Balanced And/Or trees and linear threshold functions *

Hervé Fournier[†] Danièle Gardy[†] Antoine Genitrini[†]

Abstract

We consider random balanced Boolean formulas, built on the two connectives *and* and *or*, and a fixed number of variables. The probability distribution induced on Boolean functions is shown to have a limit when letting the depth of these formulas grow to infinity. By investigating how this limiting distribution depends on the two underlying probability distributions, over the connectives and over the Boolean variables, we prove that its support is made of linear threshold functions, and give the speed of convergence towards this limiting distribution.

Keywords: Random And/Or formulas; Balanced formulas; Boolean functions; Boolean linear thresholds

1 Introduction

Consider a random balanced formula built on both connectives *and* and *or* and a fixed number of variables. What is the typical function computed by such a large formula? This is the question addressed in this paper, at the confluence of two lines of research.

The first one concerns (not necessarily balanced) formulas, using these two connectives, and the Boolean functions they define. They have received substantial attention under the guise of *And/Or* trees (a formula is easily represented as a tree). Starting with an early work by Paris et al. [9], there have been several attempts to use *And/Or* trees to define a probability distribution on the set of Boolean functions; see Lefman and Savický [8], later extended by Chauvin et al. [3]. Some papers [14, 5, 7] have tried to give further information on this probability distribution, most often for the constant functions. For this model, the limiting distribution is shown to weight any Boolean function expressible in the system, and seems to be biased towards low complexity functions.

The second line of research is the one initiated by Valiant [13] to obtain small monotone Boolean formulas computing the *majority* function. The so-called *probabilistic amplification* method, based on the construction of balanced formulas using a single Boolean connective in order to increase the probability of some Boolean functions, was further used to obtain small formulas for other functions. After Valiant's result, the first set of Boolean functions to be exhibited was threshold functions by Boppana [1], later extended to any read-once functions by Dubiner and Zwick [4]. Gupta and Mahajan [6] improved the construction of Valiant for *majority*. Another approach was taken by Savický [10], and later by Brodsky and Pippenger [2], who consider the construction mechanisms (growth processes, in the terms of Brodsky) as worth studying *per se*, and then try and classify them according to their defining connective.

Here we are interested in a variant, when we no longer work with a single Boolean connective, but have a random choice of two binary connectives, *and* and *or*. That is, we consider fully balanced trees whose internal nodes are independently labelled with *And/Or* according to a Bernouilli distribution, and leaves independently labelled with a fixed number of variables according to a probability vector μ_0. We show that, letting the depth of trees grow to infinity, this induces a limiting probability distribution π on the set of Boolean functions, and we characterize it. As in the above mentioned model of balanced trees build over a single connective, π is concentrated on a small number of functions. Indeed, we show that the support of π is made of linear thresholds whose defining hyperplanes are normal to μ_0. This is to be compared with the unbalanced *And/Or* tree model.

The paper is organized as follows. We define our model of random *And/Or* expressions and describe the limiting distribution it induces on Boolean functions, in Section 2. A proof of this result is given in Section 3. At last some analysis on the speed of convergence is presented in Section 4.

2 The limiting probability distribution for balanced And/Or trees

Our basic objects are Boolean expressions (or formulas) built on k variables $x_1, ..., x_k$, and on the Boolean

*This work was partially supported by the ANR project *SADA* (Structures Aléatoires Discrètes et Algorithmes).

†PRiSM, CNRS UMR 8144, Université de Versailles Saint-Quentin en Yvelines, 45 av. des États-Unis, 78035 Versailles Cedex, France. Email: {herve.fournier, daniele.gardy, antoine.genitrini}@prism.uvsq.fr.

connectives *and* and *or*. Let us notice here that a Boolean expression can be seen as a tree, with the connectives and the literals labelling respectively the internal nodes and the leaves. We shall say that an expression is *balanced* if all leaves have the same depth.

Let $k > 0$, $p \in [0,1]$ and μ_0 a probability distribution on $H_0 = \{x_1, \ldots, x_k\}$. For all $n \geqslant 1$, let $H_n = \{h_1 \wedge h_2,\ h_1 \vee h_2 \mid h_1, h_2 \in H_{n-1}\}$, and μ_n the probability distribution on H_n defined by $\mu_n(h_1 \wedge h_2) = p\ \mu_{n-1}(h_1)\mu_{n-1}(h_2)$, and $\mu_n(h_1 \vee h_2) = (1-p)\ \mu_{n-1}(h_1)\mu_{n-1}(h_2)$. Hence, H_n is the set of balanced expressions of depth n whose internal nodes are independently labelled by $\wedge$ or $\vee$ according to a Bernouilli distribution of parameter p, and whose leaves are independently labelled by the variables $x_1, \ldots, x_k$ according to the probability distribution μ_0.

Now any Boolean expression defines a Boolean function in a natural way (note however that a function is defined by an infinite number of expressions). We denote by $\mathcal{B}_k$ the set of Boolean functions on the k variables $x_1, ..., x_k$. Our sequence (μ_n) induces a sequence (π_n) of probability distributions on $\mathcal{B}_k$ in the following way: For all $f \in \mathcal{B}_k$,

$$\pi_n(f) = \sum_{\{h \in H_n \ \mid\ h \text{ defines } f\}} \mu_n(h).$$

Our aim is to show the existence and describe the limiting distribution of the sequence (π_n). Notice that in the special case where μ_0 is uniform and $p = 1/2$, $\pi_n(f)$ is the proportion of expressions from H_n that define f.

Before presenting our main result, we first need a couple of definitions. We denote by $\prec$ the usual (strict) partial order on $\{0,1\}^k$; that is, $(a_1, \ldots, a_k) \prec (b_1, \ldots, b_k)$ if $a_i \leqslant b_i$ for all i and $a \neq b$. Given $a, b \in \{0,1\}^k$, the greatest lower bound of $\{a, b\}$ is denoted by $\inf\{a, b\}$. For a given probability distribution μ_0 on $\{x_1, \ldots, x_k\}$, we define the *weight* of a point $a = (a_1, \ldots, a_k) \in \{0,1\}^k$ as

$$\omega(a) = \mu_0(x_1).a_1 + \ldots + \mu_0(x_k).a_k.$$

Notice that $\omega(a)$ is a real number of the interval $[0,1]$. Given $\alpha = (\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k$ and $\theta \in \mathbb{R}$, the *linear threshold function* $T_{\alpha,\theta}$ is the Boolean function on $\{0,1\}^k$ defined by

$$T_{\alpha,\theta}(a) = 1 \ \Leftrightarrow\ \alpha_1.a_1 + \ldots + \alpha_k.a_k \geqslant \theta.$$

We are now ready to state:

THEOREM 2.1. *Let $k \geqslant 1$, $p \in [0,1]$, and μ_0 a probability distribution on $\{x_1, \ldots, x_k\}$ such that $\mu_0(x_i) > 0$ for all i. The sequence of distributions (π_n) has a limiting probability distribution, which is described as follows:*

- *First suppose that the probability of $\wedge$ is $p > 1/2$. Then the support of the limiting distribution is reduced to the single function $x_1 \wedge \ldots \wedge x_k$. In the same way, if the probability of $\wedge$ is $p < 1/2$, the support is reduced to the single function $x_1 \vee \ldots \vee x_k$.*

- *If $\wedge$ and $\vee$ are equally likely ($p = 1/2$), then the limiting probability distribution is concentrated on linear threshold functions of the form $T_{\mu_0,U}$, where U is a random variable uniform in $[0,1]$. More precisely, the limiting distribution π can be described as follows: Let $\theta_0 = 0 < \theta_1 < \theta_2 < \cdots < \theta_s = 1$ be the different weights of the all points of $\{0,1\}^k$; for $i \in \{1, \ldots, s\}$, $\pi(T_{\mu_0,\theta_i}) = \theta_i - \theta_{i-1}$.*

The limiting distribution in the case $p = 1/2$ has a natural geometric interpretation. Let $h_0, h_1, \ldots, h_s$ be the different real affine hyperplanes normal to μ_0 and intersecting the hypercube $\{0,1\}^k$, such that h_{i+1} is the hyperplane just above h_i. Notice that each affine hyperplane not containing $(0, \ldots, 0)$ defines a linear threshold in the following way: Points of $\{0,1\}^k$ lying in the open halfspace containing $(0, \ldots, 0)$ evaluate to 0, while the other points of $\{0,1\}^k$ evaluate to 1. The limiting distribution π is concentrated on the linear threshold functions defined by hyperplanes $h_1, \ldots, h_s$; moreover, the probability of the threshold function defined by h_i is proportional to the Euclidean distance $\mathrm{d}(h_i, h_{i-1})$ between h_i and h_{i-1} (i.e. it is equal to $\mathrm{d}(h_i, h_{i-1})/\mathrm{d}(h_0, h_s)$). This is illustrated by Figure 1 in the case of two variables.

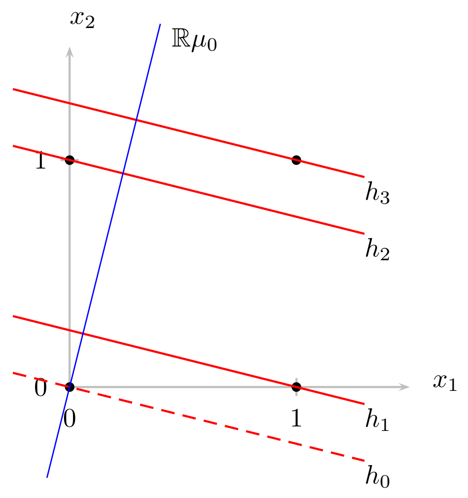

Figure 1: Defining hyperplanes of linear threshold functions in the limiting distribution of (π_n) for the parameters $p = \frac{1}{2}$, $\mu_0(x_1) = \frac{1}{5}$ and $\mu_0(x_2) = \frac{4}{5}$.

The uniform case in an immediate consequence of Theorem 2.1:

Corollary 2.1. *For $p = 1/2$ and μ_0 uniform on $H_0 = \{x_1, \dots, x_k\}$, the sequence (π_n) has a limiting distribution which is uniform on the k threshold functions $x_1 + x_2 + \ldots + x_k \geqslant i$ for $i \in \{1, \dots, k\}$.*

It is an easy extension of Theorem 2.1 to consider symbols computing any Boolean function as labels for the leaves. The sequence of probability induced on Boolean functions by complete *And/Or* trees of increasing heights still admits a limiting distribution, which is obtained by substitution and merging in the limiting distribution described above. As a special case we obtain:

Corollary 2.2. *For $p = 1/2$ and μ_0 uniform on $H_0 = \{x_1, \bar{x}_1, \dots, x_k, \bar{x}_k\}$, the sequence (π_n) has a limiting distribution which is uniform on the two constant functions* True *and* False.

Proof. Consider π'_0 uniform on $H'_0 = \{x_1, y_1, \dots, x_k, y_k\}$. By Corollary 2.1, the sequence (π'_n) induced on Boolean functions admits a limiting distribution which is uniform on the $2k$ threshold functions $x_1 + y_1 + \ldots + x_k + y_k \geqslant i$, $i \in \{1, \dots, 2k\}$. Substituting y_j with $\bar{x}_j$ in these functions, we get the limiting distribution of the sequence (π_n) obtained for the initial distribution μ_0 on H_0: it is uniform over the $2k$ functions $x_1 + \bar{x}_1 + \ldots + x_k + \bar{x}_k \geqslant i$, $i \in \{1, \dots, 2k\}$. However, notice that $x_j + \bar{x}_j$ always evaluates to 1. Hence, the above functions are the constant functions *True* (for $i \in \{1, \dots, k\}$) and *False* (for $i \in \{k + 1, \dots, 2k\}$). □

It is interesting to note that the limiting distribution is concentrated on a small number of functions, at most 2^k; in particular, this number of functions is equal to k (resp. 2) in the uniform case with positive literals only (resp. with positive and negative literals). This is to be compared with the unbalanced tree model, which weights all (monotone) functions.

3 Proof of convergence

In the whole section, we assume that μ_0 is a probability distribution on $H_0 = \{x_1, \dots, x_k\}$ satisfying $\mu_0(x_i) > 0$ for all $i \in \{1, \dots, k\}$.

Remark that if an expression F computes a Boolean function f, substituting $\wedge$ for $\vee$ and $\vee$ for $\wedge$ in F gives a formula F' computing the dual f' of f, defined by $f'(x_1, \dots, x_k) = \bar{f}(\bar{x}_1, \dots, \bar{x}_k)$. Hence, cases p and $1 - p$ are dual and we shall only consider the case $p \geqslant 1/2$ in proofs.

We shall consider in turn the case of a non uniform distribution over the connectives (Proposition 3.1), and the case of a uniform distribution over the connectives (Proposition 3.2). Theorem 2.1 follows from both these propositions.

Lemma 3.1. *Let $p \in [0, 1]$. For $a \in \{0, 1\}^k$, let us define the following sequence:*

$$u_n = \mathbb{P}_{f \sim \pi_n} [f(a) = 1] . \tag{3.1}$$

The sequence (u_n) satisfies:

$$\begin{cases} u_0 = \omega(a) \\ u_{n+1} = (2p - 1)u_n^2 + 2(1 - p)u_n. \end{cases}$$

In particular:

- *For $p > 1/2$ and $a \neq (1, \dots, 1)$, $u_n \to 0$;*
- *For $p = 1/2$, the sequence (u_n) is constant, equal to $\omega(a)$.*

Proof. From the definition of π_n, u_n is the probability that an expression $F \in H_n$, chosen at random according to μ_n, evaluates to 1 at point a. If we write $a = (a_1, \dots, a_k)$, the initial distribution μ_0 yields $u_0 = \sum_{i=1}^{k} a_i . \mu_0(x_i) = \omega(a)$. The recurrence equation on (u_n) is obtained by studying the label of the root of an expression $F \sim \mu_{n+1}$, (i.e. $F \in H_{n+1}$ is chosen at random according to the distribution μ_{n+1}): F evaluates to 1 in a if and only if the root of F is labelled by $\wedge$, and both left and right subexpressions evaluate to 1 in a; or if the root of F is labelled by $\vee$, and the left and right subexpressions do not both evaluate to 0 in a. We get the following equation:

$$u_{n+1} = p\ u_n^2 + (1 - p)\ (1 - (1 - u_n)^2).$$

This gives the recurrence relation stated in the lemma.

Assume now $p > 1/2$ and $a \neq (1, \dots, 1)$. From the hypothesis on μ_0, we know that $u_0 = \omega(a) < 1$. The study of the graph of the real function $x \mapsto (2p - 1)x^2 + 2(1 - p)x$ shows that $u_n \to 0$. The case $p = 1/2$ follows at once. □

We first consider the case of a non-uniform distribution over the connectives.

Proposition 3.1. *If $p > 1/2$, the sequence (π_n) has a limit, concentrated on the function $x_1 \wedge \cdots \wedge x_k$.*

Proof. Notice that the expressions we build, all compute monotone, non constant functions. Thus, every expression built by our system evaluates to 1 at point

$(1,\ldots,1)$. Let $a \in \{0,1\}^k \setminus \{(1,\ldots,1)\}$. By Lemma 3.1, we know that

$$\mathbb{P}_{f\sim\pi_n}[f(a)=1] \to 0.$$

Hence, any function f different from $x_1 \wedge \cdots \wedge x_k$ satisfies $\pi_n(f) \to 0$. Therefore, $\pi_n(x_1 \wedge \cdots \wedge x_k) \to 1$. $\square$

We now consider the case of a uniform distribution over the connectives.

LEMMA 3.2. *Let $p = 1/2$, and $a, b \in \{0,1\}^k$ be two distinct points such that $\omega(a) \leqslant \omega(b)$. Let*

$$v_n = \mathbb{P}_{f\sim\pi_n}[f(a)=1 \text{ and } f(b)=0]. \tag{3.2}$$

The sequence (v_n) satisfies:

$$\begin{cases} v_0 = \omega(a) - \omega(\inf\{a,b\}) \\ v_{n+1} = v_n(1-\omega(b)+\omega(a)-v_n). \end{cases}$$

In particular, $v_n \to 0$. Furthermore, the following holds for all n:

$$\begin{aligned} &\mathbb{P}_{f\sim\pi_n}[f(a)=0 \text{ and } f(b)=1] \\ &\quad - \mathbb{P}_{f\sim\pi_n}[f(a)=1 \text{ and } f(b)=0] \\ &\quad = \omega(b)-\omega(a). \end{aligned}$$

Proof. For $\alpha, \beta \in \{0,1\}$, let:

$$v_n^{(\alpha,\beta)} = \mathbb{P}_{f\sim\pi_n}[f(a)=\alpha \text{ and } f(b)=\beta].$$

Note that $v_n^{(\alpha,\beta)}$ is the probability that an expression $F \in H_n$, chosen at random according to μ_n, evaluates to α at point a and to β at point b. For $a = (a_1,\ldots,a_k)$ and $b = (b_1,\ldots,b_k)$, we have

$$\begin{aligned} v_0^{(1,0)} &= \sum_{i=1}^k a_i(1-b_i)\ \mu_0(x_i) \\ &= \sum_{i=1}^k a_i\ \mu_0(x_i) - \sum_{i=1}^k a_i b_i\ \mu_0(x_i) \\ &= \omega(a) - \omega(\inf\{a,b\}). \end{aligned}$$

Recurrence equations on the four sequences $v_n^{(\alpha,\beta)}$ are obtained by inspection of the connective (either $\wedge$ or $\vee$) labelling the root of a random expression $F \sim \mu_{n+1}$. For $v_{n+1}^{(0,0)}$, we obtain:

$$\begin{aligned} v_{n+1}^{(0,0)} &= \frac{1}{2}(v_n^{(0,0)})^2 \\ &+\frac{1}{2}\left((v_n^{(0,0)})^2 + 2v_n^{(0,0)}v_n^{(0,1)} + 2v_n^{(0,0)}v_n^{(1,0)}\right) \\ &+\frac{1}{2}\left(2v_n^{(0,0)}v_n^{(1,1)} + 2v_n^{(0,1)}v_n^{(1,0)}\right). \end{aligned}$$

We get the three other equations in a similar way. Using the fact that $v_n^{(0,0)} + v_n^{(0,1)} + v_n^{(1,0)} + v_n^{(1,1)} = 1$ for all n, the system can be simplified as follows:

$$\begin{cases} v_{n+1}^{(0,0)} = v_n^{(0,0)} + v_n^{(1,0)}v_n^{(0,1)} \\ v_{n+1}^{(1,1)} = v_n^{(1,1)} + v_n^{(1,0)}v_n^{(0,1)} \\ v_{n+1}^{(0,1)} = v_n^{(0,1)} - v_n^{(1,0)}v_n^{(0,1)} \\ v_{n+1}^{(1,0)} = v_n^{(1,0)} - v_n^{(1,0)}v_n^{(0,1)} \end{cases} \tag{3.3}$$

Of course $v_n^{(\alpha,\beta)} \in [0,1]$ for all n. It follows from Equations 3.3 that the sequences $(v_n^{(0,0)})$ and $(v_n^{(1,1)})$ are non-decreasing; since they are bounded by 1, they converge. In the same way, both sequences $(v_n^{(0,1)})$ and $(v_n^{(1,0)})$ converge since they are non-increasing and bounded by 0. For $\alpha, \beta \in \{0,1\}$, let $\ell^{(\alpha,\beta)} = \lim v_n^{(\alpha,\beta)}$. By letting Equations 3.3 grow to infinity, we obtain: $\ell^{(1,0)}.\ell^{(0,1)} = 0$.

Now, by subtracting the last two Equations (3.3), we obtain that the sequence $(v_n^{(0,1)} - v_n^{(1,0)})$ is constant; it is equal to $v_0^{(0,1)} - v_0^{(1,0)} = (\omega(b) - \omega(\inf\{a,b\})) - (\omega(a) - \omega(\inf\{a,b\})) = \omega(b) - \omega(a)$. It follows that $\ell^{(0,1)} - \ell^{(1,0)} = \omega(b) - \omega(a) \geqslant 0$. Now recall that $\ell^{(1,0)}.\ell^{(0,1)} = 0$; we get $\ell^{(1,0)} = 0$, i.e. $v_n^{(1,0)} \to 0$. $\square$

PROPOSITION 3.2. *For $p = 1/2$, the sequence (π_n) admits a limiting distribution, which is the law of $T_{\theta,U}$ where U is uniform in $[0,1]$.*

Proof. Consider two distinct points $a, b \in \{0,1\}^k$ such that $\omega(a) \leqslant \omega(b)$. By Lemma 3.2, we know that

$$\mathbb{P}_{f\sim\pi_n}[f(a)=1 \text{ and } f(b)=0] \to 0.$$

As a consequence, any function f satisfies $\pi_n(f) \to 0$ if it does not fulfill the following condition:

$$\text{for all } a, b \in \{0,1\}^k,\ \omega(a) \leqslant w(b) \Rightarrow\ f(a) \leqslant f(b).$$

Notice that the functions satisfying the above condition are exactly linear threshold functions of the form $T_{\mu_0,\theta}$ for some $\theta \in \mathbb{R}$.

Let $\theta_0 = 0 < \theta_1 < \theta_2 < \cdots < \theta_s = 1$ be the different weights of the all points of $\{0,1\}^k$. For $0 \leqslant i \leqslant s$, let $a_i \in \{0,1\}^k$ be a point of weight θ_i. Let $j \in \{0,\ldots,s\}$; by Lemma 3.1, we know that

$$\mathbb{P}_{f\sim\pi_n}[f(a_j)=1] = \omega(a_j) = \theta_j$$

for all n. Thus it follows that

$$\pi_n(T_{\mu_0,\theta_0}) + \ldots + \pi_n(T_{\mu_0,\theta_j}) \to \theta_j.$$

By induction on j, for all $j \in \{1,\ldots,s\}$, we have $\pi_n(T_{\mu_0,\theta_j}) \to \theta_j - \theta_{j-1}$. This ends the proof since $\sum_{j=1}^s(\theta_j - \theta_{j-1}) = 1$. $\square$

4 Analysis of the convergence speed

We shall analyse the speed of convergence with respect to the number of iterations, for a fixed number of variables, and fixed values of p and μ_0. Let

$$\|\pi_n - \pi\| = \max_{f \in \mathcal{B}_k} |\pi_n(f) - \pi(f)|.$$

If $\|\pi_n - \pi\| = 2^{-\Theta(n)}$, we shall say the speed of convergence is *linear*. The speed of convergence will be called *logarithmic* if $\|\pi_n - \pi\| = \Theta(1/n)$. The system under study never exhibits a convergence speed faster than linear in non trivial cases.

Again, we assume that μ_0 satisfies $\mu_0(x_i) > 0$ for all $i \in \{1, \ldots, k\}$. Notice that in the case $k = 1$, all expressions compute the same function x_1. Hence, we assume $k > 1$ in the remaining of this section. The case $p \in \{0, 1\}$ corresponds to the use of a single connective; it is easily shown that $\|\pi_n - \pi\| = 2^{-\Theta(2^n)}$ in this case. At last, recall that we shall only deal with $p \geqslant 1/2$ by duality. We shall make use of the following fact, a corollary of the mean value theorem, whose proof can be found in undergraduate textbooks:

Fact 4.1. *Let $f : [a, b] \to [a, b]$ be a real function of differentiability class C^2. Assume there exists $c < 1$ such that $|f'(x)| < c$ for all $x \in [a, b]$. Let $x_0 \in [a, b]$. The sequence (x_n) defined by $x_{n+1} = f(x_n)$ converges towards the single fixed point ℓ of f. Moreover, if $f'(\ell) \neq 0$ and $x_0 \neq \ell$, then there exists $\lambda \neq 0$ such that $x_n - \ell \sim \lambda f'(\ell)^n$.*

We first consider the case of a non uniform distribution over the connectives.

Lemma 4.1. *Let $1/2 < p < 1$, and $a \in \{0, 1\}^k \setminus \{(0, \ldots, 0), (1, \ldots, 1)\}$. The sequence (u_n) defined by (3.1) satisfies $u_n = \Theta\left((2 - 2p)^n\right)$.*

Proof. From the hypothesis on μ_0, we have $0 < \omega(a) < 1$. The result follows from Fact 4.1, using the recurrence equation given in Lemma 3.1. □

Proposition 4.1. *For $1/2 < p < 1$, the convergence speed of (π_n) is linear.*

Proof. For a constant function f, we already know that $\pi_n(f) = 0$ for all n. Let f be a non constant function different from $x_1 \wedge \ldots \wedge x_k$. We know from Theorem 2.1 that $\pi(f) = 0$. Let $a \neq (1, \ldots, 1)$ such that $f(a) = 1$. Of course $0 \leqslant \pi_n(f) \leqslant \mathbb{P}_{f \sim \pi_n}[f(a) = 1]$ since $f(a) = 1$. Lemma 4.1 gives

$$\mathbb{P}_{f \sim \pi_n}[f(a) = 1] = O((2 - 2p)^n).$$

It follows that $|\pi_n(f) - \pi(f)| = 2^{-O(n)}$.

We now deal with the function g computing $x_1 \wedge \ldots \wedge x_k$. Of course $\pi_n(g) = 1 - \sum_{f \neq g} \pi_n(f)$. Since there are only a finite number of functions independent of n, there exists some constant $C > 0$ (independent of n and f) such that $\pi_n(f) \leqslant 2^{-C.n}$ for all $f \neq g$. By using the first part of the proof, we get $1 - \pi_n(g) \leqslant |\mathcal{B}_k| \cdot 2^{-C.n}$. We know from Theorem 2.1 that $\pi(g) = 1$. This gives $|\pi_n(g) - \pi(g)| = 2^{-O(n)}$. Altogether, we have shown that $\|\pi_n - \pi\| = 2^{-O(n)}$.

Let $a \in \{0, 1\}^k \setminus \{(0, \ldots, 0), (1, \ldots, 1)\}$ – it exists since we assumed $k > 1$. From the hypothesis on μ_0, we have $0 < \omega(a) < 1$. Since $g(a) = 0$, we have $\pi_n(g) \leqslant \mathbb{P}_{f \sim \pi_n}[f(a) = 0] = 1 - \mathbb{P}_{f \sim \pi_n}[f(a) = 1]$. Recall now from Theorem 2.1 that $\pi(g) = 1$. Hence, we have obtained

$$\mathbb{P}_{f \sim \pi_n}[f(a) = 1] \leqslant \pi(g) - \pi_n(g).$$

From Lemma 4.1, we know that $\mathbb{P}_{f \sim \pi_n}[f(a) = 1] = 2^{-\Omega(n)}$. Hence we have obtained $|\pi_n(g) - \pi(g)| = 2^{-\Omega(n)}$. The lower bound on $\|\pi_n - \pi\|$ is proved. □

We now consider the case of a uniform distribution over the connectives.

Lemma 4.2. *Let $p = 1/2$, and $a, b \in \{0, 1\}^k$ distinct points such that $\omega(a) \leqslant \omega(b)$. The asymptotic behaviour of (v_n) defined by (3.2) is the following:*

- *If $a \prec b$, then (v_n) is constant, equal to 0;*
- *If $\omega(a) = \omega(b)$, then $v_n \sim 1/n$;*
- *In all other cases, $v_n = \Theta((1 - \omega(b) + \omega(a))^n)$.*

Proof. If $a \prec b$, of course $v_n = 0$ for all n because our system only builds monotone functions.

Suppose now that $\omega(a) = \omega(b)$. From Lemma 3.2, we have $v_0 = \omega(a) - \omega(\inf\{a, b\})$. Because $a \neq b$ and $\omega(a) \leqslant \omega(b)$, we have $a \neq (1, \ldots, 1)$; it follows from the hypothesis on μ_0 that $\omega(a) < 1$. Hence $v_0 < 1$. Since $a \not\prec b$, we have $\inf\{a, b\} \prec a$. The hypothesis on μ_0 yields $\omega(\inf\{a, b\}) < \omega(a)$. It follows that $0 < v_0 < 1$. Lemma 3.2 gives the equation $v_{n+1} = v_n(1 - v_n)$. By induction, $v_n \in]0, 1[$ for all n. Let $w_n = 1/v_n$. We have $w_{n+1} = w_n + 1 + \frac{1}{w_n - 1}$. Thus $w_n \geqslant n$ for all n. Hence, $w_{n+1} - w_n = 1 + \frac{1}{w_n - 1} \leqslant 1 + \frac{1}{n-1}$. It follows that $w_n \leqslant n + \ln(n - 2) + w_2$. Finally we conclude that $w_n \sim n$, i.e. $v_n \sim 1/n$.

We now deal with the last case: assume $a \not\prec b$ and $\omega(a) < \omega(b)$. The initial term satisfies $v_0 = \omega(a) - \omega(\inf\{a, b\}) \in]0, 1 - \omega(b) + \omega(a)[$. From the recurrence equation on (v_n) given by Lemma 3.2 and Fact 4.1, we conclude that $v_n = \Theta((1 - \omega(b) + \omega(a))^n)$. □

PROPOSITION 4.2. *For $p = 1/2$, the convergence speed of (π_n) is linear if all points of $\{0,1\}^k$ have distinct weights; otherwise it is logarithmic.*

Proof. Let f be a Boolean function that does not belong to the support of π. If f is not monotone or constant, then $\pi_n(f) = 0$ for all n; we assume in the following that this is not the case. Then there exist two points $a, b \in \{0,1\}^k$ such that $\omega(a) \leqslant \omega(b)$, $f(a) = 1$ and $f(b) = 0$. Of course

$$0 \leqslant \pi_n(f) \leqslant \mathbb{P}_{g\sim\pi_n}[g(a) = 1 \text{ and } g(b) = 0]$$

since $f(a) = 1$ and $f(b) = 0$. With the asymptotic behaviour of $\mathbb{P}_{g\sim\pi_n}[g(a) = 1 \text{ and } g(b) = 0]$ given in Lemma 4.2, we obtain $\pi_n(f) = O(1/n)$ if $\omega(a) = \omega(b)$, and $\pi_n(f) = O((1-\omega(b)+\omega(a))^n)$ if $\omega(a) < \omega(b)$; i.e. $\pi_n(f) = 2^{-O(n)}$ in this last case.

Now let f be a function belonging to the support of π. Let $a \in \{0,1\}^k$ a point of maximal weight in $f^{-1}(0)$, and $b \in \{0,1\}^k$ a point of minimal weight in $f^{-1}(1)$. From Theorem 2.1, $f = T_{\mu_0,\omega(b)}$ and $\pi(f) = \omega(b) - \omega(a)$. Let $\mathcal{F}^{(0,1)}$ be the set of functions computing 0 for a and 1 for b, and $\mathcal{F}^{(1,0)}$ be the set of functions computing 1 for a and 0 for b. Of course $\pi_n(\mathcal{F}^{(0,1)}) = \pi_n(f) + \pi_n(\mathcal{F}^{(0,1)} \setminus \{f\})$. By Lemma 3.2, it holds for all n that:

$$\pi_n(\mathcal{F}^{(0,1)}) - \pi_n(\mathcal{F}^{(1,0)}) = \omega(b) - \omega(a).$$

Now recall that $\pi(f) = \omega(b) - \omega(a)$. Hence we have obtained:

$$\pi_n(f) - \pi(f) = \pi_n(\mathcal{F}^{(1,0)}) - \pi_n(\mathcal{F}^{(0,1)} \setminus \{f\}).$$

Observe now that no function in $\mathcal{F}^{(1,0)}$ belongs to the support of π, because $\omega(a) < \omega(b)$. In the same way, no function in $\mathcal{F}^{(0,1)} \setminus \{f\}$ is in the support of π (since f is the only linear threshold function of the form $T_{\mu_0,\cdot}$ such that $f(a) = 0$ and $f(b) = 1$). With the bounds already obtained in the first paragraph, we obtain that $|\pi_n(f) - \pi(f)| = 2^{-O(n)}$ if all points have different weights; otherwise we get $|\pi_n(f) - \pi(f)| = O(1/n)$. Altogether, we have shown that $\|\pi_n - \pi\| = 2^{-O(n)}$ if all points have different weights; and $\|\pi_n - \pi\| = O(1/n)$ otherwise.

We shall now prove some lower bound on $\|\pi_n - \pi\|$. First suppose that all points have distinct weights. Since $k > 1$, there exist $a, b \in \{0,1\}^k$, such that $0 < \omega(a) < \omega(b) < 1$, and $\{c \in \{0,1\}^k \mid \omega(a) < \omega(c) < \omega(b)\} = \emptyset$. Let $\mathcal{F}^{(0,1)}$ be the set of functions computing 0 for a and 1 for b, and $\mathcal{F}^{(1,0)}$ the set of functions computing 1 for a and 0 for b. Let $f = T_{\mu_0,\omega(b)}$. Of course we have $\pi_n(\mathcal{F}^{(0,1)}) = \pi_n(f) + \pi_n(\mathcal{F}^{(0,1)} \setminus \{f\})$. In the same way as we already did, we can rewrite this

$$\begin{aligned}\pi_n(\mathcal{F}^{(1,0)}) &= (\pi_n(f) - \pi(f)) + \pi_n(\mathcal{F}^{(0,1)} \setminus \{f\}) \\ &= \sum_{g\in\mathcal{F}^{(0,1)}} (\pi_n(g) - \pi(g)) \\ &\leqslant \sum_{g\in\mathcal{F}^{(0,1)}} |\pi_n(g) - \pi(g)|.\end{aligned}$$

It follows that $\|\pi_n - \pi\| \geqslant \pi_n(\mathcal{F}^{(1,0)})/|\mathcal{B}_k|$ for all n. From the asymptotic behaviour of the sequence $(\pi_n(\mathcal{F}^{(1,0)}))$ given in Lemma 3.2 in the case where all points have distinct weights, we get $\|\pi_n - \pi\| = 2^{-\Omega(n)}$.

Suppose now that there exist points $a, b \in \{0,1\}^k$ with different weights. Once again let $\mathcal{F}^{(1,0)}$ be the set of functions computing 1 for a and 0 for b. Of course

$$\pi_n(\mathcal{F}^{(1,0)}) = \sum_{f\in\mathcal{F}^{(1,0)}} \pi_n(f) \leqslant \sum_{f\in\mathcal{F}^{(1,0)}} |\pi_n(f) - \pi(f)|$$

since $\pi(f) = 0$ for all $f \in \mathcal{F}^{(1,0)}$ (all functions in the support of π are constant on any set of points of a given weight). As Lemma 3.2 gives $\pi_n(\mathcal{F}^{(1,0)}) \sim 1/n$ in that case, we have proved that $\|\pi_n - \pi\| = \Omega(1/n)$. □

It would be interesting to study the convergence speed with respect to the number of variables and the initial distribution. However, we do not expect anything fast (e.g., that would allow to recover the existence of polynomial size monotone formulas for the majority function). Note that linear thresholds obtained here are not known to have monotone polynomial size formulas; for more on this see Servedio [11, 12].

References

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

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

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

[4] M. Dubiner and U. Zwick. Amplification by read-once formulas. *SIAM Journal on Computing*, 26(1):15–38, 1997.

[5] D. Gardy and A. Woods. Lower bounds on probabilities for Boolean functions. In C. Martinez, editor, *First International Conference on the Analysis of Algorithms*, pages 139–146, Barcelona (Spain), June 2005. Discrete Mathematics and Theoretical Computer Science, DMTCS Proceedings AD.

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

[7] J. Kozik. Subcritical pattern languages for And/Or trees. In *5th International Conference on Mathematics and Computer Science*, Blaubeuren, Germany, september 2008. Discrete Mathematics and Theoretical Computer Science, DMTCS Proceedings AD.

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

[9] J. B. Paris, A. Vencovská, and G. M. Wilmers. A natural prior probability distribution derived from the propositional calculus. *Annals of Pure and Applied Logic*, 70:243–285, 1994.

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

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

[12] R. A. Servedio. Every linear threshold function has a low-weight approximator. *Computational Complexity*, 16(2):180–209, 2007.

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

[14] A. Woods. On the probability of absolute truth for And/Or formulas. *Bulletin of Symbolic Logic*, 12(3), 2005.