The growing tree distribution on Boolean functions.*

Brigitte Chauvin† Danièle Gardy‡ Cécile Mailler§

Abstract

We define a probability distribution over the set of Boolean functions of k variables induced by the tree representation of Boolean expressions. The law we are interested in is inspired by the growth model of Binary Search Trees: we call it the *growing tree* law. We study it over different logical systems and compare the results we obtain to already known distributions induced by the tree representation: Catalan trees, Galton-Watson trees and balanced trees.

1 Introduction

A Boolean function of k variables is a function $f : \{0,1\}^k \longrightarrow \{0,1\}$ where 0 and 1 may be interpreted as the truth values $False$ and $True$. Our aim is to build a probability distribution over the set of Boolean functions and to study it.

The uniform distribution over the set of Boolean functions of k variables - denoted by $\mathcal{F}_k$ - has been studied by Shannon [Sha49]. If we define the complexity of a Boolean function as the minimal number of connectives needed to represent this function by a Boolean expression, then a Boolean function chosen uniformly at random has, asymptotically when k tends to infinity, an exponential complexity. As the maximal complexity is also of exponential order, roughly speaking, an average Boolean function is asymptotically of maximal complexity. This phenomenon is called the *Shannon effect*.

Lefmann and Savický [LS97] and later Chauvin *et al.* [CFGG04] studied random distributions induced by tree representation of Boolean functions. Indeed, complete binary trees - i.e. trees whose internal nodes have either zero or two sons - with nodes labelled with connectives - for example $\wedge$ and $\vee$ - and with leaves labelled with literals $x_1, \bar{x_1}, \ldots, x_k, \bar{x_k}$, represent Boolean functions, and a random distribution over the set of such trees induces a random distribution over $\mathcal{F}_k$. Both articles define the *Catalan trees* distribution. let the size of a binary tree be the number of its internal nodes[1], $\mathbb{U}_{n,k}$ the uniform distribution over labelled binary trees of size n over k variables, and $\mu_{n,k}$ the distribution induced by $\mathbb{U}_{n,k}$ on $\mathcal{F}_k$. The limit distribution μ_k of $\mu_{n,k}$ when n tends to infinity exists and is called the Catalan trees distribution.

Chauvin *et al.* study another distribution over binary trees induced by a critical Galton-Watson process, which are randomly labelled afterwards. It gives a distribution over $\mathcal{F}_k$ denoted by π_k. It has been shown [CFGG04, GG10] that every Boolean function is weighted, but Boolean functions with lower complexity are more weighted by both μ_k and π_k.

In the present paper we now consider another distribution induced by the tree representation: the *growing tree* distribution, inspired by the Binary Search Tree growing process. We then label the binary tree according to two different models: the $\wedge/\vee$ model - studied in [LS97, CFGG04] - and the implication model - used to compare intuitional and classical logics in [FGGZ07, KZ04].

In section 2, we define precisely the growing tree model and the two labelling models, before stating our main results in the following section: Theorems 1 and 2 (proved in section 4) give the convergence of the growing tree distribution to a distribution which support is included into the set of constant functions in both labelling models, and Theorem 3 (proved in section 6) deals with the proportion of *simple tautologies* among tautologies. Finally, some extensions of the two labelling models are presented respectively in sections 5 and 6.

2 Growing tree

First, let us define a new distribution over unlabelled trees of size n - the *growing tree* distribution - via a random process stopped at step n.

DEFINITION 1. *The growing process $(\mathcal{T}_i)_{i\in\mathbb{N}}$ is defined by:*

*This work was partially supported by the ANR project NT09-432755 Boole.

†Laboratoire de Mathématiques de Versailles and INRIA Rocquencourt project Algorithms. Université de Versailles St-Quentin en Yvelines, 45 avenue des Etats-Unis, 78035 Versailles, France. Email: brigitte.chauvin@math.uvsq.fr

‡Laboratoire PRISM. Université de Versailles St-Quentin en Yvelines, 45 avenue des Etats-Unis, 78035 Versailles, France. Email: daniele.gardy@prism.uvsq.fr.

§Laboratoire de Mathématiques de Versailles. Université de Versailles St-Quentin en Yvelines, 45 avenue des Etats-Unis, 78035 Versailles, France. Email: cecile.mailler@math.uvsq.fr.

[1] Let us note that a binary tree with n internal nodes has $n+1$ leaves.

- *$\mathcal{T}_0$ is reduced to its root.*
- *Given $\mathcal{T}_i$, we choose uniformly at random a leaf of the tree and make it grow by giving it two sons. The new tree is $\mathcal{T}_{i+1}$.*

The random variable $\mathcal{T}_n$ is called the growing tree of size n. To define a distribution over the set $\mathcal{E}_k$ of random Boolean expressions over k variables, we have to choose a rule to label randomly the nodes. We choose to study two rules: the $\wedge/\vee$ model, which is complete - i.e. each Boolean function can be expressed in this logical system, and the implication model, which is simpler but not complete, and nevertheless useful to compare intuitional and classical logics [FGGZ07, KZ04].

Definition 2. **The $\wedge/\vee$ model.** *Given a growing tree of size n, we label it according to the following rules:*

- *Each internal node is labelled by $\wedge$ or $\vee$ with probability $\frac{1}{2}$ and $\frac{1}{2}$.*
- *Each leaf is labelled by a literal chosen uniformly at random in the set $\{x_1, \bar{x_1}, \ldots, x_k, \bar{x_k}\}$.*
- *All the labellings are independent from each other.*

The implication model. *Given a growing tree of size n, we label it according to the following rules:*

- *Each internal node is labelled by $\rightarrow$.*
- *Each leaf is labelled independently from the others by a literal chosen uniformly at random in the set $\{x_1, \ldots, x_k\}$.*

Each model defines a distribution, respectively $\mathbb{P}^{\wedge\vee}_{n,k}$ and $\mathbb{P}^{\rightarrow}_{n,k}$ - denoted by $\mathbb{P}_{n,k}$ when there is no possible confusion - over $\mathcal{E}_k$, the set of Boolean expressions over k variables. Let us define a surjective mapping Φ from $\mathcal{E}_k$ to $\mathcal{F}_k$ as follows[2]:

$\Phi(\gamma) = f$ if and only if γ represents (or computes) f.

The image of $\mathbb{P}_{n,k}$ by Φ over $\mathcal{F}_k$ is a distribution over $\mathcal{F}_k$ denoted by $p_{n,k}$: $\forall f \in \mathcal{F}_k$, $p_{n,k}(f) = \mathbb{P}_{n,k}(\{\gamma \in \mathcal{E}_k \text{ such that } \Phi(\gamma) = f\})$.

Our aim is now to study the behaviour of $p_{n,k}$ when the size n of the random tree tends to infinity: does it tend to a limit distribution p_k? What are the properties of this distribution p_k if it exists? Is there any Shannon effect on p_k when k tends to infinity?

[2] Of course, this is not a one-to-one mapping since a same function can be represented by different expressions. For example, $\Phi(x_1 \wedge \bar{x_1}) = \Phi(x_2 \wedge \bar{x_2}) = False$.

3 Main results

Surprisingly, the growing tree model is a very simple model. Indeed, in both labelling models, the asymptotic distribution p_k exists and its support is included into the set of the constant functions. Moreover, the speed of the convergence is of order $O\left(\frac{1}{\ln n}\right)$. We prove the following theorems:

Theorem 1. (Growing tree - $\wedge/\vee$ model) *In the case of the $\wedge/\vee$ labelling model, we have, when n tends to infinty: $p_{n,k} \longrightarrow p_k = \frac{1}{2}\delta_{True} + \frac{1}{2}\delta_{False}$. Moreover, $\|p_{n,k} - p_k\|_\infty = O\left(\frac{1}{\ln n}\right)$ when n tends to inifnity.*

Theorem 1 can be extended to a more general labelling model, as shown in section 5.

Theorem 2. (Growing tree - implication model) *In the case of the implication labelling model, we have, when $n \longrightarrow +\infty$: $p_{n,k} \longrightarrow p_k = \delta_{True}$. Moreover, $\|p_{n,k} - p_k\|_\infty = O\left(\frac{1}{\ln n}\right)$ when $n \longrightarrow +\infty$.*

To sum up, the asymptotic distribution p_k does not depend on k and there is obviously no Shannon effect, as the average complexity of a function chosen at random according to p_k is the complexity of a constant: 1.

Remark 1. *The difference between the two theorems comes from the fact that the function $False$ cannot be represented by an expression built with the single connective $\rightarrow$ and with the positive literals $\{x_1, \ldots, x_k\}$. A function can be expressed in this model if and only if there exists $i \in [\![1, k]\!]$ and $g \in \mathcal{F}_k$ such as $f = x_i \vee g$.*

In the Catalan trees and Galton-Watson models, an important part of the study in the implication labelling model was to consider *simple tautologies*, which are "simple" Boolean expressions that compute $True$:

Definition 3. ([FGGZ07]) *In the implication labelling model, every Boolean expression can be written as: $A_1 \rightarrow (A_2 \rightarrow \ldots (A_p \rightarrow \alpha))$. The subtrees $A_1, \ldots, A_p$ are called the* premises *of the Boolean expression and α is called the* goal*. A simple tautology is a Boolean expression which has a premise reduced to a simple leaf, labelled by α.*

Let ST_k be the set of simple tautologies over k variables.

It has been shown [FGGZ07] that either in the Catalan trees or in the Galton-Watson model, roughly speaking, every tautology is a simple tautology, asymptotically when k tends to infinity. The following theorem states that in the growing tree model, we get a different behaviour:

Theorem 3. *We have: $\mathbb{P}_{n,k}(ST_k) \xrightarrow{n\rightarrow+\infty} 1 - \mathsf{e}^{-1/k} \sim \frac{1}{k}$ when $k \rightarrow +\infty$.*

Since $\mathbb{P}_{n,k}(\{tautologies\}) = p_{n,k}(True) \xrightarrow{n\to+\infty} 1$, simple tautologies are not the only ones charged by the growing tree law, asymptotically when k tends to infinity.

4 Proofs of Theorems 1 and 2

We present two different proofs for Theorems 1 and 2: one using an analytic combinatorics approach, and a probabilistic approach via Yule trees. The first one is the one already used to study the Catalan trees or the Galton-Watson model: it is a general approach but it does not give the convergence speed easily. The probabilistic approach is more powerful: it is shorter and gives the speed of the convergence quite easily, but it is specific to the growing tree model. In both approaches, the proofs of Theorems 1 and 2 are almost the same. Therefore, we only present the proof of Theorem 1, which is the most intricate one, since there are two connectives instead of one.

4.1 The analytic combinatorics approach. The idea of this proof is to use generating functions and analytic combinatorics methods, as presented, e.g., by Flajolet and Sedgewick [FS09]. Briefly, we consider a sequence (for example $(p_{n,k}(f))_{n\geq 0}$) as the sequence of the coefficients of a power series. Thus, the asymptotic behaviour of the power series near its dominant singularity can give some clues about the asymptotic behaviour of the initial sequence.

Thus, given a Boolean function f, let us introduce its generating function, defined as:

$$\phi_f(z) = \sum_{n=0}^{+\infty} p_{n,k}(f) z^n$$

where $p_n(f)$ is the probability that the random growing tree $\mathcal{T}_n$ of size n computes f. Now, $\mathcal{T}_n$ computes f if and only if

- $n = 0$, $f = \alpha$ is a literal, and the root of $\mathcal{T}_0$ (which is also its single node) is labelled by α; or
- $n \neq 0$, the left subtree of $\mathcal{T}_n$ computes g, the right subtree of $\mathcal{T}_n$ computes h, the root of $\mathcal{T}_n$ is labelled by $\diamond \in \{\wedge, \vee\}$ and $f = g \diamond h$.

Moreover the subtrees of $\mathcal{T}_n$ are also growing trees, and the probability that the left subtree has size i is $\frac{1}{n}$. Indeed, we can show it by induction: if $n = 2$, then the left subtree has size 1 with probability 1. Now, let us assume that the size of the left subtree of $\mathcal{T}_n$ - denoted by $\mathcal{L}_n$ - follows the uniform law over $\{1, \ldots, n-1\}$. Thus,

$$\begin{aligned}\mathbb{P}(|\mathcal{L}_{n+1}| = i) &= \frac{i-1}{n}\mathbb{P}(|\mathcal{L}_n| = i-1) \\ &\quad + \left(1 - \frac{i}{n}\right)\mathbb{P}(|\mathcal{L}_n| = i) \\ &= \frac{i-1}{n}\cdot\frac{1}{n-1} + \frac{n-i}{n}\cdot\frac{1}{n-1} = \frac{1}{n}.\end{aligned}$$

Thus, the size of the left subtree $\mathcal{L}_{n+1}$ of $\mathcal{T}_{n+1}$ follows the uniform law over $\{1, \ldots, n\}$. Therefore, by conditioning on the size of the left subtree, we obtain the following formula:

(4.1)
$$\begin{aligned}p_{n+1,k}(f) &= \frac{1}{2}\sum_{g\wedge h=f}\sum_{i=0}^{n}\frac{1}{n+1}p_{i,k}(g)p_{n-i,k}(h) \\ &\quad + \frac{1}{2}\sum_{g\vee h=f}\sum_{i=0}^{n}\frac{1}{n+1}p_{i,k}(g)p_{n-i,k}(h).\end{aligned}$$

By multiplying (4.1) by z^{n+1} and summing for $n \geq 0$, we get a relationship between the 2^{2^k} different generating functions:

(4.2)
$$\begin{aligned}2\phi_f(z) - 2p_{0,k}(f) &= \int \sum_{g\wedge h=f}\phi_g(z)\phi_h(z)dz \\ &\quad + \int \sum_{g\vee h=f}\phi_g(z)\phi_h(z)dz.\end{aligned}$$

Deriving formula (4.2), we finally obtain the

LEMMA 1. *$\forall f \in \mathcal{F}_k$, we have:*

$$2\phi'_f(z) = \sum_{g\wedge h=f}\phi_g(z)\phi_h(z) + \sum_{g\vee h=f}\phi_g(z)\phi_h(z).$$

Lemma 1 gives a differential system satisfied by the 2^{2^k} generating functions for the 2^{2^k} Boolean functions of $\mathcal{F}_k$. Studies of the Catalan trees and of the Galton-Watson model by this method both lead to very similar systems, except that they are both algebraic systems (cf. [CFGG04]). In those cases, the Drmota-Lalley-Woods theorem allowed to conclude easily since it applies for algebraic systems ([Drm97, Lal93, Woo97]). In our case, this theorem cannot apply due to the differential operator. Luckily, we can still obtain a solution of the system from Lemma 1.

First we observe obvious symmetries that simplify the system. Indeed, in the growing tree model, the two connectives $\wedge$ and $\vee$ have the same probability. E.g., the functions $x_1 \wedge x_2$ and $x_1 \vee x_2$ have the same probability to be computed by a growing tree: they thus have the same generating function. Moreover, all the literals have the same probability to appear as labels of each leaf: for

example, $x_1 \wedge x_2$ and $x_1 \wedge x_3$ have the same probability to be computed by a growing tree.

We can thus define equivalency classes of Boolean functions through the following operations :

- permutation of variables,
- negation of a variable,
- negation of the function.

One class is $\{False, True\}$: let us denote by ϕ_V the generating function of both $False$ and $True$, and by $\phi_1, \dots, \phi_q$ the generating functions of the q other equivalency classes - a detailed study of these classes can be found in an article by Harrison [Har63]. As an example, there are 16 Boolean functions over 2 variables, and we define 4 equivalency classes that are:

$True$	x	$x \wedge y$	x XOR y
$False$	$\bar{x}$	$x \wedge \bar{y}$	x XOR $\bar{y}$
	y	$\bar{x} \wedge y$	
	$\bar{y}$	$\bar{x} \wedge \bar{y}$	
		$x \vee y$	
		$x \vee \bar{y}$	
		$\bar{x} \vee y$	
		$\bar{x} \vee \bar{y}$	

If we consider $k = 3$, there are 256 Boolean functions and 14 equivalency classes; and for $k = 4$, there are 65 536 Boolean functions and 222 equivalency classes. Therefore, the simplification is is significant for numerical computation, even if the equivalence $f \sim \bar{f}$ is enough for the following theoretical proof.

By replacing in the system obtained from Lemma 1 each generating function by the generating function of its equivalency class, we can reduce it to a system of $q+1$ differential equations:

$$\left\{ \begin{array}{rcl} \phi_V' & = & P_V(\phi_V, \phi_1, \dots, \phi_q); \\ \phi_1' & = & P_1(\phi_V, \phi_1, \dots, \phi_q); \\ & \vdots & \\ \phi_q' & = & P_q(\phi_V, \phi_1, \dots, \phi_q). \end{array} \right. \tag{4.3}$$

Then we introduce $\sigma_i = \phi_i \circ \phi_V^{-1}$ for all $i \in [\![1, q]\!]$: indeed, ϕ_V is strictly increasing on the real line and thus invertible on its neighborhood. We have:

$$\left\{ \begin{array}{rcl} \sigma_1'(u) & = & \dfrac{P_1(u, \sigma_1(u), \dots, \sigma_q(u))}{P_V(u, \sigma_1(u), \dots, \sigma_q(u))}; \\ & \vdots & \\ \sigma_q'(u) & = & \dfrac{P_q(u, \sigma_1(u), \dots, \sigma_q(u))}{P_V(u, \sigma_1(u), \dots, \sigma_q(u))}. \end{array} \right. \tag{4.4}$$

where $P_1, \dots, P_q$ and P_V are homogeneous polynomials of degree 2.

To study the solutions of system (4.4), we note that the u^2 monomial only appears in $P_V(u, \sigma_1(u), \dots, \sigma_q(u))$: if both subtrees compute a constant ($True$ or $False$) then, the whole tree computes a constant. The following lemma can be applied to the system (4.4).

LEMMA 2. *If $Y : \mathbb{R} \longrightarrow \mathbb{R}^n$ satisfies the differential equation*

$$Y' = f(x, Y) \text{ with } \lim_{\|Z\|_\infty \longrightarrow \infty} \lim_{x \longrightarrow \infty} f(x, Z) = 0$$

and if $f = (f_1, \dots, f_n)$ with $f_1, \dots, f_n > 0$, then each coordinate of $Y(x)$ is of order $o(x)$.

Proof. This lemma results from standard arguments: we detail a proof in appendix for completeness sake.

Thanks to Lemma 2, we conclude that for all $i \in [\![1, q]\!]$, we have:

$$\sigma_i(u) = o(u) \text{ as } u \longrightarrow +\infty. \tag{4.5}$$

Let us remind that system (4.3) gives $\phi_V' = P_V(\phi_V, \phi_1, \dots, \phi_q)$ with $\phi_i = \sigma_i \circ \phi_V$. Thus, $\phi_V' = \mathcal{G}(\phi_V)$, where $\mathcal{G}(w) = P_V(w, \sigma_1(w), \dots, \sigma_q(w))$. Therefore, ϕ_V satisfies the hypothesis of

LEMMA 3. *If y satisfies the differential equation $y' = G(y)$ where G is non negative and $G(x) \sim cx^2$ when x tends to infinity, then there exists x_0 such that*

$$y(x) \sim \frac{1}{c\,(x_0 - x)} \text{ when } x \longrightarrow x_0.$$

Proof. See Appendix.

Thanks to Lemma 3, we conclude that there exists a constant c and a real number u_0 such that, for real values of u:

$$\phi_V(u) \sim \frac{1}{c\,(u_0 - u)} \text{ when } u \longrightarrow u_0.$$

FACT 1. *Let us remark that, since all coefficients of ϕ_V are less than 1, $u_0 \geq 1$.*

Moreover, thanks to (4.5), we have, for real values of u:

$$\phi_i(u) \overset{u \longrightarrow u_0}{=} o\left(\frac{c}{(u - u_0)}\right).$$

These asymptotic behaviours only hold on the real line. Consequently, we cannot use the classical transfer lemma of Flajolet and Odlyzko [FO90] (detailed in [FS09, page 389]), but, thanks to a standard tauberian

theorem (see for example [Har49, page 155]), we obtain that:

(4.6)
$$\begin{cases} \sum_{i=1}^{n} p_{i,k}(True)u_0^i = \sum_{i=1}^{n} p_{i,k}(False)u_0^i \sim cn \\ \sum_{i=1}^{n} p_{i,k}(f)u_0^i = o(n) \text{ for all } f \notin \{True, False\}. \end{cases}$$

when $n \longrightarrow \infty$. Therefore,

$$\sum_{f\in\mathcal{F}_k}\sum_{i=1}^{n} p_{i,k}(f)u_0^i = \sum_{i=1}^{n}\left(\sum_{f\in\mathcal{F}_k} p_{i,k}(f)\right)u_0^i = \sum_{i=1}^{n} u_0^i$$
$$= \begin{cases} u_0\dfrac{u_0^{n+1}-1}{u_0-1} & \text{if } u_0 > 1 \\ n & \text{if } u_0 = 1 \end{cases}$$

in addition, thanks to (4.6), we have that, when n tends to infinity:

$$\begin{aligned} &\sum_{f\in\mathcal{F}_k}\sum_{i=1}^{n} p_{i,k}(f)u_0^i \\ &\quad\sim \sum_{i=1}^{n} p_{i,k}(True)u_0^i + \sum_{i=1}^{n} p_{i,k}(False)u_0^i \\ &\quad\sim 2cn. \end{aligned}$$

We can thus conclude that $u_0 = 1$ and $c = \frac{1}{2}$ and

$$\sum_{i=1}^{n} p_{i,k}(True) \sim \frac{n}{2} \text{ when } n \longrightarrow \infty. \tag{4.7}$$

To conclude, let us remind that, thanks to (4.1),

$$\begin{aligned} p_{n+1,k}(True) &= \frac{1}{2}\sum_{g\wedge h=True}\sum_{i=0}^{n}\frac{1}{n+1}p_{i,k}(g)p_{n-i,k}(h) \\ &+\frac{1}{2}\sum_{g\vee h=True}\sum_{i=0}^{n}\frac{1}{n+1}p_{i,k}(g)p_{n-i,k}(h) \\ &\geq \frac{1}{2(n+1)}\sum_{i=1}^{n} p_{i,k}(True)p_{n-i,k}(True) \\ &+\frac{1}{n+1}\sum_{i=1}^{n} p_{i,k}(True)(1-p_{n-i,k}(True) \\ &+\frac{1}{2(n+1)}\sum_{i=1}^{n} p_{i,k}(True)p_{n-i,k}(True) \end{aligned}$$

where the first term of the sum stands for the probability to compute $True$ knowing that the root is labelled by $\wedge$ and the other terms are less than the same probability knowing that the root is labelled by $\vee$. We finally obtain that:

$$p_{n+1,k}(True) \geq \frac{1}{n+1}\sum_{i=1}^{n} p_{n,i}(True) \xrightarrow{n\to\infty} \frac{1}{2}$$

thanks to (4.7). Thus we have proved $p_{n,k}(True) \longrightarrow \frac{1}{2}$ when n tends to infinity, which is the first assertion of Theorem 1. The convergence speed would follow from a second-order evaluation of the differential system; however it can be more simply obtained from the probabilistic approach presented below.

4.2 The probabilistic approach. The idea of this proof - due to Pittel [Pit84] - is to embed the discrete process of the growing tree into continuous time by using exponential clocks.

Instead of growing step by step at times $(1, 2, \ldots, n, \ldots)$, the tree grows at random continuous times: each leaf grows independantly from the others after an exponentially distributed time. We thus define a continuous process of trees - denoted by $(\mathcal{Y}_t)_{t\geq 0}$ and named the *Yule tree* (c.f. Definition 4). The link with the (discrete) growing tree is the following: if we consider the sequence of the different values taken by the continuous process (it is a sequence of trees), then this sequence is a growing tree. This property is due to the use of independant and exponentially distributied time of growth. Thus, studying the continuous time process will give information about the (discrete) growing tree.

Moreover, the Yule tree has a property which was wrong in discrete time and which plainly justifies the continuous time embedding : it gives independance between the right and left subtrees at each node of the tree. It is the key of the following proof.

DEFINITION 4. *A* Yule tree *is a continuous time process of binary trees $(\mathcal{Y}_t)_{t\geq 0}$ growing according to the following rules:*

- *$\mathcal{Y}_0$ is a single root;*
- *each leaf of $\mathcal{Y}_t$ gives birth to two sons at the end of a random time following an exponential law of parameter 1, independently from the other leaves.*

DEFINITION 5. *A* labelled Yule tree *is a continuous time process $(\mathcal{Z}_t)_{t\geq 0}$ of labelled binary trees, which evolves according to the following rules:*

- *the underlying binary tree is a Yule tree;*
- *each new leaf is labelled by a literal chosen uniformly at random into $\{x_1, \bar{x_1}, \ldots, x_k, \bar{x_k}\}$;*
- *each new internal node is labelled by $\wedge$ or $\vee$ uniformly at random;*
- *each labelling is independent from the others.*

Let us denote by $\mathtt{P}_t$ the image by Φ[3] of the law of $\mathcal{Z}_t$.

[3] let us remind that *Phi* is the surjective mapping from $\mathcal{E}_k$ to $\mathcal{F}_k$ such that $\Phi(\gamma) = f$ if and only if γ represents (or computes) f

FACT 2. *For all $t \geq 0$, let us denote by $n(t)$ the number of internal nodes of $\mathcal{Z}_t$. Then, $\mathcal{Z}_t$ has the same law as $\mathcal{T}_{n(t)}$: it is a growing tree of size $n(t)$.*

To prove Theorem 1, the idea - inspired from an article about balanced binary trees [FGG09] - is to consider the probability that two different assignments have distinct images by the random Boolean function, and to prove that it tends to 0 as t tends to infinity. Therefore, only constant functions - i.e. $True$ and $False$ - will be charged by the asymptotic distribution.

Let $a = (a_1, \ldots, a_k)$ and $b = (b_1, \ldots, b_k)$ be two distinct elements of $\{0,1\}^k$, which means two different assignments of the k variables. Let α and β be two elements of $\{0,1\}$. For all $t \geq 0$, we denote $\mathtt{P}_t^{\alpha\beta}(a,b) = \mathtt{P}_t\left(f(a) = \alpha \text{ and } f(b) = \beta\right)$.

FACT 3. *Thanks to the symmetries between $\wedge$ and $\vee$ and the variables and their negations, we get $\mathtt{P}_t^{01}(a,b) = \mathtt{P}_t^{10}(a,b)$ and $\mathtt{P}_t^{00}(a,b) = \mathtt{P}_t^{11}(a,b)$. Indeed, the probability to compute f or its negation $\bar{f}$ are the same since $\wedge$ and $\vee$ occur with the same probabilty at each internal node and a variable and its negation occur with the same probability at each leaf.*

Conditionning on the time when the root's clock (which has an exponential law of parameter 1) rings, we get:

$$\begin{aligned}\mathtt{P}_t^{10} = & \sum_{i=1}^{k} \frac{\mathtt{e}^{-t}}{2k}\left(\mathbb{1}_{\{a_i=1 \text{ and } b_i=0\}} + \mathbb{1}_{\{a_i=0 \text{ and } b_i=1\}}\right) \\ & + \frac{1}{2}\int_0^t \left(\mathtt{P}_{t-s}^{11}\mathtt{P}_{t-s}^{10} + \mathtt{P}_{t-s}^{10}(\mathtt{P}_{t-s}^{11} + \mathtt{P}_{t-s}^{01})\right. \\ & \left. + \mathtt{P}_{t-s}^{01}(\mathtt{P}_{t-s}^{00} + \mathtt{P}_{t-s}^{10}) + \mathtt{P}_{t-s}^{00}\mathtt{P}_{t-s}^{10}\right)\mathtt{e}^{-s}ds\end{aligned}$$

where the fisrt term of the sum stand for the probability that $f(a) = 1$ and $f(b) = 0$ knowing that the Yule tree is still reduced to its root at time t, and the second term is the probability of the same event, knowing that the root has given birth to two sons before time t. In the second term, we look at the different possibilities to get $f(a) = 1$ and $f(b) = 0$, depending on the value of the root's label ($\wedge$ or $\vee$ with probability $\frac{1}{2}$) and on the values of the two subtrees for the affectations a and b. Simplifying, we get:

$$\mathtt{P}_t^{10} = \frac{\mathtt{e}^{-t}}{2k}c_{a,b} + \mathtt{e}^{-t}\int_0^t \left(\mathtt{P}_s^{10} - (\mathtt{P}_s^{10})^2\right)\mathtt{e}^s ds$$

where $c_{a,b} = \sum_{i=1}^{k}\left(\mathbb{1}_{\{a_i=1 \text{ and } b_i=0\}} + \mathbb{1}_{\{a_i=0 \text{ and } b_i=1\}}\right)$ is a constant depending only on a and b. Let $\pi_{a,b}(t) = \mathtt{P}_t^{10}(a,b)$. We have:

$$\mathtt{e}^t\pi_{a,b}(t) = \frac{c_{a,b}}{2k} + \int_0^t \left(\pi_{a,b}(s) - \pi_{a,b}(s)^2\right)\mathtt{e}^s ds. \tag{4.8}$$

Therefore, we have the following result on $\pi_{a,b}(t)$:

PROPOSITION 1.
- *If $a \neq b$ then $\pi_{a,b}(t) = \frac{1}{t+t_0}$ where $t_0 = \frac{2k}{c_{a,b}}$.*
- *If $a = b$, then $\pi_{a,a}(t)$ is the constant function equal to zero.*

Thus, $\pi_{a,b}(t) = \mathtt{P}_t^{10}(a,b) \longrightarrow 0$ for all a, b in $\{0,1\}^k$.

Proof. We can easily show that $\pi_{a,b}$ is differentiable, and thus, thanks to (4.8), we get $\pi'_{a,b} + \pi^2_{a,b} = 0$. Let us remark that if $a \neq b$ then there exist $i \in [\![1,k]\!]$ such that $a_i \neq b_i$, i.e. $a_i = 1$ and $b_i = 0$ or $a_i = 0$ and $b_i = 1$. Therefore, $c_{a,b} \geq \mathbb{1}_{\{a_i=1 \text{ and } b_i=0\}} + \mathbb{1}_{\{a_i=0 \text{ and } b_i=1\}} = 1$ and thus $\pi_{a,b}(0) = \frac{c_{a,b}}{2k} > 0$, thus $\pi_{a,b}(t) = \frac{1}{t+t_0}$ where $t_0 = \frac{2k}{c_{a,b}}$.

If $a = b$, $\pi_{a,a}(0) = 0$ and we get that $\pi_{a,a}(t) = 0$ for all t: a single element a cannot have two different images by a function f.

To conclude about the convergence of $\mathtt{P}_t$ when t tends to infinity, we only have to note that:

$$\begin{aligned}\mathtt{P}_t(\mathcal{F}_k \setminus \{True, False\}) &\leq \sum_{(a,b),a\neq b} \mathtt{P}_t(f(a) = 1 \text{ et } f(b) = 0) \\ &\leq 2^k(2^k - 1)\sup_{(a,b)} \mathtt{P}_t^{10}(a,b) \\ &\leq \frac{2^k(2^k-1)}{t}\end{aligned}$$

Then, for all function $f \notin \{True, False\}$, we have: $\lim_{t\to+\infty} \mathtt{P}_t(f) = 0$. Moreover, $\mathtt{P}_t(\{True, False\}) \geq \left(1 - \frac{2^k(2^k-1)}{t}\right)$, which leads to $\lim_{t\to+\infty} \mathtt{P}_t(True) + \mathtt{P}_t(False) \geq 1$, i.e. $\lim_{t\to+\infty} \mathtt{P}_t(True) = \lim_{t\to+\infty} \mathtt{P}_t(False) = \frac{1}{2}$. Thus, $\mathtt{P}_t$ tends to a limit distribution $p_k = \frac{1}{2}\delta_{True} + \frac{1}{2}\delta_{False}$ with a convergence speed of order $\frac{1}{t}$:

$$\|\mathtt{P}_t - p_k\|_\infty \leq \frac{2^k(2^k-1)}{t}. \tag{4.9}$$

FACT 4. *If T_n is the random variable defined by $T_n = \inf\{t \geq 0 \text{ such that } n(t) = n\}$ then $|T_n - \ln n|$ tends to zero almost surely as n tend to infinity.*

We are now able to prove the following Proposition; which, giving the convergence speed, ends the proof of Theorem 1:

PROPOSITION 2. *For large enough n: $\|p_{n,k} - p_k\|_\infty \leq \frac{2^k(2^k-1)}{\ln n - \epsilon} = O\left(\frac{1}{\ln n}\right)$, where ϵ is a non-negative constant.*

Proof. We have $p_{n,k} = \mathtt{P}_{T_n}$ almost surely. Since for all

$t \in \mathbb{R}^+$, $1 = \sum_{n\geq 1} \mathbb{1}_{\{T_n \leq t < T_{n+1}\}}$, we get:

$$\begin{aligned}\|\mathtt{P}_t - p_k\|_\infty &= \sum_{n\geq 1} \|\mathtt{P}_t - p_k\|_\infty \mathbb{1}_{\{T_n \leq t < T_{n+1}\}} \\ &= \sum_{n\geq 1} \|\mathtt{P}_{T_n} - p_k\|_\infty \mathbb{1}_{\{T_n \leq t < T_{n+1}\}} \\ &= \sum_{n\geq 1} \|p_{n,k} - p_k\|_\infty \mathbb{1}_{\{T_n \leq t < T_{n+1}\}}.\end{aligned}$$

We can thus deduce from (4.9) that:

$$\sum_{n\geq 1} \|p_{n,k} - p_k\|_\infty \mathbb{1}_{\{T_n \leq t < T_{n+1}\}} \leq \frac{2^k(2^k-1)}{t},$$

which implies $\forall n \geq 1$, $\|p_{n,k} - p_k\|_\infty \mathbb{1}_{\{T_n \leq t < T_{n+1}\}} \leq \frac{2^k(2^k-1)}{t}$. Given $\epsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that $\forall n \geq n_0$, $|T_n - \ln n| \leq \epsilon$. Therefore, for all $n \geq n_0$,

$$\|p_{n,k} - p_k\|_\infty \mathbb{1}_{\{T_n - \ln n \leq t - \ln n \leq T_{n+1} - \ln n\}} \leq \frac{2^k(2^k-1)}{t}.$$

Moreover, there exists $n_1 \in \mathbb{N}$ such that $\forall n \geq n_1$, $|\ln\left(\frac{n+1}{n}\right)| \leq \epsilon$, which gives, for all $n \geq n_2 = \max(n_0, n_1)$,

$$\|p_{n,k} - p_k\|_\infty \mathbb{1}_{\{-\epsilon \leq t - \ln n \leq 2\epsilon\}} \leq \frac{2^k(2^k-1)}{t}.$$

thus, for all $n \geq n_2$, for all $t \in \mathbb{R}^+$, we have:

$$\begin{aligned}\|p_{n,k} - p_k\|_\infty \mathbb{1}_{\{-\epsilon \leq t - \ln n \leq 2\epsilon\}} &\leq \frac{2^k(2^k-1)}{t} \mathbb{1}_{\{-\epsilon \leq t - \ln n \leq 2\epsilon\}} \\ &\leq \frac{2^k(2^k-1)}{\ln n - \epsilon}\end{aligned}$$

Let us fix $t = \ln n$, then, for all $n \geq n_2$, we obtain $\|p_{n,k} - p_k\|_\infty \leq \frac{2^k(2^k-1)}{\ln n - \epsilon}$.

5 Extensions of Theorem 1

In this section, we consider the extension of our results to more general models. In the first model, we to bias the law over the literals in both labelling models (c.f. Definition 2); in the second, we bias the law over the connectives in the $\wedge/\vee$ model, and in the third we study the $\wedge/\vee$ model with only positive literals. These last two labelling models have been studied by Fournier *et al.* [FGG09] in the case of balanced binary trees (binary trees whose leaves are all at the same level). The results we obtain here are very similar to those obtained in the case of balanced trees.

Biasing the law over the literals. We label each node by $\wedge$ or $\vee$ with probability $\frac{1}{2}$ independently from each other. But we now label each leaf according to a law ν such that $\forall i \in [\![1,k]\!]$, $\nu(x_i) = \nu(\bar{x_i}) > 0$, independently from each other. In this case, since the symmetry between the variables and their negations still holds, the behaviour of the induced probability law p_n over $\mathcal{F}_k$ is the same as in the uniform case - when ν is the uniform law over $\{x_1, \bar{x_1}, \ldots, x_k, \bar{x_k}\}$ - studied just before.

Indeed, in both proofs developped beforehand, the modifications appear only in constants - $p_{0,k}(f)$ in (4.2) and $c_{a,b}$ in (4.8) - which disappear when we take the derivative of the equations. Therefore, the result is the same as for the uniform case.

Biaising the law over connectives in the $\wedge/\vee$ model. We define the *biased* model as follows:

- each internal node is labelled according to the law $q\delta_\wedge + (1-q)\delta_\vee$ with $q \in [0,1]$ independently from the other nodes,
- each leaf is labelled according to a law ν over $\{x_1, \bar{x_1}, \ldots, x_k, \bar{x_k}\}$ such that $\forall i \in [\![1,k]\!]$, $\nu(x_i) = \nu(\bar{x_i}) > 0$, independently from the others.

This process defines a new induced distribution $p_{n,k}$ over $\mathcal{F}_k$ whose behaviour is determined in the following:

THEOREM 4. *In the biased model, if $\mathbb{P}(\wedge) = q$, then:*

- *If $q > \frac{1}{2}$, then $p_{n,k} \longrightarrow \delta_{False}$.*
- *If $q < \frac{1}{2}$, then $p_{n,k} \longrightarrow \delta_{True}$.*

Moreover, the convergence speed is of order $O\left(\frac{1}{n^{|2q-1|}}\right)$ in both cases.

REMARK 2. *It is interesting to note that in the balanced case $q = \frac{1}{2}$ (c.f. Theorem 1), the convergence speed is of order $\frac{1}{\ln n}$ while it is of order $\frac{1}{n^{|2q-1|}}$ in Theorem 4.*

Proof. We can again develop two different proofs that are very close to the proofs of Theorem 1 : we develop hereafter the probabilistic one.

The cases $q > \frac{1}{2}$ and $q < \frac{1}{2}$ are symmetric and can be treated in the same way. In the proof, we only consider the $q > \frac{1}{2}$ case. As in the uniform $\wedge/\vee$ model, we can choose between two proofs: the analytic combinatorics one and the probabilistic one. We present the approach via Yule trees since it gives easily the convergence speed.

As before, we consider a labelled Yule tree $(\mathcal{E}_t)_{t\geq 0}$ which induces a law $\mathtt{P}_t$ over $\mathcal{F}_k$ for all $t \geq 0$. Let $a = (a_1, \ldots, a_k) \in \{0,1\}^k$ be an assignment of the k variables. We prove again that the probability that the image of a by a random Boolean function of law $\mathtt{P}_t$ is 1, tends to 0 when t tend to infinity. Therefore, let us study $\pi_a(t) := \mathtt{P}_t(f(a) = 1)$.

Conditionning by the time when the root's clock rings, with a law $\mathcal{E}xp(1)$, we get:

$$\pi_a(t) = \mathsf{e}^{-t}\sum_{i=1}^{k}(\nu(a_i)\mathbb{1}_{a_i=1} + \nu(\bar{a_i})\mathbb{1}_{a_i=0}) + \int_0^t \big[q\,\pi_a(t-s)^2 + (1-q)(2\pi_a(t-s) - \pi_a(t-s)^2)\big]\,\mathsf{e}^{-s}ds,$$

$$\mathsf{e}^t\pi_a(t) = \frac{1}{2} + \int_0^t \left((2q-1)\pi_a(s)^2 + 2(1-q)\pi_a(s)\right)e^s ds.$$

Deriving and taking into account $p \neq \frac{1}{2}$, we get: $\pi_a + \pi_a' = (2q-1)\pi_a^2 + 2(1-q)\pi_a$, from which we deduce $\pi_a' = (2q-1)(\pi_a^2 - \pi_a)$, and finally that $\pi_a(t) = 1 - \frac{1}{\mathsf{e}^{(2q-1)t}+1}$ since $\pi_a(0) = \frac{1}{2}$. Thus,

$$\mathtt{P}_t(\mathcal{F}_k \setminus \{False\}) \leq \sum_a \pi_a(t) \leq 2^k\left(1 - \frac{1}{\mathsf{e}^{(2p-1)t}+1}\right)$$

Thus, since $q > \frac{1}{2}$, $\lim_{t\to+\infty}\mathtt{P}_t(\mathcal{F}_k \setminus \{False\}) = 0$ and we have:

$$\|p_{n,k} - \delta_{False}\|_\infty \leq 2^k\left(1 - \frac{1}{\mathsf{e}^{(2q-1)T_n}+1}\right) = O\left(\frac{1}{n^{2p-1}}\right),$$

thanks to similar aguments as in the proof od Proposition 2.

The $\wedge/\vee$ model with positive literals. A last labelling model for growing trees - the *positive* model - is as follows:

- each internal node is labelled according to the law $q\delta_\wedge + (1-q)\delta_\vee$ with $q \in [0,1]$ independently from the other nodes,
- each leaf is labelled according to a law μ over $\{x_1, \ldots, x_k\}$ independently from the others.

This process defines a new induced law - still denoted $p_{n,k}$ - over $\mathcal{F}_k$ whose behaviour is determined in the following Theorem:

THEOREM 5. *In the positive model, we have:*

- *If $q > \frac{1}{2}$, then $p_{n,k} \longrightarrow \delta_{x_1\wedge\ldots\wedge x_k}$.*
- *If $q < \frac{1}{2}$, then $p_{n,k} \longrightarrow \delta_{x_1\vee\ldots\vee x_k}$.*

And the convergence speed is in both cases of order $O\left(\frac{1}{n^{|2p-1|}}\right)$.

Proof. As in the biased model, the proofs for $q > \frac{1}{2}$ and $q < \frac{1}{2}$ are very similar. We assume $q > \frac{1}{2}$ in the proof. Here again, we only developped the probabilistic approach. By the same computation as in the proof of Theorem 4, we get:

$$\pi_a(t) = \mathtt{P}_t(f(a)=1) = 1 + \frac{1}{\lambda\mathsf{e}^{(2q-1)t} - 1} \text{ for all } t \geq 0$$

or $\pi_a = 1$. If $a = (1,\ldots,1)$ then $\pi_a(0) = \sum_{i=1}^k \mathbb{1}_{a_i=1} = 1$ and $\pi_a(t) = 1$. Thus $\mathtt{P}_t(f(1,\ldots,1)=1) = 1$. Otherwise, if $a \neq (1,\ldots,1)$, then since $q > \frac{1}{2}$, we have $\lim_{t\to\infty}\mathtt{P}_t(f(a)=1) = 0$. Thus, the asymptotic law of the $p_{n,k}$ exists and only charges the function $x_1\wedge\ldots\wedge x_k$.

Actually, Theorem 5 is not complete since we have not studied the case $q = \frac{1}{2}$ which is a natural extension of the $\wedge/\vee$ model. Surprisingly, this last case is the most complicated of the whole study. To state our last theorem, we have to present the definition of a *threshold* function - first introduced in [Ser04] and used in [FGG09]. We show that the asymptotic distribution of the $p_{n,k}$ exists and that its support is included in a finite set of threshold functions.

DEFINITION 6. ([FGG09]) *Let $a = (a_1,\ldots,a_k) \in \{0,1\}^k$. The* weight *of a relatively to the distribution ν is the real number $\omega_\nu(a) = \nu(x_1)a_1 + \ldots + \nu(x_k)a_k$.*

DEFINITION 7. ([FGG09]) *A Boolean function f is a* threshold function *if there exists a real number $\theta \geq 0$ such that $\forall(a_1,\ldots,a_k) \in \{0,1\}^k$, $f(a_1,\ldots,a_k) = 1 \Leftrightarrow \omega_\nu(a) \geq \theta$. We denote by $T_{\nu,\theta}$ the threshold function associated to the constant θ and to the distribution ν.*

THEOREM 6. *Let us number the different elements of $\{0,1\}^k$ in order of increasing weight ω_ν: $\omega_\nu(a^{(1)}) \leq \omega_\nu(a^{(2)}) \leq \ldots \leq \omega_\nu(a^{(2^k)})$. Then, $p_{n,k} \overset{n\to+\infty}{\longrightarrow} \sum_{j=1}^{2^k}\left(\omega_\nu(a^{(j)}) - \omega_\nu(a^{(j-1)})\right)\delta_{T_{\nu,\omega_\nu(a^{(j)})}}$ where $\omega_\nu(a^{(0)}) := 0$.*

Said differently, $p_{n,k}$ tends to an asymptotic distribution law p_k that satisfies: $p_k(T_{\nu,\omega_\nu(a^{(j)})}) = \omega_\nu(a^{(j)}) - \omega_\nu(a^{(j-1)})$ and, if f is a Boolean function different from $T_{\nu,\omega_\nu(a^{(j)})}$ for all $j \in [\![1, 2^k]\!]$, then $p_k(f) = 0$.

Proof. The proof is once again based on Yule trees: we did not handle a proof based on analytic combinatorics. The probabilistic approach is natural in this case, since it is an extension of the proof developped in [FGG09] in the case of balanced trees. Let $\mathcal{E}_t$ be a Yule tree, $a = (a_1,\ldots,a_k)$ and $b = (b_1,\ldots,b_k)$ in $\{0,1\}^k$ be two assignments of the k variables, and α, β in $\{0,1\}$. For all $t \geq 0$, let $\pi_{\alpha\beta}(t) = \mathtt{P}_t(f(a) = \alpha \text{ and } f(b) = \beta)$. Let us compute π_{10} by conditioning on the time when the

root's clock rings.

$$\pi_{10}(t) = \mathsf{e}^{-t}\sum_{i=1}^{k} a_i(1-b_i)\nu(x_i) + \int_0^t \frac{1}{2}\begin{pmatrix} \pi_{11}(t-s)\pi_{10}(t-s) \\ +\pi_{10}(t-s)(\pi_{10}(t-s)+\pi_{11}(t-s)) \\ +\pi_{10}(t-s)(\pi_{10}(t-s)+\pi_{00}(t-s)) \\ +\pi_{10}(t-s)\pi_{00}(t-s) \end{pmatrix}\mathsf{e}^{-s}ds$$

This gives

$$\pi_{10}(t)\mathsf{e}^t = \sum_{i=1}^k a_i(1-b_i)\omega_\nu(x_i) + \int_0^t \left(\pi_{10}(s)^2 + \pi_{10}(s)\pi_{11}(s) + \pi_{10}(s)\pi_{00}(s)\right)\mathsf{e}^s ds.$$

By differentiating and using the obvious relation $\pi_{11} + \pi_{10} + \pi_{01} + \pi_{00} = 1$, we get: $\pi'_{10} = -\pi_{10}\pi_{01}$. Doing the same computation for π_{00}, π_{01} and π_{11}, we obtain the differential system:

$$\left\{\begin{array}{rcl} \pi'_{10} &=& -\pi_{10}\pi_{01}; \\ \pi'_{01} &=& -\pi_{10}\pi_{01}; \\ \pi'_{11} &=& \pi_{10}\pi_{01}; \\ \pi'_{00} &=& \pi_{10}\pi_{01}. \end{array}\right. \tag{5.10}$$

Thanks to (5.10), we can see that $\pi_{10}(t)$ and $\pi_{01}(t)$ are decreasing functions of t; since they are both positive, they have a limit as $t \longrightarrow +\infty$. In the same way, π_{11} and π_{00} are increasing and thus convergent. Let us denote $l_{\alpha\beta} = \lim_{t\to\infty}\pi_{\alpha\beta}(t)$.

Since $\pi_{\alpha\beta}$ is monotone and convergent for t tends to $+\infty$, its derivative tends to zero as $t \longrightarrow +\infty$. thus, taking the limit in system (5.10), we get:

$$l_{10}l_{01} = 0. \tag{5.11}$$

Moreover, $\pi_{10} - \pi_{01}$ is a constant; then,

$$l_{10} - l_{01} = \pi_{10}(0) - \pi_{01}(0) = \omega_\nu(a) - \omega_\nu(b). \tag{5.12}$$

Thus, if $\omega_\nu(a) \geq \omega_\nu(b)$, then, thanks to (5.11) and (5.12), we get: $l_{01} = 0$. If $\omega_\nu(a) \geq \omega_\nu(b)$, then $\mathtt{P}_t(f(a) = 0 \text{ and } f(b) = 1) \longrightarrow 0$ as $t \longrightarrow +\infty$. Said differently, if there exists a and b such that $\omega_\nu(a) \geq \omega_\nu(b)$ and $f(a) = 0$ and $f(b) = 1$, then $p_{n,k}(f) \longrightarrow 0$ as $n \longrightarrow +\infty$. The only Boolean functions weighted by $p_{n,k}$ when n tends to infinity are those verifying $\forall a, b$ such that $\omega_\nu(a) \geq \omega_\nu(b) \Rightarrow f(a) \geq f(b)$. And those functions are threshold functions: only threshold functions can be weighted by the asymptotic law of the $p_{n,k}$, if this law exists.

The calculations we made in the non-uniform positive model can be done again in this case to prove that $\mathtt{P}_t(f(a) = 1)$ is a constant for all a. Thus $\mathtt{P}_t(f(a) = 1) = \omega_\nu(a)$ and for all $j \in [\![1, 2^k]\!]$,

$$p_{n,k}(T_{\nu,\omega_\nu(a^{(1)})}) + \ldots + p_{n,k}(T_{\nu,\omega_\nu(a^{(j)})}) \longrightarrow \omega_\nu(a^{(j)}).$$

Thus $p_{n,k}(T_{\nu,\omega_\nu(a^{(j)})}) \longrightarrow \omega_\nu(a^{(j)}) - \omega_\nu(a^{(j-1)})$, and as $\sum_{j=1}^{2^k}\omega_\nu(a^{(j)}) - \omega_\nu(a^{(j-1)}) = 1$, we indeed proved Theorem 6.

6 Simple tautologies

6.1 Proof of Theorem 3

Proof. The proof has two steps: first, we compute the law of the number f_n of premises that are reduced to a simple leaf in a growing tree of size n - we call them *nice premises*; and second, we calculate the probability to get a simple tautology by conditioning over the number of nice premises.

The first step can be handled by modelling the system by a Pólya urn. Indeed, let us consider an urn containing three kinds of balls, representing three kinds of leaves of the tree. The white balls, standing for the nice premises; one red ball, standing for the goal of the Boolean expression; and some black balls standing for the other leaves. When the growing tree grows, we choose one of its leaves (i.e. one of the balls) uniformly at random, and

- if we choose the red ball, then we put it back into the urn and add a white ball (i.e. a nice premise);
- if we choose a white ball, then we remove it from the urn and add two black balls into the urn;
- if we choose a black ball, then we put it back into the urn and add another black ball.

Morcrette [Mor10] has shown that (see also [FGP05] for a general approach by analytic combinatorics method):

$$\mathbb{P}(f_n = q) = \frac{1}{q!}\left(\mathsf{e}^{-1} - \sum_{j\geq n+1-q}\frac{(-1)^j}{j!}\right). \tag{6.13}$$

Let us now calculate $\mathbb{P}_{n,k}(ST_k)$ by conditioning over the number of nice premises: $\mathbb{P}_{n,k}(ST_k) = \sum_{q=1}^n \mathbb{P}(f_n = q)\left(1 - \left(1-\frac{1}{k}\right)^q\right)$ since $\left(1 - \left(1-\frac{1}{k}\right)^q\right)$ is the probability that one of the nice premises is labelled by the same label as the goal of the Boolean expression. Let $c = \left(1 - \frac{1}{k}\right)$.

$$\begin{aligned} \mathbb{P}_{n,k}(ST_k) &= \sum_{q=1}^n \frac{1}{q!}\left(\mathsf{e}^{-1} - \sum_{j\geq n+1-q}\frac{(-1)^j}{j!}\right)(1-c^q) \\ &= \sum_{q=1}^n \frac{\mathsf{e}^{-1}}{q!}(1-c^q) - \sum_{q=1}^n \frac{(1-c^q)}{q!}\sum_{j\geq n+1-q}\frac{(-1)^j}{j!} \\ &= \mathsf{e}^{-1}(\mathsf{e} - 1 - \mathsf{e}^c + 1) - \mathsf{e}^{-1}\sum_{q=n+1}^{\infty}\frac{(1-c^q)}{q!} - R_n \\ &= 1 - \mathsf{e}^{-1/k} - S_n - R_n \end{aligned}$$

where $R_n = \sum_{q=1}^{n} \frac{(1-c^q)}{q!} \sum_{j\geq n+1-q} \frac{(-1)^j}{j!}$ and $S_n = \mathsf{e}^{-1} \sum_{q=n+1}^{\infty} \frac{(1-c^q)}{q!}$. Let show that R_n and S_n tend to zero as n tends to infinity: $\sum_{j\geq n+1-q} \frac{(-1)^j}{j!}$ is an alternating series, thus $|\sum_{j\geq n-q} \frac{(-1)^j}{j!}| \leq \frac{1}{(n+1-q)!}$ and:

$$\begin{aligned}
|R_n| &\leq \sum_{q=1}^{n} \frac{(1-c^q)}{q!(n+1-q)!} \\
&\leq \frac{1}{(n+1)!} \sum_{q=1}^{n} \binom{n+1}{q} (1-c^q) \\
&\leq \frac{\left(2^{n+1} - (1+c)^{n+1}\right)}{(n+1)!} \\
&\xrightarrow{n\to\infty} 0.
\end{aligned}$$

Moreover, S_n is the remainder of a convergent series, thus $S_n \xrightarrow{n\to\infty} 0$.

6.2 Negative literals Simple tautologies have been studied in another labelling model: an implication model where negative literals are allowed [FGGZ10]. We can prove, just as we did for the classical implication model, that $p_{n,k}$ tends to δ_{True} when n tends to infinity. In this new labelling model, there are two kinds of simple tautologies: simple tautologies of first kind, defined in the same way as in the classic labelling model (c.f. Definition 3), and simple tautologies of second kind:

Definition 8. ([FGGZ10]) *A tautology of second kind is a Boolean expression in which two nice premises are labelled respectively with a variable and its negation. We denote by ST_k^1 (resp. ST_k^2) the set of simple tautologies of first kind (resp. second kind).*

It has been shown that in both the Catalan trees and in the Galton-Watson model, all the tautologies are simple tautologies of either first or second kind, asymptotically when k tends to infinity [FGGZ10]. We show that it is not the case in the growing tree model:

Theorem 7. *We have $\mathbb{P}_{n,k}(ST_k^1) \xrightarrow{n\to+\infty} 1 - \mathsf{e}^{-1/2k} \sim \frac{1}{2k}$, when $k \to +\infty$;*
and $\mathbb{P}_{n,k}(ST_k^2) \xrightarrow{n\to+\infty} 1 - \frac{1}{\mathsf{e}}(2\mathsf{e}^{1/2k} - 1)^k \sim \frac{1}{4k}$, when $k \to +\infty$.

Therefore, in the implication model with positive and negative literals, there are again other tautologies charged by the growing tree distribution, asymptotically when k tends to infinity.

Proof. The first statement of Theorem 7 can be shown in the same way as Theorem 3. Let us consider the second one. The equation (6.13) still holds: we only have to compute the probability that two nice premises among q are labelled by a variable and its negation. The idea is to reformulate this problem in terms of a birthday problem [JK97]. We have to assign q balls (labels of the nice premises) into k urns (variables). The balls are either white or black (positive or negative literals) with probability one half independently from the others. The probability that at least one urn contains at least one black ball and one white ball is the probability to get a simple tautology of second kind.

By a symbolic method, we get that, if $\alpha_{r,q}$ is the number of assignments of q balls into k urns that realize r times a simple tautology of second kind, then $\Phi(t,z) := \sum_{r,q} \alpha_{r,q} z^r \frac{t^q}{q!} = (z(\mathsf{e}^t - 1)^2 + 2\mathsf{e}^t - 1)^k$. As $\Phi(t,0)$ is the generating function of the number of assignments of the q balls that do not realize a simple tautology of second kind, we get that: $\mathbb{P}_{n,k}(\overline{ST_k^2} | f_n = q) = \frac{[\frac{t^q}{q!}]\Phi(t,0)}{(2k)^q} = \frac{\alpha_{0,q}}{(2k)^q}$, which, thanks to (6.13) gives, after calculation:

$$\mathbb{P}_{n,k}(\overline{ST_k^2}) = \underbrace{\mathsf{e}^{-1} \sum_{q=0}^{n} \frac{\alpha_{0,q}}{q!\,(2k)^q}}_{Q_n} - \underbrace{\sum_{q=0}^{n} \frac{\alpha_{0,q}}{q!\,(2k)^q} \sum_{j\geq n+1-q} \frac{(-1)^j}{j!}}_{R_n}.$$

We then can easily show that R_n tends to zero as n tends to infinity, and that $Q_n \xrightarrow{n\to+\infty} \frac{1}{\mathsf{e}}\Phi(\frac{1}{2k}, 0)$. Thus,

$$\mathbb{P}_{n,k}(\overline{ST_k^2}) \xrightarrow{n\to+\infty} \frac{1}{\mathsf{e}}\Phi(\frac{1}{2k}, 0) = \frac{1}{\mathsf{e}}(2\mathsf{e}^{1/2k} - 1)^k \sim \frac{1}{4k},$$

when $k \to +\infty$.

7 Conclusions and perspectives

We have studied in this paper the behaviour of the growing tree under different labelling systems. This behaviour is very different from the Catalan trees and the Galton-Watson trees [CFGG04], but very similar to the balanced trees behaviour [FGG09]. Indeed, Theorems 1, 2, 4, 5 and 6 are true for balanced trees. The similarity may be intuitively explained by the fact that the growing trees have a saturation level of order $\ln n$, i.e. a saturation level tending to infinity as the size of the tree is growing to infinity: roughly speaking, a big growing tree contains a big balanced tree. On the contrary, Catalan trees and Galton-Watson trees have a saturation level of order $\Theta(1)$. But the precise relationship between growing trees and balanced trees still needs to be made precise.

To sum up about the methods used in this paper, we have to note that the analytic combinatorics approach is useful when the asymptotic law only charges constants (Theorems 1, 2 and 4). For the other results, we prefer the probabilistic one, as it is more natural and give

proofs that are very similar to those developped in [FGG09] for balanced trees.

Acknoledgements We are grateful to Michael Drmota for fruitful discussions about this work, and to the referees for their interesting and useful remarks about this work.

References

[CFGG04] B. Chauvin, Ph. Flajolet, D. Gardy, and B. Gittenberger. And/or trees revisited. *Combinatorics, Probability and Computing*, 13(4-5):475–497, Juillet-Septembre 2004.

[Drm97] M. Drmota. Systems of fonctional equations. *Random Structures and Algorithms*, 10(1-2):103–124, 1997.

[FGG09] H. Fournier, D. Gardy, and A. Génitrini. Balanced and/or trees and linear threshold functions. In *5th SIAM Workshop on Analytic and Combinatorics (ANALCO)*, pages 51–57, 2009.

[FGGZ07] H. Fournier, D. Gardy, A. Génitrini, and M. Zaionc. Classical and intuitionistic logics are asymptotically identical. In *Computer Science Logic*, pages 177–193. Springer, 2007.

[FGGZ10] H. Fournier, D. Gardy, A. Génitrini, and M. Zaionc. Simple tautologies over implication with negative literal. *Journal Mathematical Logic Quarterly*, 56(4):388–396, 2010.

[FGP05] Ph. Flajolet, J. Gabarró, and H. Pekari. Analytic urns. *Annals of Probability*, 33(3):1200–1233, 2005.

[FO90] Ph. Flajolet and A.M. Odlyzko. Singularity analysis of generating functions. *SIAM Journal on discrete mathematics*, 3(2):216–240, 1990.

[FS09] Ph. Flajolet and R. Sedgewick. *Analytic combinatorics*. Cambridge University Press, 2009.

[GG10] A. Génitrini and B. Gittenberger. No Shannon effect on probability distributions on Boolean fonctions induced by Boolean expressions. In *Analysis of Algorithms*, Wien, Austria, Juillet 2010. DMTCS proceedings.

[Har49] G. H. Hardy. *Divergent series*. Oxford University Press, 1949.

[Har63] M. A. Harrison. The number of classes of Boolean functions under groups containing negation. *I.E.E.E Trans. Elect. Comput.*, 12:559–561, 1963.

[JK97] N.L. Johnson and S. Kotz. *Urns models and their application*. Wiley and Sons, 1997.

[KZ04] Z. Kostrzycka and M. Zaionc. Statistics of intuitionnistic versus classical logic. *Studia Logica*, 76(3):307–328, 2004.

[Lal93] S. P Lalley. Finite range random walk on free groups and homogeneous trees. *The Annals of Probability*, 21(4):2087–2130, 1993.

[LS97] H. Lefmann and P. Savický. Some typical properties of large and/or Boolean formulas. *Random Structures and Algorithms*, 10:337–351, 1997.

[Mor10] B. Morcrette. Combinatoire analytique et modèles d'urnes. Master's thesis, MPRI - Inria Rocquencourt, 2010.

[Pit84] B. Pittel. On growing random binary trees. *Journal of Mathematical Analysis and Applications*, 103(2):461–480, 1984.

[Ser04] R. A. Servedio. Monotone Boolean formulas can approximate monotone linear threshold functions. *Discrete Applied Mathematics*, 142(1-3):181–187, 2004.

[Sha49] C. E. Shannon. The synthesis of two-terminal switching circuits. *Bell System Technical*, 28:59–98, 1949.

[Woo97] A. R. Woods. Coloring rules for finite trees, and probabilities of monadic second order sentences. *Random Structures and Algorithms*, 10(4):453–485, 1997.

Appendix

Proof of Lemma 2

Proof. Let $\epsilon > 0$. We have $\lim_{\|Z\|_\infty \longrightarrow \infty} \lim_{x \longrightarrow \infty} f(x, Z) = 0$; thus there exists y_0 such that $\forall Z$ such that $\|Z\| \geq y_0$,

$$\lim_{x \longrightarrow \infty} f(x, Z) = 0 < \frac{\epsilon}{2},$$

thus there exists y_0 such that for all Z satisfying $\|Z\| \geq y_0$, there exists $x_0(Z)$ such that $\forall x \geq x_0(Z)$, $\|f(x, Z)\|_\infty < \epsilon$.

Let Y be a solution of the differential equation $Y'(x) = f(x, Y(x))$. We assume $f_1, \ldots, f_n > 0$: therefore, each component of Y is strictly increasing.

First case: $\forall x$, $\|Y(x)\|_\infty \leq y_0$. Therefore, $Y(x)$ is bounded, and $Y(x)$ is indeed of order $o(x)$.

Second case: $\exists x_1$ such that $\forall x \geq x_1$, $\|Y(x)\|_\infty \geq y_0$. Let us denote $x_2 = \max(x_0, x_1)$. Then, $\forall x \geq x_2$, $\|f(x, Y)\|_\infty < \epsilon$. By interpreting the following computations *component by component*, we obtain:

$$\begin{aligned} Y(x) &= Y(x_2) + \int_{x_2}^{x} f\underbrace{(x, Y)}_{\leq \epsilon} dx \\ &\leq Y(x_2) + \epsilon(x - x_2). \end{aligned}$$

Said differently, in both cases, $Y(x) = o(x)$ *component by component.*

Proof of Lemma 3

Proof. We have $\frac{dy}{G(y)} = dx$, then

$$\int_{y_0}^{y(x)} \frac{dy}{G(y)} = x - \tilde{x_0} \text{ with } y_0, \tilde{x_0} \text{ such that } y(\tilde{x_0}) = y_0,$$

thus $\displaystyle\int_{y_0}^{\infty} \frac{dy}{G(y)} - \int_{y(x)}^{\infty} \frac{dy}{G(y)} = c_0 - \int_{y(x)}^{\infty} \frac{dy}{G(y)} = x - \tilde{x_0}$.

Since

$$\int_{y(x)}^{\infty} \frac{dy}{G(y)} \sim \frac{c}{y(x)} \text{ when } y(x) \longrightarrow \infty,$$

we deduce $\frac{c}{y(x)} \sim c_0 + \tilde{x_0} - x$ when $y(x) \longrightarrow \infty$, said differently,

$$y(x) \sim \frac{1}{c(x_0 - x)} \text{ when } x \longrightarrow x_0.$$