover the commutativity of \( \land \) and \( \lor \). However, non-plane trees can be treated in exactly the same way, albeit the technical level of the computations considerably increases. How this can be accomplished is demonstrated for two formula size models in [13] and for implicational trees in [14].

What remains to do: In this paper we were able to show that the probability of constant functions is bounded between two positive constants. We expect that these probabilities converge for \( n \) tending to infinity (such a result would imply convergence of the probabilities of any given Boolean function), but this requires a much deeper understanding of the structure of tautologies which could not be provided in this paper.

**References**

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

[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*, 27:490–519, 2005.

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

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

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

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

[7] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge U.P., Cambridge, 2009.

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

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

[10] 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):317–349, 2012.

[11] D. Gardy. Random Boolean expressions. In *Colloquium on Computational Logic and Applications*, volume AF, pages 1–36. DMTCS Proceedings, 2006.

[12] A. Genitrini and B. Gittenberger. No Shannon effect on probability distributions on Boolean functions induced by Boolean expressions. In *Proc. of Analysis of Algorithms*, pages 303–316, Wien, Austria, July 2010. DMTCS Proceedings AM.
[13] A. Genitrini, B. Gittenberger, V. Kraus, and C. Mailler. Associative and commutative tree representations for Boolean functions. Submitted. http://arxiv.org/abs/1305.0651.

[14] 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, 20 pages, (electronic), 2012.

[15] A. Genitrini, B. Gittenberger, and C. Mailler. No Shannon-effect induced by And/Or trees. DMTCS Proceedings, 2014. to appear.

[16] A. Gupta and S. Mahajan. Using amplification to compute majority with small majority gates. Computational Complexity, 6(1):46–63, 1997.

[17] Jakub Kozik. Subcritical pattern languages for and/or trees. In Fifth Colloquium on Mathematics and Computer Science, Discrete Math. Theor. Comput. Sci. Proc., AI, pages 437–448. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2008.

[18] S. P. Lalley. Finite range random walk on free groups and homogeneous trees. The Annals of Probability, 21, 1993.

[19] H. Lefmann and P. Savický. Some typical properties of large And/Or Boolean formulas. Random Structures and Algorithms, 10:337–351, 1997.

[20] O. B. Lupanov. Complexity of formula realization of functions of logical algebra. Problemy Kibernetiki, 3:61-80, 1960. English translation: Problems of Cybernetics, Pergamon Press, 3:782–811, 1962.

[21] O. B. Lupanov. On the realization of functions of logical algebra by formulae of finite classes (formulae of limited depth) in the basis $\land$, $\lor$, $\neg$. Problemy Kibernetiki, 6:5-14, 1961. English translation: Problems of Cybernetics, Pergamon Press, 6:1–14, 1965.

[22] 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):243–285, 1994.

[23] J. Riordan and C. E. Shannon. The number of two terminal series-parallel networks. Journal of Mathematics and Physics, 21:83–93, 1942.

[24] P. Savický. Random Boolean formulas representing any Boolean function with asymptotically equal probability. Discrete Mathematics, 83:95–103, 1990.

[25] C. E. Shannon. Communication theory of secrecy systems. Bell System Tech. J., 28:656–715, 1949.

[26] L. Valiant. Short monotone formulae for the majority function. Journal of Algorithms, 5:363–366, 1984.

[27] A. R. Woods. Coloring rules for finite trees, and probabilities of monadic second order sentences. Random Structures and Algorithms, 10(4):453–485, 1997.