

The growing trees distribution on Boolean functions.*

Brigitte Chauvin[†], Danièle Gardy[‡] and Cécile Mailler[§]

October 1, 2010

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 trees* 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.

Keywords

Boolean functions, probability distribution, binary trees, BST model of growth, Yule trees.

*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: danièle.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 Introduction

A Boolean function of k variables is a function $f : \{0,1\}^k \rightarrow \{0,1\}$ 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 $x_1, \bar{x}_1, \dots, x_k, \bar{x}_k$ 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. The size of a binary tree is the number of its internal nodes¹. Let $\mathcal{U}_{n,k}$ be the uniform distribution over labelled binary trees of size n over k variables and $\mu_{n,k}$ the distribution induced by $\mathcal{U}_{n,k}$ 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 trees* distribution. This distribution is inspired by the Binary Search Tree growing process. We then label the binary trees 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 part 2, we define precisely the growing trees model and the two labelling models, before stating our main results in the following part: Theorems 1 and 2 (proved in part 4) give the convergence of the growing trees distribution to a distribution which support is included into the set of constant functions in both labelling models, and Theorem 3 (proved in part 6) deals with the proportion of *simple tautologies* among tautologies. Finally, some extensions of the two labelling models are presented respectively in parts 5 and 6.

2 Growing trees

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

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

- \mathcal{T}_0 is reduced to its root.
- Given \mathcal{T}_i , we choose uniformly at random a leave 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 of them: 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 leave is labelled by a literal chosen uniformly at random in the set $\{x_1, \bar{x}_1, \dots, x_k, \bar{x}_k\}$.

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

- 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 $\{x_1, \dots, x_k\}$.

Each model defines a distribution, respectively $\mathbb{P}_{n,k}^{\wedge/\vee}$ and $\mathbb{P}_{n,k}^{\rightarrow}$ - 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 follows²:

$$\Phi(\gamma) = f \text{ if and only if } \gamma \text{ represents (or computes) } f.$$

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

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?

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 $O(\frac{1}{\ln n})$. We prove the following theorems:

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

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

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

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 fonction chosen at random according to p_k is the complexity of a constant: 1.

Remark: 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 $\{x_1, \dots, x_k\}$. A function can be expressed in this model if and only if there exists $i \in \llbracket 1, k \rrbracket$ 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: $A_1 \rightarrow (A_2 \rightarrow \dots (A_p \rightarrow \alpha))$. The subtrees A_1, \dots, 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 trees model, we get a different behaviour:

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

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

²Of course, this is not a one-to-one mapping since a same fonction can be represented by different expressions. For example, $\Phi(x_1 \wedge \bar{x}_1) = \Phi(x_2 \wedge \bar{x}_2) = False$.

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 trees 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 complicated 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]. Given a Boolean function f , we define its generating function as:

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

where $p_n(f)$ is the probability that the 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 $\frac{1}{n}$. Therefore, by conditioning on the size of the left subtree, we obtain the following formula:

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

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

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

Deriving formula (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 verified 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. [CFG04]). 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 trees 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 with the same generating function. One class is $\{False, True\}$: let us denote by ϕ_V the generating function of both *False* and *True*, and by ϕ_1, \dots, ϕ_q the generating functions of the q other equivalency classes - a detailed study of these classes can be found in an article by Harrison [Har64].

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:

$$\begin{cases} \phi'_V &= P_V(\phi_V, \phi_1, \dots, \phi_q); \\ \phi'_1 &= P_1(\phi_V, \phi_1, \dots, \phi_q); \\ &\vdots \\ \phi'_q &= P_q(\phi_V, \phi_1, \dots, \phi_q). \end{cases} \quad (3)$$

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

$$\begin{cases} \sigma'_1(u) &= \frac{P_1(u, \sigma_1(u), \dots, \sigma_q(u))}{P_V(u, \sigma_1(u), \dots, \sigma_q(u))}; \\ &\vdots \\ \sigma'_q(u) &= \frac{P_q(u, \sigma_1(u), \dots, \sigma_q(u))}{P_V(u, \sigma_1(u), \dots, \sigma_q(u))}. \end{cases} \quad (4)$$

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

To study the solutions of system (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. Thus, the following lemma can be applied to the system (4).

Lemma 2. *If $Y : \mathbb{R} \rightarrow \mathbb{R}^n$ verifies the differential equation*

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

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. \square

Thanks to Lemma 2, we claim that for all $i \in \llbracket 1, q \rrbracket$, we have:

$$\sigma_i(u) = o(u) \text{ as } u \rightarrow +\infty. \quad (5)$$

Let us remind that system (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 verifies the hypothesis of

Lemma 3. *If y verifies the differential equation $y' = G(y)$ where G is non negative and $G(x) \stackrel{x \rightarrow \infty}{\sim} cx^2$, then there exists x_0 such that*

$$y(x) \stackrel{x \rightarrow x_0}{\sim} \frac{1}{c(x_0 - x)}.$$

Proof. See Appendix. \square

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

$$\phi_V(u) \stackrel{u \rightarrow u_0}{\sim} \frac{1}{c(u_0 - u)}.$$

By applying a classical transfer lemma [FO90] (detailed in [FS09, page 389]), we state that:

$$p_{n,k}(True) = p_{n,k}(False) = [u^n] \phi_V(u) \stackrel{n \rightarrow \infty}{\sim} \frac{\gamma}{u_0^n}, \quad (6)$$

where γ is a constant. Moreover, thanks to (5), we have: $\phi_i(u) \stackrel{u \rightarrow u_0}{\sim} o\left(\frac{c}{(u-u_0)}\right)$; applying again the transfer lemma, we get: $[u^n] \phi_i(u) \stackrel{n \rightarrow \infty}{\sim} o(1)$. Hence, $u_0 = 1$, finally, since the sum of the 2^{2^k} generating functions is $\frac{1}{1-z}$ and using (6), we obtain $\gamma = \frac{1}{2}$. We have hence proved

$$p_{n,k}(True) = [u^n] \phi_V(u) \stackrel{n \rightarrow +\infty}{\sim} \frac{1}{2}, \quad (7)$$

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. This embedding gives us independence between the right and left subtrees at each node of the tree.

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 $(\mathcal{E}_t)_{t \geq 0}$ 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 $\{x_1, \bar{x}_1, \dots, x_k, \bar{x}_k\}$;
- each new internal node is labelled by \wedge or \vee uniformly at random;
- each labelling is independent from the others.

Let us denote by \mathbf{P}_t the image by Φ of the law of \mathcal{E}_t .

Fact: For all $t \geq 0$, let us denote by $n(t)$ the number of internal nodes of \mathcal{E}_t . Then, \mathcal{E}_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 $a = (a_1, \dots, a_k)$ and $b = (b_1, \dots, b_k)$ 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 $\mathbf{P}_t^{\alpha\beta}(a, b) = \mathbf{P}_t(f(a) = \alpha \text{ and } f(b) = \beta)$.

Fact: Thanks to the symmetries between \wedge and \vee and the variables and their negations, we get $\mathbf{P}_t^{01}(a, b) = \mathbf{P}_t^{10}(a, b)$ and $\mathbf{P}_t^{00}(a, b) = \mathbf{P}_t^{11}(a, b)$.

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

$$\begin{aligned} \mathbf{P}_t^{10} &= \sum_{i=1}^k \frac{e^{-t}}{2k} (\mathbb{1}_{\{a_i=1 \text{ et } b_i=0\}} + \mathbb{1}_{\{a_i=0 \text{ et } b_i=1\}}) \\ &\quad + \frac{1}{2} \int_0^t (\mathbf{P}_{t-s}^{11} \mathbf{P}_{t-s}^{10} + \mathbf{P}_{t-s}^{10} (\mathbf{P}_{t-s}^{11} + \mathbf{P}_{t-s}^{01}) + \mathbf{P}_{t-s}^{01} (\mathbf{P}_{t-s}^{00} + \mathbf{P}_{t-s}^{10}) + \mathbf{P}_{t-s}^{00} \mathbf{P}_{t-s}^{10}) e^{-s} ds \\ &= \frac{e^{-t}}{2k} c_{a,b} + e^{-t} \int_0^t (\mathbf{P}_s^{10} - (\mathbf{P}_s^{10})^2) e^s ds \end{aligned}$$

where $c_{a,b}$ is a constant depending only on a and b . Let $\pi_{a,b}(t) = \mathbf{P}_t^{10}(a, b)$. We have:

$$e^t \pi_{a,b}(t) = \frac{c_{a,b}}{2k} + \int_0^t (\pi_{a,b}(s) - \pi_{a,b}(s)^2) e^s ds. \quad (8)$$

We can easily show that $\pi_{a,b}$ is differentiable, and thus, thanks to (8), we get $\pi'_{a,b} + \pi_{a,b}^2 = 0$. Let us remark that if $a \neq b$ then $\pi_{a,b}(0) = \frac{c_{a,b}}{2k} \neq 0$, thus $\pi_{a,b}(t) = \frac{1}{t-t_0}$ where $t_0 = \frac{2k}{c_{a,b}}$. Thus, $\mathbf{P}_t^{10}(a, b) \longrightarrow 0$ for all $a, b \in \{0, 1\}^k$.

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

To conclude, we only have to note that:

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

Moreover, $\mathbb{P}_t(\{True, False\}) \geq \frac{1}{2} \left(1 - \frac{2^k(2^k - 1)}{t - \frac{1}{2}}\right)$, which leads to $\lim_{t \rightarrow +\infty} \mathbb{P}_t(True) + \mathbb{P}_t(False) \geq 1$, i.e. $\lim_{t \rightarrow +\infty} \mathbb{P}_t(True) = \lim_{t \rightarrow +\infty} \mathbb{P}_t(False) = \frac{1}{2}$. Thus, \mathbb{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 $\frac{1}{t}$: $\|\mathbb{P}_t - p_k\|_\infty \leq \frac{2^k(2^k - 1)}{t - \frac{1}{2}}$.

Fact: 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.

Moreover, $p_n = \mathbb{P}_{T_n}$ almost surely and we thus can easily see that for large enough n : $\|p_{n,k} - p_k\|_\infty \leq \frac{2^k(2^k - 1)}{\ln n - \frac{3}{4}} = O\left(\frac{1}{\ln n}\right)$.

5 Extensions of Theorem 1

In this section, we consider the extension of our results to more general models. The first way is to bias the law over the literals in both labelling models (c.f. Definition 2); the second is to bias the law over the connectives in the \wedge/\vee model, and the third to 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 all leaves are at the same level). The results we have are very similar to those obtained in the case of balanced trees.

Biasing the law over the literals.

Let us study a slightly different labelling model. We still 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 $\forall i \in \llbracket 1, k \rrbracket$, $\nu(x_i) = \nu(\bar{x}_i) > 0$, 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 $\{x_1, \bar{x}_1, \dots, x_k, \bar{x}_k\}$ - studied just before.

Indeed, in both proofs developed before, the modifications appear only in constants - $p_{0,k}(f)$ in (2) and $c_{a,b}$ in (8) - which disappear when we take the derivative of the equations. Therefore, the result is the same as for the uniform case.

Biasing the law over connectives in the \wedge/\vee model.

Let us introduce another labelling model over growing trees of size n : the *biased* model.

- Each internal node is labelled according to the law $p\delta_\wedge + (1 - p)\delta_\vee$ with $p \in [0, 1]$ independently from the other nodes.
- Each leaf is labelled according to a law ν over $\{x_1, \bar{x}_1, \dots, x_k, \bar{x}_k\}$ such that $\forall i \in \llbracket 1, k \rrbracket$, $\nu(x_i) = \nu(\bar{x}_i) > 0$, 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 $\mathbb{P}(\wedge) = p$, then:*

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

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

Remark: It is interesting to note that in the balanced case $p = \frac{1}{2}$ (c.f. Theorem 1), the convergence speed is of order $\frac{1}{\ln n}$ while it is of order $\frac{1}{n}$ in Theorem 4.

Proof. We can again develop two different proofs that are very close to the proofs of Theorem 1. See Appendix for a detailed version of the probabilistic approach. \square

The \wedge/\vee model with positive literals

Let us introduce a last labelling model for growing trees: the *positive* model.

- Each internal node is labelled according to the law $p\delta_{\wedge} + (1-p)\delta_{\vee}$ with $p \in [0, 1]$ independently from the other nodes.
- Each leaf is labelled according to a law μ over $\{x_1, \dots, x_k\}$ 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 $p > \frac{1}{2}$, then $p_{n,k} \longrightarrow \delta_{x_1 \wedge \dots \wedge x_k}$.
- If $p < \frac{1}{2}$, then $p_{n,k} \longrightarrow \delta_{x_1 \vee \dots \vee x_k}$.

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

Proof. For this theorem we only develop the probabilistic approach (very similar to the probabilistic proof of Theorem 1) that can be read in appendix. \square

In fact, Theorem 5 is not complete since we have not studied the case $p = \frac{1}{2}$ 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 $a = (a_1, \dots, a_k) \in \{0, 1\}^k$. The weight of a relatively to the distribution ν is the real number $\omega_{\nu}(a) = \nu(x_1)a_1 + \dots + \nu(x_k)a_k$.*

Definition 7 ([FGG09]). *A Boolean function f is a threshold function if there exists a real number $\theta \geq 0$ such that $\forall (a_1, \dots, a_k) \in \{0, 1\}^k$, $f(a_1, \dots, a_k) = 1 \Leftrightarrow \omega_{\nu}(a) \geq \theta$. 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} : $\omega_{\nu}(a^{(1)}) \leq \omega_{\nu}(a^{(2)}) \leq \dots \leq \omega_{\nu}(a^{(2^k)})$. Then, $p_{n,k} \xrightarrow{n \rightarrow +\infty} \sum_{j=1}^{2^k} (\omega_{\nu}(a^{(j)}) - \omega_{\nu}(a^{(j-1)})) \delta_{T_{\nu, \omega_{\nu}(a^{(j)})}}$ 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 $j \in \llbracket 1, 2^k \rrbracket$, 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 developed in [FGG09] in the case of balanced trees. Let \mathcal{E}_t be a Yule tree, $a = (a_1, \dots, a_k)$ and $b = (b_1, \dots, b_k)$ in $\{0, 1\}^k$ be two

assignments of the k variables, and α, β in $\{0, 1\}$. For all $t \geq 0$, let $\pi_{\alpha\beta}(t) = \mathbb{P}_t(f(a) = \alpha \text{ and } f(b) = \beta)$. Let us compute π_{10} by conditioning on the time when the root's clock rings.

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

This gives $\pi_{10}(t)e^t = \sum_{i=1}^k a_i(1 - b_i)\omega_\nu(x_i) + \int_0^t (\pi_{10}(s)^2 + \pi_{10}(s)\pi_{11}(s) + \pi_{10}(s)\pi_{00}(s)) e^s ds$. By deriving 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:

$$\begin{cases} \pi'_{10} &= -\pi_{10}\pi_{01}; \\ \pi'_{01} &= -\pi_{10}\pi_{01}; \\ \pi'_{11} &= \pi_{10}\pi_{01}; \\ \pi'_{00} &= \pi_{10}\pi_{01}. \end{cases} \quad (9)$$

Thanks to (9), 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 $t \rightarrow +\infty$. In the same way, π_{11} and π_{00} are increasing and thus convergent. Let us denote $l_{\alpha\beta} = \lim_{t \rightarrow \infty} \pi_{\alpha\beta}(t)$.

Since $\pi_{\alpha\beta}$ is monotone and convergent for t tends to $+\infty$, its derivative tends to zero as $t \rightarrow +\infty$. thus, taking the limit in system (9), we get:

$$l_{10}l_{01} = 0. \quad (10)$$

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

$$l_{10} - l_{01} = \pi_{10}(0) - \pi_{01}(0) = \omega_\nu(a) - \omega_\nu(b). \quad (11)$$

Thus, if $\omega_\nu(a) \geq \omega_\nu(b)$, then, thanks to (10) and (11), we get: $l_{01} = 0$. If $\omega_\nu(a) \geq \omega_\nu(b)$, then $\mathbb{P}_t(f(a) = 0 \text{ and } f(b) = 1) \rightarrow 0$ as $t \rightarrow +\infty$. 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 $p_{n,k}(f) \rightarrow 0$ as $n \rightarrow +\infty$. 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 $\mathbb{P}_t(f(a) = 1)$ is a constant for all a . Thus $\mathbb{P}_t(f(a) = 1) = \omega_\nu(a)$ and

$$p_{n,k}(T_{\nu, \omega_\nu(a^{(1)})}) + \dots + p_{n,k}(T_{\nu, \omega_\nu(a^{(j)})}) \rightarrow \omega_\nu(a^{(j)}) \text{ for all } j \in \llbracket 1, 2^k \rrbracket.$$

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

6 Proof and extension of Theorem 3

6.1 Proof of Theorem 3

Proof. The proof has two steps: first, we compute the law of f_n the number of premises that are reduced to a simple leave in a growing tree of size n - we call them *nice premises*. This 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.

This urn can be studied by analytic combinatorics methods [FGP05] and it has been shown by Morcrette [Mor10] that:

$$\mathbb{P}(f_n = q) = \frac{1}{q!} \left(e^{-1} - \sum_{j \geq n+1-q} \frac{(-1)^j}{j!} \right). \quad (12)$$

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

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

where $R_n = \sum_{q=1}^n \frac{(1-c^q)}{q!} \sum_{j \geq n+1-q} \frac{(-1)^j}{j!}$ and $S_n = \sum_{q=n+1}^{\infty} \frac{(1-c^q)}{q!}$. 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 $|\sum_{j \geq n+1-q} \frac{(-1)^j}{j!}| \leq \frac{1}{(n+1-q)!}$ and:

$$|R_n| \leq \sum_{q=1}^{n+1-1} \frac{(1-c^q)}{q!(n+1-q)!} \leq \frac{1}{(n+1)!} \sum_{q=1}^n \binom{n+1}{q} (1-c^q) \leq \frac{(2^{n+1} - (1+c)^{n+1})}{(n+1)!} \xrightarrow{n \rightarrow \infty} 0.$$

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

6.2 An extension of Theorem 3

Simple tautologies have been studied in another labelling model: an implication model where negative literals are allowed [FGGZ10]. We can prove, in the same way 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 trees model:

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

Therefore, in the implication model with positive and negative literals, there are again other tautologies charged by the growing trees 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 (12) 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) which are white or black (positive or negative literals) with probability one half independently from the others into k urns (variables). 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 $\Phi(t, z) := \sum_{r,q} \alpha_{r,q} z^r \frac{t^q}{q!} = (z(e^t - 1)^2 + 2e^t - 1)^k$. 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 (12) gives, after calculation:

$$\mathbb{P}_{n,k}(\overline{TS_2}) = e^{-1} \underbrace{\sum_{q=0}^n \frac{\alpha_{0,q}}{q! (2k)^q}}_{Q_n} - \underbrace{\sum_{q=0}^n \frac{\alpha_{0,q}}{q! (2k)^q} \sum_{j \geq n+1-q} \frac{(-1)^j}{j!}}_{R_n}.$$

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

$$\mathbb{P}_{n,k}(\overline{TS_2}) \xrightarrow{n \rightarrow +\infty} \frac{1}{e} \Phi(\frac{1}{2k}, 0) = \frac{1}{e} (2e^{1/2k} - 1)^k \underset{k \rightarrow +\infty}{\sim} \frac{1}{4k}.$$

□

7 Conclusions and perspectives

We have studied in this paper the behaviour of the growing trees under different labelling systems. This behaviour is very different from the Catalan trees model and the Galton-Watson model [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 enlightened.

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 developed in [FGG09] for balanced trees.

Acknowledgements

We are grateful to Michael Drmota for fruitful discussions 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 functional 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 functions induced by Boolean expressions. In *Analysis of Algorithms*, Wien, Austria, Juillet 2010. DMTCS proceedings.
- [Har64] M. A. Harrison. On the classification of Boolean functions by the general linear and affine groups. *Journal of the Society for Industrial and Applied Mathematics*, 12(2):285–299, 1964.
- [JK97] N.L. Johnson and S. Kotz. *Urn models and their application*. Wiley and Sons, 1997.
- [KZ04] Z. Kostrzycka and M. Zaionc. Statistics of intuitionistic 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. We assume $f_1, \dots, f_n > 0$: therefore, each component of Y is strictly increasing. Let $\epsilon > 0$. We have $\lim_{\|Z\|_\infty \rightarrow \infty} \lim_{x \rightarrow \infty} f(x, Z) = 0$; thus,

$$\exists y_0 \text{ such that } \forall \|Y\| \geq y_0, \lim_{x \rightarrow \infty} f(x, Y) = 0 < \frac{\epsilon}{2}.$$

, thus,

$$\exists y_0 \text{ such that } \forall \|Y\| \geq y_0, \exists x_0(Y) \text{ such that } \forall x \geq x_0(Y), \|f(x, Y)\|_\infty < \epsilon.$$

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:

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

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 $\int_{y_0}^\infty \frac{dy}{G(y)} - \int_{y(x)}^\infty \frac{dy}{G(y)} = c_0 - \int_{y(x)}^\infty \frac{dy}{G(y)} = x - \tilde{x}_0$. Since

$$\int_{y(x)}^\infty \frac{dy}{G(y)} \underset{y(x) \rightarrow \infty}{\sim} \frac{c}{y(x)},$$

we deduce $\frac{c}{y(x)} \underset{y(x) \rightarrow \infty}{\sim} c_0 + \tilde{x}_0 - x$, said differently,

$$y(x) \underset{x \rightarrow x_0}{\sim} \frac{1}{c(x_0 - x)}.$$

□

Proof of Theorem 4

Proof. The cases $p > \frac{1}{2}$ and $p < \frac{1}{2}$ are symmetric and can be treated in the same way. In the proof, we only consider the $p > \frac{1}{2}$ 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 P_t over \mathcal{F}_k for all $t \geq 0$. Let $a = (a_1, \dots, a_k) \in \{0, 1\}^k$ 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 P_t is 1, tends to 0 when t tend to infinity. Therefore, let us study $\pi_a(t) := P_t(f(a) = 1)$.

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

$$\begin{aligned}\pi_a(t) &= \frac{e^{-t}}{2k} \sum_{i=1}^k (\mathbb{1}_{a_i=1} + \mathbb{1}_{a_i=0}) + \int_0^t [p\pi_a(t-s)^2 + (1-p)(2\pi_a(t-s) - \pi_a(t-s)^2)] e^{-s} ds, \\ e^t \pi_a(t) &= \frac{1}{2} + \int_0^t ((2p-1)\pi_a(s)^2 + 2(1-p)\pi_a(s)) e^s ds.\end{aligned}$$

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

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

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

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

□

Proof of Theorem 5

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

$$\pi_a(t) = \mathbb{P}_t(f(a) = 1) = 1 + \frac{1}{\lambda e^{(1-2p)t} - 1} \text{ for all } t \geq 0 \text{ or } \pi_a = 1.$$

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