No Shannon effect on probability distributions on Boolean functions induced by random expressions†

Antoine Genitrini[1] and Bernhard Gittenberger[2]

[1]*Laboratoire PRiSM, CNRS UMR 8144 and Université de Versailles Saint-Quentin-en-Yvelines, 45 avenue des États-Unis, 78035 Versailles, France. Email: antoine.genitrini@prism.uvsq.fr.*
[2]*Technische Universität Wien, Wiedner Hauptstrasse 8-10/104, A-1040 Wien, Austria.*
Email: gittenberger@dmg.tuwien.ac.at.

The Shannon effect states that "almost all" Boolean functions have a complexity close to the maximal possible for the uniform probability distribution. In this paper we use some probability distributions on functions, induced by random expressions, and prove that this model does not exhibit the Shannon effect.

Keywords: Boolean functions; Implicational expressions; Complexity; Limiting ratio; Galton-Watson trees; Probability distribution; Analytic combinatorics.

1 Introduction

A Boolean function in k variables is a function $f : \{0,1\}^k \to \{0,1\}$ where 0 and 1 may be interpreted as the truth values `false` or `true`. Each such function can be represented by a Boolean expression built of the k variables and connectives taken from the set $\{\wedge, \vee, \to\}$. The complexity of a Boolean function f is the minimal number of variables needed to build a Boolean expression which represents f. Roughly speaking, the *Shannon effect* is the following phenomenon: If we choose uniformly at random a Boolean function in k variables, then asymptotically almost surely (as k tends to infinity) the function will have a complexity which is exponential in k. The most complex functions in k variables also have exponential complexity. So a random function has, in some sense, almost maximal complexity.

Now, instead of drawing a random function we turn to its representation. Fix a set of connectives, for instance a subset of $\{\wedge, \vee, \to\}$, and an integer k. Then write at random a Boolean expression in k variables using the connectives of the specified set. What is the "typical" function you get? What is its complexity? Do we observe the *Shannon effect* here, *i.e.*, is the complexity of the "typical" function almost the largest possible? What is the mean complexity of the Boolean functions? Note that the distribution obtained in

†This research was partially supported by the A.N.R. project *BOOLE*, the P.H.C. Amadeus project *Random logical trees and related structures*, the ÖAD Amadeus project 03/10, and the FWF grant NFN S9604.

that way is different from the previous one. If, e.g., the chosen set of connectives is a proper subset of $\{\wedge, \vee, \rightarrow\}$, then the system is incomplete, *i.e.*, there are functions which do not have a representation and therefore their probability is zero.

The efforts to define non-uniform probability distributions (induced by random Boolean expressions, or formulae) on the set of Boolean functions, date back to the mid nineties. The starting point is generally the description of expressions as *trees* of a suitable shape and suitably labelled. The first investigations in this direction were carried out by Paris et al. [PVW94] on *And/Or* trees, *i.e.*, expressions built on the two connectives $\wedge$ and $\vee$; the underlying model was that of binary Catalan trees. The study of these trees was further pursued by Lefmann and Savický [LS97] who proved by a pruning argument the existence of a probability distribution induced by random expressions. Moreover, they established important lower and upper bounds for the probability of any Boolean function in terms of its complexity. At the same time, Woods [Woo97] proved independently the existence of a limiting distribution for general formulae. The term limiting probability in this context has to be understood as follows: Consider the ratio of the number of formulae of size n that compute a fixed Boolean function f among all formulae of size n and let the size grow to infinity. It is possible to show that the limit of this ratio exists for a wide variety of logical systems (see Gardy's survey [Gar06]), and that we can thus define a probability distribution on the set of Boolean functions.

The survey paper of Gardy [Gar06] presents an overview of the probability distributions induced by random Boolean expressions on Boolean functions and of the way we can obtain them using the tools of analytic combinatorics: enumeration of formulae/trees by generating functions, the Drmota-Lalley-Woods theorem (see [FS09, p. 482]) for solving an algebraic system of algebraic quadratic equations and asymptotics. Chauvin *et al.* [CFGG04] consider *And/Or* trees, too. They improved the bounds established by Lefmann and Savický and then introduce a second construction of a probability distribution on functions, whose underlying expressions are built by a critical Galton-Watson branching process. More recently, Kozik [Koz08] proved the order of convergence of the probability of a fixed function, when the number of variables tends to infinity, for both probability distributions. Fournier *et al.* [FGGG08] examined the problem in the logical system of implication and established a relationship between the complexity and the probability of a fixed function.

By considering the uniform distribution on Boolean functions, the *Shannon effect* states that asymptotically almost all Boolean functions have a tree complexity with an order of magnitude $2^k/\log k$ which is close to the maximal possible complexity that is of order $\Theta(2^k)$. This classical result was discovered in [RS42, Sha49a]. Further investigations were carried out by Lupanov [Lup62, Lup65]; a proof based on combinatorial counting arguments can be found in Flajolet and Sedgewick's book [FS09].

The main goal of this paper is to disprove the existence of the Shannon effect in probability distribution on functions induced by random expressions. We will examine two probability distributions: For the first one, we will focus on the logical system with one connective (implication). In order to show our result we will prove that a certain class of functions with small complexity has a positive limiting probability. The second is based on Galton-Watson branching processes. We consider here logical systems with an arbitrary set of connectives and are able to characterize the set of functions which attains the total mass in the limiting distribution.

The present paper is organized as follows. In Section 2 we describe the model and state the main result, namely that a subfamily of functions whose complexity is at most quadratic has a strictly positive probability when the number of variables tends to infinity in large implicational trees. The next section develops the tool of expanding trees in a suitable way which will be one of the main ingredients of the proof. Section 5 is dedicated to a second probability distribution, based on decorated Galton-Watson trees. Finally, we present possible perspectives in Section 6.

2 Model and main result

First we start with a rapid description of the expressions under consideration, then we detail the way they induce a probability distribution on Boolean functions. We will state our main result, saying that there is no Shannon effect in the probability distribution on Boolean functions induced by large *implication* trees. Then we will define the crucial tool of expansions of trees and finally, we will prove the main result by computing some limiting ratios.

For the first distribution, we consider expressions built with the single connective of *implication* (denoted by $\to$) and k positive literals $\{x_1, \ldots, x_k\}$, *i.e.*, there is no negation of variables. These expressions can be represented as complete binary and planar trees whose internal nodes are labelled by the single connective and the leaves by some literals. The set of expressions of this logical system is denoted by $\mathcal{E}_k$. Each expression, or tree, is associated to a specific Boolean function. For any expression, we will say that this expression *computes* or *represents* the associated function. The subset of functions that are represented by some expressions of $\mathcal{E}_k$ will be denoted by $\mathcal{B}_k$. The logical system of implication with positive literals is not complete, so $\mathcal{B}_k$ is a subset of all Boolean functions in k variables.

We define the *size* of any expression of $\mathcal{E}_k$ as the number of leaves of its tree representation. Let $f \in \mathcal{B}_k$. The *complexity* of f is the size of the smallest trees computing it. If f actually depends on the variable x, then we say that x is an *essential variable* for f. Otherwise x is called an *inessential* variable for the function f.

Let C_n be the number of complete binary unlabelled trees with n leaves, *i.e.*, $C_n = \frac{1}{n}\binom{2n-2}{n-1}$, the $(n-1)$-th Catalan number. The generating function that enumerates full binary unlabelled trees, where z marks the leaves, is denoted by $F(z)$ and satisfies:

$$F(z) = \frac{1 - \sqrt{1 - 4z}}{2}.$$

Fact 1 *The Catalan numbers satisfy* $C_{n+1} \leqslant \frac{4^n}{\sqrt{\pi} n^{3/2}}$ *for all* $n \in \mathbb{N}$ *and* $C_{n+1} \sim \frac{4^n}{\sqrt{\pi} n^{3/2}}$, *as* $n \to \infty$.

Catalan numbers are well presented in the book of Flajolet and Sedgewick [FS09, p. 6-7].

We define the *limiting ratio* of a subset of expressions $\mathcal{A} \subset \mathcal{E}_k$ as

$$\mu_k(\mathcal{A}) = \lim_{n\to\infty} \frac{\#\{A \in \mathcal{A} \ : \ |A| = n\}}{\#\{A \in \mathcal{E}_k \ : \ |A| = n\}}, \text{ if this limit exists.}$$

For a Boolean function f, we define $\mu_k(f) = \mu_k(\{A \in \mathcal{E}_k \ : [\![A]\!] = f\})$, where $[\![A]\!]$ is the Boolean function represented by the expression A. The results of Drmota [Drm97], Lalley [Lal93] and Woods [Woo97] give us an easy way to prove that the limiting ratio of each Boolean function is defined in the system $\mathcal{E}_k$ (*i.e.*, for all Boolean functions f the limit defining $\mu_k(f)$ exists). These theorems are nicely described in Flajolet and Sedgewick [FS09].

In the following, we will denote the generating function enumerating all trees in this logical system $F_k(z)$. The variable z marks the leaves, so

$$F_k(z) = \frac{1 - \sqrt{1 - 4kz}}{2}.$$

Let us state the main theorem of this section.

Theorem 2 *Let $R = 9\pi k^2/16$. Then the probability of all functions of complexity at most R is larger than or equal to $9/64$, when the number of variables k tends to infinity. Therefore there is no Shannon effect in the logical system built only on implication.*

This theorem proves that a family of functions with small complexity (polynomial in k) has a non-negligible probability. So, the probability distribution induced by *implication* trees cannot exhibit the Shannon effect.

Corollary 3 *In the logical system of implication $\{\rightarrow\}$ and literals $\{x_1, \bar{x}_1, \ldots, x_k, \bar{x}_k\}$ (this system is complete,* i.e.*, all functions are expressible), the probability distribution cannot exhibit the Shannon effect.*

To prove the corollary, we use Theorem 2, with $2k$ positive literals instead of k. Then for all $i \in \{1, \ldots, k\}$ we identify x_{k+i} to $\bar{x}_i$.

3 Expansions in *implication* trees

The goal of this part consists in defining some family of large trees obtained using a smaller tree. One of the property of these trees is that they compute the same function as the smaller one, and consequently, these trees do not represent functions with larger complexity. Therefore we will introduce the concept of (ν, A)-expansions of trees (see below). Certain subclasses of (ν, A)-expansions were used in [FGGG08] for studying the relation between probability and complexity of functions in the implicational system.

Any tree in our logical system can be viewed as a finite sequence of the form $(A_1, \ldots, A_\ell, \lambda)$ with binary trees A_i and a leaf λ. The decomposition is as follows. Start at the root and go to the right-most leaf of the tree. The subtrees sticking out to the left of this path are the trees A_i. Since each node corresponds to an implication $a \rightarrow b$ and this implication is equivalent to $\bar{a} \vee b$, the function computed by the whole tree is just the disjunction of the negations of the functions computed by the subtrees A_i and the label of the leaf λ. In such a decomposition we call λ the *goal* and $A_1, \ldots, A_\ell$ the *premises* of the tree. Recursively, we define the premises and the goal of any subtree T' of T (the goal of T' is its rightmost leaf and its premises are the left subtrees of T' we meet by following its right branch).

For any tree T and any leaf λ of T, we define Δ_λ to be the minimal left subtree of T whose goal is λ (if λ is the goal of T, we define $\Delta_\lambda = T$). For the rest of the paper we will abusively use Δ_x instead of Δ_λ, where x is the label of λ.

Definition 4 *Let T and A be two trees and ν an arbitrary node of T. A (ν, A)-*expansion *of T is defined to be the tree obtained by replacing the subtree B of T which is rooted at ν by the tree $A \rightarrow B$.*

The general scheme of a (ν, A)-*expansion* is depicted in Figure 1. As the tree B is replaced by $A \rightarrow B$, in the right tree the connective $\rightarrow$ is in the place where the root of B was before the expansion.

As a concrete example of the above definition consider the tree T of the expression $x \rightarrow (y \rightarrow z)$. Let ν_1 the (internal) node corresponding to the second implication and ν_2 the node corresponding to the literal y. Furthermore, let A be an arbitrary tree. Then the (ν_1, A)-expansion of T is $x \rightarrow (A \rightarrow (y \rightarrow z))$ and the (ν_2, A)-expansion of T is $x \rightarrow ((A \rightarrow y) \rightarrow z)$ (see Figure 2).

Lemma 5 *Let T be a tree and x one of the labels of its leaves. Furthermore, let A be a tree with a premise of size one, labelled by x. Then for every (internal or external) node η of Δ_x the (η, A)-expansion of T computes the same Boolean function as T.*

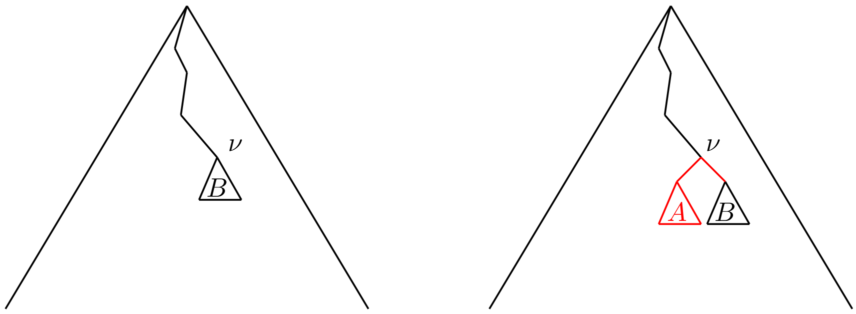

Fig. 1: The right tree is the (ν, A)-expansion of the left one.

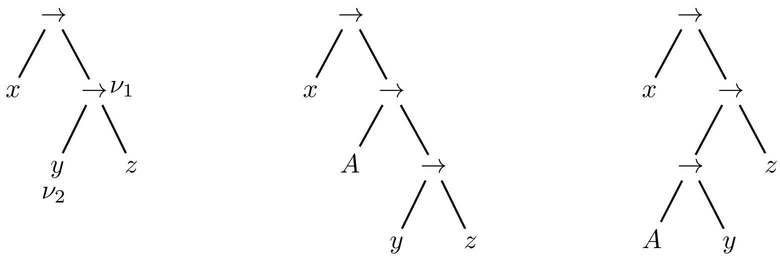

Fig. 2: The left tree is T, the other trees are the (ν_1, A)-expansion and (ν_2, A)-expansion of T.

Such an expansion will be called an *expansion of type "premise x"*.

Proof: Let T be a tree and x the label of one of its leaves. Furthermore, let A be a tree with a premise of size one and labelled by x. Denote by η one of the nodes of Δ_x and B the subtree rooted in η. Set Δ'_x equal to the (η, A)-expansion of Δ_x.

If x is set equal to 1, then Δ'_x computes 1, and so does the tree Δ_x. Otherwise, if $x = 0$, then the tree A, which contains a premise labelled with x computes 1. Consequently, the tree $A \to B$ computes the same function as B and thus Δ'_x and Δ_x compute the same function. This completes the proof. □

Let x be a fixed variable and let us determine $E_k^{prem.\ x}(r)$ the number of expansions of type "premise x" of all labelled trees of size r.

Lemma 6 *Let* $\rho = (k-1)/(2k-1)^2$.

$$E_k^{prem.\ x}(r) = \rho^{-r} \cdot \left(\frac{1}{2(2k-1)} - \frac{1}{2k(k-1)} \sum_{\ell=0}^{r-1} (k\rho)^{\ell+1} \binom{2\ell}{\ell} \right).$$

Proof: Let $T(y,z)$ and $U(y,z)$ be two generating functions. Both enumerate trees with the variable z marking leaves. For the generating function T, the variable y marks every node which belongs to a Δ_x. For U, y is marking nodes such that these nodes belong to at least two distinct (and therefore nested) Δ_x. In fact, by differentiating $T(y,z)$ with respect to y, and then by evaluating y to 1, we get a generating function in the variable z whose coefficient of z^r is the number of expansions (counted with multiplicities) of type "premise x" in all trees of size r.

But we are interested in the number of possible expansions of type "premise x" in all trees of size r (counted *without* multiplicities). This is given by the following value:

$$E_k^{prem.\ x}(r) = [z^r]\left(\frac{\partial T}{\partial y}(y,z)_{|y=1} - \frac{\partial U}{\partial y}(y,z)_{|y=1}\right).$$

To simplify the following equations, we will denote by $T_y(z)$ the function $\frac{\partial T}{\partial y}(y,z)_{|y=1}$. The same kind of notations $T_z(y)$ and $U_y(z)$ will be used. Thus $E_k^{prem.\ x}(r) = T_y(z) - U_y(z)$.

We first establish a functional equation for T (by decomposing it according to its right branch):

$$T(y,z) = \frac{(k-1)z}{1-T(y,z)} + \frac{yz}{1-T(y,y^2z)}.$$

Both terms are respectively obtained when the goal is different from x, resp. is equal to x. In the second term, in each subtree obtained by $T(y,y^2z)$, expansions are possible in every node and moreover an expansion is possible in the node which is the father of this subtree. Consequently, if the subtree has size s, then exactly $2s$ expansions are possible relatively to this subtree. After differentiation and evaluation at $y=1$ we obtain:

$$\begin{aligned} T_y(z) &= \frac{z(1-F_k(z)+2zF_k'(z))}{(1-F_k(z))^2} \\ &= \frac{z}{\sqrt{1-4kz}} + \frac{4kz^2}{(1-4kz)(1+\sqrt{1-4kz})}. \end{aligned}$$

In the same way we get:

$$U(y,z) = \frac{(k-1)z}{1-U(y,z)} + \frac{z}{1-T(y,z)}.$$

The variable y marks the nodes that belong to two distinct Δ_x. In fact, if the goal of the whole tree is distinct from x, then we recursively enumerate the premises. Otherwise, if the goal of the tree is x, then we want to enumerate nodes that belong to a second Δ_x in each premise.

$$U_y(z) = \frac{zT_y(z)}{(1-F_k(z))^2-(k-1)z}.$$

So finally,

$$T_y(z) - U_y(z) = \frac{(1-2k)z}{2((2k-1)^2z-k+1)} + \frac{z}{2((2k-1)^2z-k+1)\sqrt{1-4kz}}.$$

The constant $\rho = (1-(2k-1)^{-2})/(4k)$ is the smallest singularity of the function $T_y(z) - U_y(z)$. To obtain the coefficient of z^r in the previous generating function, we use the Cauchy product of both generating functions $g(z) = 1/((2k-1)^2z-k+1)$ and $h(z) = 1/\sqrt{1-4kz}$. Let r be an integer, then

$$[z^r]g(z) = \frac{\rho^{-r}}{1-k}z^r \text{ and } [z^r]h(z) = (r+1)k^rC_{r+1}z^r,$$

where C_{r+1} is the rth Catalan number. Computing the Cauchy product $g(z)h(z)$ we get

$$[z^{r-1}]g(z)h(z) = \frac{-\rho^{-r}}{k(k-1)} \sum_{\ell=0}^{r-1} (k\rho)^{\ell+1} \binom{2\ell}{\ell}.$$

Thus we conclude

$$E_k^{prem.\ x}(r) = \rho^{-r} \cdot \left(\frac{1}{2(2k-1)} - \frac{1}{2k(k-1)} \sum_{\ell=0}^{r-1} (k\rho)^{\ell+1} \binom{2\ell}{\ell} \right).$$

□

For any tree T and any node ν of T, we define Δ_ν^2 to be the minimal left subtree of T which strictly contains Δ_ν, if it exists. As before we will abusively use Δ_x^2 instead of Δ_ν^2, where x is the label of ν, for the rest of the paper.

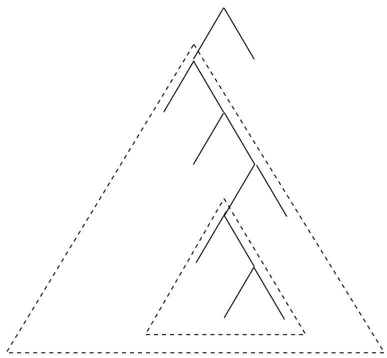

Fig. 3: The left subtrees Δ_ν and Δ_ν^2 associated to a node ν of a tree.

Lemma 7 *Let T be a tree and ν one of its left leaves (a leaf which is a left son) labelled with x. Moreover, let A be a tree whose goal is labelled by x. Then for every (internal or external) node η of Δ_ν^2 which is distinct from ν, the (η, A)-expansion of T computes the same Boolean function as T.*

Such an expansion will be called an *expansion of type "goal x"*.

Proof: Let T be a tree and ν one of its left leaves labelled with x. As ν is a left son, the existence of Δ_ν^2 is obvious. Let A be a tree whose goal is labelled by x. Let η be one of the node of Δ_ν^2, distinct from ν, and B be the tree rooted in η. If x is set equal to 1, then the expanded tree $\Delta_\nu^{2'}$ (the (η, A)-expansion of Δ_ν^2) computes the same function as the tree Δ_ν^2 since A computes 1. Otherwise, if $x = 0$, then Δ_ν^2 computes 1 and consequently the tree $\Delta_\nu^{2'}$ computes 1, too. This completes the proof. □

Let x be a fixed variable and let us determine $E_k^{goal\ x}(r)$ the number of expansions of type "goal x" of all labelled trees of size r.

Lemma 8 *For all $r \in \mathbb{N}$ we have*

$$E_k^{goal\ x}(r) \geqslant E_k^{prem.\ x}(r) - rk^{r-1}C_r.$$

Proof: We want to establish a lower bound of the number of expansions of type "goal x" of trees of size r. Let T be a tree with a leaf ν labelled by x. Using Lemma 7 we know that we can expand the tree by an expansion of type "goal x" in every node of $\Delta_\nu^2 \setminus \{\nu\}$ if ν is a left leaf.

First suppose that ν is a left node. If we consider the mirror image T' of T (it is obtained by exchanging the left and the right sons of T and then going on recursively, *i.e.*, apply the whole procedure to the sons). We denote by ν' the image of ν in T'. The number of nodes of Δ_ν^2 in T is larger than – or possibly equal to (in the case where ν is the first premise of Δ_ν^2) – the number of nodes of Δ'_ν in T'. Consequently,

$$\#\{(\nu, A)\text{-expansions of type "goal } x\text{" in } T\} \geqslant \#\{(\nu', A)\text{-expansions of type "premise } x\text{" in } T'\} - 1$$

Suppose now that ν is a right node. We cannot expand with type "goal x". But if we consider its mirror image, the size of Δ'_ν in T' is 1, so the previous inequality is still valid in this case.

The mapping that changes a tree into its mirror image is bijective so the total number of expansions of type "goal x" in trees of size r is larger than or equal to the difference of the total number of expansions of type "premise x" in trees of size r and the number of nodes labelled by x in trees of size r (let us denote this number by $N(r)$).

$$\begin{aligned} N(r) &= C_r \sum_{\ell=1}^{r} \binom{r}{\ell} (k-1)^{r-\ell} \\ &\leqslant rk^{r-1}C_r. \end{aligned}$$

This completes the proof. □

4 Lower bounds and proof of Theorem 2

Lemma 9 *Let $R = \frac{9\pi}{16}k^2$. For all $r < R$ we have*

$$E_k^{prem.\ x}(r) \geqslant \left(\frac{1}{2(2k-1)} - \frac{1}{8k(k-1)} - \frac{\sqrt{r}}{4\sqrt{\pi}k(k-1)} \right) (4k)^r.$$

Proof: Let $R = \frac{9\pi}{16}k^2$ and $r < R$. Using Fact 1 and the fact that $k\rho < 1/4$, we get

$$\sum_{\ell=0}^{r-1} (k\rho)^{\ell+1} \binom{2\ell}{\ell} \leqslant \frac{1}{4} + \sum_{\ell=1}^{r-1} \frac{\ell+1}{4\sqrt{\pi}\,\ell^{3/2}}$$

The function $(x+1) \cdot x^{-3/2}$ is decreasing, so

$$\sum_{\ell=0}^{r-1} (k\rho)^{\ell+1} \binom{2\ell}{\ell} \leqslant \frac{1}{4} + \frac{1}{4\sqrt{\pi}} \int_1^r \frac{x+1}{x^{3/2}} dx \leqslant \frac{1}{4} + \frac{\sqrt{r}}{2\sqrt{\pi}}.$$

Consequently, the lemma is proved. □

Lemma 10 *Let $R = \frac{9\pi}{16}k^2$. For all $r < R$ we have*

$$E_k^{goal\ x}(r) \geqslant \left(\frac{1}{2(2k-1)} - \frac{1}{8k(k-1)} - \frac{\sqrt{r}}{4\sqrt{\pi}k(k-1)} - \frac{1}{4\sqrt{\pi r}k} \right)(4k)^r.$$

Proof: Using Fact 1,

$$rk^{r-1}C_r \leqslant \frac{(4k)^{r-1}}{\sqrt{\pi r}}.$$

Now, using Lemmas 8 and 9, we get the result. □

We are now ready to prove Theorem 2.

Proof of Theorem 2: Let $R = 9\pi k^2/16$ and $\mathcal{B}_k^R$ be the set of Boolean functions on k variables whose complexity is less than R. For a tree T of size $r < R$, let $\mathcal{E}_T^x$ be the family of trees obtained by one expansion of type "premise x" such that the left subtree A grafted to T satisfies the following conditions: the structure of A is $x \to (T_1 \to T_2)$ or $T_1 \to (x \to T_2)$ and both sizes of T_1 and T_2 are larger than R. Trees of $\mathcal{E}_T^x$ are computing the same function as T, because they belong to expansions of type "premise x" of T. For a tree T of size $r < R$, let $\mathcal{F}_T^x$ be the family of trees obtained by one expansion of type "goal x" such that the left subtree A grafted to T satisfies the following conditions: the structure of A is $T_1 \to T_2$ such that both sizes of T_1 and T_2 are larger than R, the first premise of T_2 has size at least 2 and the goal of T_2 is x. Trees of $\mathcal{F}_T^x$ are computing the same function as T, because they belong to expansions of type "goal x" of T. We remark that both families $\mathcal{E}_T^x$ and $\mathcal{F}_T^x$ are disjoint.

Then,

$$\mu_k(\mathcal{B}_k^R) \geqslant \sum_{r=1}^{\lfloor R \rfloor} \sum_{\text{all variables } x} \sum_{T \in \mathcal{T}_k, |T| = r} \mu_k(\mathcal{E}_T^x) + \mu_k(\mathcal{F}_T^x),$$

where $\lfloor R \rfloor$ denotes the integer part of R.

Let $X(z)$ be the generating functions of trees with structure $x \to (T_1 \to T_2)$ or $T_1 \to (x \to T_2)$ and such that both subtrees' sizes are larger than R. Then we get $X(z) = 2zP(z)^2$, where $P(z) = \sum_{n=\lfloor R \rfloor}^{\infty} k^n C_n z^n$. Let x_n be the coefficient of z^n in X, we have

$$x_n = 2 \sum_{l=\lfloor R \rfloor + 1}^{n-1-\lfloor R \rfloor} k^{n-1} C_l C_{n-1-l}.$$

Let $\epsilon > 0$, then by using Fact 1 there exist sufficiently large k and n such that

$$x_n \geqslant \left(2 \sum_{l=\lfloor R \rfloor + 1}^{n-1-\lfloor R \rfloor} \frac{(4k)^{n-1}}{16\pi(l-1)^{3/2}(n-2-l)^{3/2}} \right) - \epsilon.$$

The function $x \to x^{-3/2}(n-3-x)^{-3/2}$ is decreasing, so

$$x_n \geqslant 2\frac{(4k)^{n-1}}{16\pi} \int_{\lfloor R \rfloor}^{n-1-\lfloor R \rfloor} x^{-3/2}(n-3-x)^{-3/2} dx.$$

Consequently, for a tree T of size r,

$$\mu_k(\mathcal{E}_T^x) = \lim_{n\to\infty} \frac{x_{n-r}}{k^n C_n} \geqslant \frac{2}{\sqrt{\pi\lfloor R\rfloor}(4k)^{r+1}}.$$

In the same way we prove

$$\mu_k(\mathcal{F}_T^x) \geqslant \frac{4}{\sqrt{\pi\lfloor R\rfloor}(4k)^{r+1}}.$$

If we set

$$M_k^1 = \sum_{r=1}^{\lfloor R\rfloor} \sum_{\text{all variables } x} E_k^{prem.\ x}(r) \frac{2}{\sqrt{\pi\lfloor R\rfloor}(4k)^{r+1}}$$

and

$$M_k^2 = \sum_{r=1}^{\lfloor R\rfloor} \sum_{\text{all variables } x} E_k^{goal\ x}(r) \frac{1}{\sqrt{\pi\lfloor R\rfloor}(4k)^{r+1}},$$

then $\mu_k(\mathcal{B}_k^R) \geqslant M_k^1 + M_k^2$. Moreover, note that $E_k^{prem.\ x}(r)$ does not depend on the variable x,

$$\begin{aligned}
M_k^1 &\geqslant \sum_{r=1}^{\lfloor R\rfloor} \left(\frac{1}{2(2k-1)} - \frac{1}{8k(k-1)} - \frac{\sqrt{r}}{4\sqrt{\pi}k(k-1)} \right) \frac{1}{2\sqrt{\pi\lfloor R\rfloor}} \\
&\geqslant \left(\frac{1}{2(2k-1)} - \frac{1}{8k(k-1)} \right) \frac{\sqrt{\lfloor R\rfloor}}{2\sqrt{\pi}} - \frac{1}{8\pi\sqrt{\lfloor R\rfloor}k(k-1)} \int_1^{\lfloor R\rfloor+1} \sqrt{x}dx \\
\lim_{k\to\infty} M_k^1 &\geqslant \frac{3}{64}.
\end{aligned}$$

In the same way we compute a lower bound for M_k^2. By taking the limit for k tending to infinity, we finally obtain

$$\lim_{k\to\infty} \mu_k(\mathcal{B}_k^R) \geqslant \frac{9}{64}.$$

□

5 Decorated Galton-Watson trees

We shall consider the probability distribution on Boolean functions induced by a distribution on trees given by a critical Galton-Watson process, where the internal nodes are labelled uniformly at random and independently and the labels are taken from a set $\mathcal{C}$ containing c binary connectives. The external nodes are labelled uniformly at random and independently with labels taken from the set $\{x_1, \dots, x_k\}$. We shall call such trees decorated Galton-Watson trees.

In this model, the probabilities that a node has zero or two sons are equal to $1/2$. We consider the size of a tree to be its number of leaves. It is known that a tree is almost surely finite in this model (see book [AN72] to get such results). We denote the set of all expressions built with the set of connectives $\mathcal{C}$ and the k variables by $\mathcal{E}_k$.

This probability distribution has been introduced by Chauvin *et al.* in [CFGG04] on *And/Or* trees and can be obviously adapted to our case – here for labelling the internal nodes we choose (uniformly at random) among c different connectives instead of two. So for an expression $A \in \mathcal{E}_k$, we get:

$$\pi_k(A) = \mathbb{P}(\text{structure of } A) \cdot \mathbb{P}(\text{labelling of } A) = \frac{1}{2^{2|A|-1}\ c^{|A|-1}\ k^{|A|}},$$

where $|A|$ denotes the size of A. Notice that the probability $\pi_k(\mathcal{A})$ is well defined for any subset of trees $\mathcal{A} \subset \mathcal{E}_k$. We define the probability of a given Boolean function f, on k variables, as

$$\pi_k(f) = \pi_k(\{A \in \mathcal{E}_k \mid [\![A]\!] = f\}) = \sum_{[\![A]\!]=f} \pi_k(A),$$

where $[\![A]\!]$ is the Boolean function represented by the expression A.

Let us denote by $\mathcal{B}_k$ the subset of functions that are represented by some expressions of $\mathcal{E}_k$ (we recall that $\mathcal{B}_k$ is dependent on $\mathcal{C}$).

As in the model of the previous sections we say that if A is an expression representing $f \in \mathcal{B}_k$, then A *computes the function* f. The *complexity* of f is the size of the smallest trees computing it. If f depends on the variable x, then x is called an *essential variable* else an *inessential* variable for f.

A Boolean function $f \in \mathcal{B}_k$ is called a *read-once* function if its complexity is equal to the number of essential variables it depends on. The *minimal trees* of a given function are the trees computing the function whose size equals the complexity of the function. A *read-once* tree is a tree whose leaves are all labelled with distinct variables. Notice that read-once trees are exactly minimal trees computing read-once functions. We will denote by $\mathcal{RT}_k$ the set of read-once trees and by $\mathcal{R}_k$ the set of all read-once functions of $\mathcal{B}_k$.

Theorem 11 *The probability of all read-once functions tends to 1 almost surely, when the number k of variables tends to infinity,* i.e.,

$$\lim_{k\to\infty} \pi_k(\mathcal{R}_k) = 1.$$

Let us give another interpretation of this theorem. When k tends to infinity and you choose a function at random according to the probability distribution induced by decorated Galton-Watson trees, then this function is read-once, almost surely. Obviously, there is no Shannon effect in this model since the complexity of read-once functions is at most k. We prove last Theorem now.

Fact 12 Dominated convergence theorem:
Suppose that $(f_n)_{n\in\mathbb{N}}$ is a sequence of measurable functions, such that f_n tends pointwise to a function f almost everywhere as n tends to infinity. If $|f_n| \leqslant g$ for all n, where g is integrable, then f is integrable and

$$\lim_{n\to\infty} \int f_n d\mu = \int f d\mu.$$

Proof of Theorem 11: To obtain a lower bound of the probability of all read-once functions, when the number k of variables, we compute the following limit: $\lim_{k\to\infty} \pi_k(\mathcal{RT}_k)$. Let us denote $\mathcal{R}_k^\gamma$ the set of read-once functions of complexity γ, and $\mathcal{RT}_k^\gamma$ the corresponding set of minimal trees. Thus we obtain

$$\lim_{k\to\infty} \pi_k(\mathcal{RT}_k) = \lim_{k\to\infty} \sum_{\gamma=1}^{k} \pi_k(\mathcal{RT}_k^\gamma).$$

Obviously $\pi_k(\mathcal{R}_k^\gamma) \geqslant \pi_k(\mathcal{RT}_k^\gamma)$. Let us compute the following probability:

$$\begin{aligned}\pi_k(\mathcal{RT}_k^\gamma) &= \sum_{\substack{A \text{ read-once tree} \\ |A|=\gamma}} \pi_k(A) \\ &= \sum_{\substack{A \text{ read-once tree} \\ |A|=\gamma}} \frac{1}{2^{2\gamma-1}\, c^{\gamma-1}\, k^\gamma} \\ &= \frac{1}{2^{2\gamma-1}\, c^{\gamma-1}\, k^\gamma} C_\gamma \cdot c^{\gamma-1} \cdot k(k-1)\cdots(k-\gamma+1),\end{aligned}$$

where C_γ denotes the $(\gamma-1)$th Catalan number.
So finally

$$\pi_k(\mathcal{RT}_k^\gamma) = \frac{2}{4^\gamma}\frac{k(k-1)\cdots(k-\gamma+1)}{k^\gamma}C_\gamma.$$

Let $\gamma \in \mathbb{N}\setminus\{0\}$, we define $g(\gamma) = \frac{2}{4^\gamma}C_\gamma$. Using Fact 1, we conclude that g is integrable. The functions (π_k) are probability distributions so they are measurable and moreover for all k and γ we have $0 \leqslant \pi_k(\mathcal{RT}_k^\gamma) \leqslant g(\gamma)$. Clearly, $\pi_k(\mathcal{RT}_k^\gamma) = 0$ for all $\gamma > k$. Using Fact 12, we obtain

$$\lim_{k\to\infty}\sum_{\gamma=1}^{k}\pi_k(\mathcal{RT}_k^\gamma) = \sum_{\gamma=1}^{\infty}\frac{2}{4^\gamma}C_\gamma.$$

Using the generating function enumerating C_γ, we get $\lim_{k\to\infty}\sum_{\gamma=1}^{k}\pi_k(\mathcal{RT}_k^\gamma) = 1$.

Finally, since the probability of read-once trees is a lower bound of the probability of the read-once functions, we conclude:

$$\lim_{k\to\infty}\pi_k(\mathcal{R}_k) = 1.$$

□

6 Conclusion and perspectives

In the model based on branching processes the situation is very clear. If you take a random function, then, roughly speaking, its complexity is linear in the number of its variables. In the model based on large Catalan trees we could show that there is a positive fraction of functions with low complexity. Thus a natural question arises: Can we hope that asymptotically almost every function has low (*i.e.*, polynomial) complexity. If so, what is the exponent of the maximal complexity which (speaking in terms of asymptotics) actually shows up? If not, is it possible to identify a class of functions which has asymptotically the total mass and which is at least easy to describe?

Another direction is of course the transfer of the result from the implicational system to other logical systems. Even in the case of And/Or trees the situation is already different. If we try a similar approach then it turns out that the class of functions has limiting ratio zero. However, we conjecture that the model of And/Or trees does not exhibit the Shannon effect. We are currently working on a generalization of the concept of *expansions* in And/Or trees and hope to get a strictly positive limiting ratio for functions of quadratic complexity.

References

[AN72] K. Athreya and P. Ney. *Branching Processes*. Springer, 1972.

[Bop85] R. B. Boppana. Amplification of Probabilistic Boolean Formulas. In *Proceedings of the 26th IEEE Symposium on Foundations of Computer Science*, pages 20–29, 1985.

[CFGG04] 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.

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

[FGGG08] H. Fournier, D. Gardy, A. Genitrini, and B. Gittenberger. Complexity and limiting ratio of Boolean functions over implication. In *33rd International Symposium on Mathematical Foundations of Computer Science (MFCS'08)*, pages 347–362, Torun, Pologne, August 2008.

[FS09] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

[Gar06] D. Gardy. Random Boolean expressions. In *Colloquium on Computational Logic and Applications*, volume AF, pages 1–36. DMTCS Proceedings, 2006.

[Koz08] J. Kozik. Subcritical pattern languages for And/Or trees. In *Fifth Colloquium on Mathematics and Computer Science*, Blaubeuren, Germany, september 2008. DMTCS Proceedings.

[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.

[Lup62] O. B. Lupanov. Complexity of formula realization of functions of logical algebra. *Problemy Kibernetiki, 3:61-80, 1960. English translation: Problems of Cybernetics, Pergamon Press*, 3:782–811, 1962.

[Lup65] O. B. Lupanov. On the realization of functions of logical algebra by formulae of finite classes (formulae of limited depth) in the basis $\wedge$, $\vee$, $\neg$. *Problemy Kibernetiki, 6:5-14, 1961. English translation: Problems of Cybernetics, Pergamon Press*, 6:1–14, 1965.

[PVW94] 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.

[RS42] J. Riordan and C. E. Shannon. The number of two terminal series-parallel networks. *Journal of Mathematics and Physics*, 21:83–93, 1942.

[Sha49a] C. E. Shannon. Communication theory of secrecy systems. *Bell System Tech. J.*, 28:656–715, 1949.

[Sha49b] C. E. Shannon. The synthesis of two-terminal switching circuits. *Bell Systems Technical J.*, 28:59–98, 1949.

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