Probabilities of Boolean Functions Given by Random Implicational Formulas

Antoine Genitrini*
Laboratoire d'Informatique de Paris 6,
CNRS UMR 7606 and
Université Pierre et Marie Curie,
4 place Jussieu,
75252 Paris cedex 05, France.
antoine.genitrini@lip6.fr

Bernhard Gittenberger†
Institute for Discrete Mathematics
and Geometry,
Technische Universität Wien,
Wiedner Hauptstrasse 8-10/104,
A-1040 Wien, Austria.
gittenberger@dmg.tuwien.ac.at

Veronika Kraus†
Institute of Bioinformatics
and Translational Research,
UMIT,
Eduard-Wallnoefer Zentrum 1,
6060 Hall in Tyrol, Austria
veronika.kraus@umit.at

Cécile Mailler*
Laboratoire de Mathématiques de Versailles,
CNRS UMR 8100 and
Université de Versailles Saint-Quentin,
45 avenue des États-Unis,
78035 Versailles, France.
cecile.mailler@math.uvsq.fr

Submitted: Dec 14, 2011; Accepted: May 31, 2012; Published: Jun 6, 2012
Mathematics Subject Classifications: 03B05, 05A15, 05A16

Abstract

We study the asymptotic relation between the probability and the complexity of Boolean functions in the implicational fragment which are generated by large random Boolean expressions involving variables and implication, as the number of variables tends to infinity. In contrast to models studied in the literature so far, we consider two expressions to be equal if they differ only in the order of the premises. A precise asymptotic formula is derived for functions of low complexity. Furthermore, we show that this model does not exhibit the Shannon effect.

*Partially supported by the A.N.R. project *BOOLE*, 09BLAN0011, and by the P.H.C. Amadeus project *Probabilities and tree representations for Boolean functions*.

†Partially supported by the P.H.C. Amadeus project *Probabilities and tree representations for Boolean functions*, and by *FWF (Austrian Science Foundation), National Research Area S9600, grant S9604* and *ÖAD, grant F03/2010*.

1 Introduction

Since almost two decades a rising interest in probability distributions on a set of Boolean functions which are induced by random Boolean expressions can be observed. To our knowledge the first systematic investigations in this direction have been done by Paris et al. [19] on expressions composed of the two connectives *AND* and *OR*. Lefmann and Savický [17] proved the existence of a limiting distribution (as the size of the expressions tends to infinity) and established estimates relating probability and complexity. These bounds were later improved in [1] and the precise asymptotic behaviour for functions of low complexity was determined by Kozik [16]. A generalisation which takes commutativity of the connectives into account and maps this property into the model was presented recently in [10]. The question for the probability of a tautology was pursued for instance in [13, 14, 15, 18] for various logical systems. An overview is presented in Gardy's survey [8].

In this paper we consider Boolean formulas built of positive variables taken from a fixed set $\{x_1, \ldots, x_n\}$ of Boolean variables and the logical connective of implication. The functions which can be represented by such formulas form the so-called implicational fragment. The implicational fragment plays an important role in the theory of programming since there are linear time satisfiability tests for such expressions (see [2]). A thorough study of a closely related model was carried out in [6]: There expressions of the form $A_1 \to (A_2 \to (A_3 \to \ldots (A_p \to \alpha)))$ have been represented by binary trees where internal nodes represent implication and external nodes (leaves) carry labels which correspond to the variables. Though technically more convenient, a drawback of using plane binary trees (i.e., binary trees embedded in the plane) is that two expressions which differ only w.r.t. the order of the premises are considered different. It is more natural to consider $x \to (y \to z)$ and $y \to (x \to z)$ to be the same expression since both represent the function $(x \wedge y) \to z$. The aim of this paper is to study a model of Boolean expressions which takes into account that the premises of an implication can be permuted. In particular, we present a quantitative relation between the probability of a function and its (formula size) complexity.

Plan of the paper: In the next section we will give a precise description of the model and state our main result. It will turn out that it is *a priori* not guaranteed that the probability measure we introduce actually exists. This issue is discussed in Section 3. In Section 4 we study the structure of tautologies. We will show that asymptotically almost all tautologies are simple, i.e. of a particular shape. Section 5 is then devoted to the asymptotic computation of the probabilities of a general Boolean function. We will succeed by following the approach of [6] and introducing certain expansion techniques which may be applied repeatedly to formulas. After all, we prove that applying one expansion to a minimal formula representing f yields a sufficiently rich class of formulas from which the asymptotic probability can be computed. In the last section we disprove the Shannon effect for our model. Roughly speaking, a model exhibits the Shannon effect if a random function asymptotically almost surely has exponential complexity (w.r.t. to the number of variables). This is, for instance, true for the set of all Boolean functions on n variables equipped with the uniform probability.

2 Model and main result

We start with a precise description of the model:

Definition 1. An **implicational formula** (or **expression**) is either a single variable taken from $\{x_1, \ldots, x_n\}$ or defined recursively as an expression of the form $\{A_1, A_2, \ldots A_k\} \to \alpha$ where $A_1, \ldots, A_k$ are themselves formulas and α is a variable. The order of the premises $A_1, \ldots, A_k$ is not relevant, i.e., all expressions $\{A_{\sigma(1)}, A_{\sigma(2)}, \ldots, A_{\sigma(k)}\} \to \alpha$ where σ is a permutation of $\{1, \ldots, k\}$ are equal. We call α the **goal** of the expression. The **size** of a formula is the number of variables it is composed of, where repetitions are counted.

Remark 2. Let us illustrate this definition by an example: Both formulas $\{(x \to y), x\} \to z$ and $\{x, (x \to y)\} \to z$ are the same object. Its size is 4 and it represents the Boolean function $(x \wedge (x \to y)) \to z$.

Expressions according to the above definition can be easily represented by trees:

Definition 3. A rooted and non-plane tree where the vertices are labelled by variables from $\{x_1, \ldots, x_n\}$ (see Figure 1) is called a **general implication tree**. The set of all general implication trees is denoted by $\mathcal{A}_n$, the set of those with m vertices by $\mathcal{A}_{n,m}$.

Obviously, general implication trees represent expressions. The label of the root is the goal and the subtrees are trees representing the premises. Clearly, the tree must be non-plane, since the order of the subtrees is irrelevant. The size of the corresponding expression equals the number of vertices of the tree.

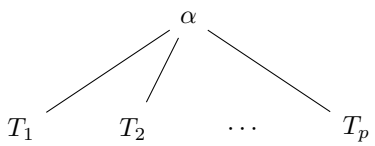


Figure 1: A general implication tree representing the expression $A_1 \to (A_2 \to (A_3 \to \ldots (A_p \to \alpha)))$ where $A_1, \ldots, A_p$ are the expressions represented by $T_1, \ldots, T_p$, respectively.

Our particular interest is the relation between probability and complexity of a Boolean function. Let $\mathcal{F}_n$ denote the implicational fragment of all Boolean functions in the variables $x_1, \ldots, x_n$. We do not require that a function in $\mathcal{F}_n$ actually depends on all the variables. E.g. the constant functions $\top$ (tautologies) and $\bot$ (contradictions) do not depend on any variable.

Definition 4. We call a variable x an **essential variable** for a function $f = f(x, x_1, \ldots, x_{n-1})$ if there exists an assignment of $True$ or $False$ to the variables $x_1, \ldots, x_{n-1}$, which we denote by $\underline{x}_0$, such that $f(True, \underline{x}_0) \neq f(False, \underline{x}_0)$.

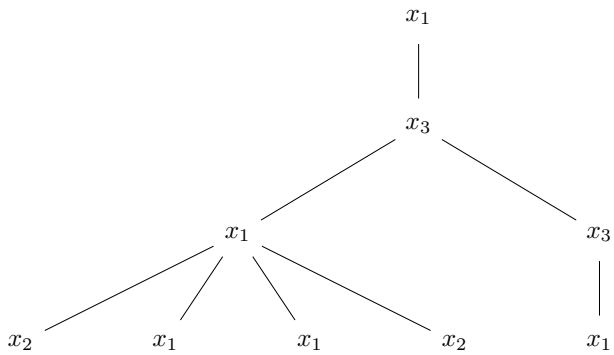


Figure 2: A general implication tree representing the Boolean function $x_3 \to x_1$.

Definition 5. Let $f \in \mathcal{F}_n$. The **complexity** $L(f)$ is defined as the minimal size of expressions representing f.

Let $\mathbb{P}_{m,n}$ be the probability distribution induced by the uniform distribution on the set of all general implication trees of size m. The probability measure $\mathbb{P}_n$ is defined as the limit of the measures $\mathbb{P}_{m,n}$ as m tends to infinity, i.e., $\mathbb{P}_n = \lim_{m\to\infty} \mathbb{P}_{m,n}$. We call $\mathbb{P}_n(f)$ the **probability** of f.

A general implication tree representing f and having $L(f)$ vertices is called a **minimal tree of** f. The set of all minimal trees of f is denoted by $\mathcal{M}_f$.

Remark 6. Note that the probability of the function is defined as the limit of a sequence of probabilities. A priori, we cannot be sure that this limit exists. We will see in the next section that this limit indeed exists.

We aim at a quantitative description of the asymptotic probability of a function, as n tends to infinity, and its relation to the complexity. Precisely, we will show:

Theorem 1. *Let n_0 be a fixed positive integer and $f \in \mathcal{F}_{n_0}$. Then we have*

$$\mathbb{P}_n(f) = \frac{\lambda_f}{n^{L(f)+1}} + \mathcal{O}\left(\frac{1}{n^{L(f)+2}}\right), \text{ as } n \to \infty,$$

where λ_f is a constant independent of n.

Remark 7. Note that we fix the function in advance. Thus the number of essential variables is fixed as well. Nevertheless, we may always regard a function in n_0 variables as an element of $\mathcal{F}_n$ for $n > n_0$. For instance, $x \to y$ is the same function as $\{z \to z, x\} \to y$.

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

In order to show our result we will use the so-called symbolic method (see [5] for a description of this method). A general implication tree can be formally described by a grammar: If $\mathcal{P}$ denotes the set of all general implication trees and $\mathcal{L}$ the set of all variables, then obviously $\mathcal{P} = \mathcal{L} \times \texttt{multiset}(\mathcal{P})$. Let us denote by $P(z) = \sum_{m\geqslant 0} P_m z^m$

the generating function of general implication trees. Then the coefficient $P_m = [z^m]P(z)$ is the number of general implication trees of size m. According to the symbolic method, the grammar can be translated into the functional equation

$$P(z) = nz \exp\left(\sum_{i\geqslant 1} \frac{P(z^i)}{i}\right).$$

Moreover, let $P_f(z) = \sum_{m\geqslant 0} P_{f,m} z^m$ denote the generating function of general implication trees representing the Boolean function f. $P_{f,m}$ is the number of all general implication trees with m vertices which represent f. Similarly as before, we get

$$P_f(z) = z\mathbb{1}_{\{f \text{ lit}\}} + z\sum_{j=1}^{n}\sum_{l=1}^{\infty}\sum_{\{g_1\wedge\ldots\wedge g_l\}\to x_j=f}\prod_{k=1}^{l}\exp\left(\sum_{i=1}^{\infty}\frac{P_{g_k}(z^i)}{i}\right)$$

where $\mathbb{1}_{\{f \text{ lit}\}}$ gives 1 if f is a variable and 0 otherwise, the index j chooses the label of the root of the tree and the index l counts the number of different Boolean functions represented by at least one subtree of the root. Thus the range of the innermost sum is the set of all choices of mutually different functions $g_1, \ldots, g_l$. For every f we get an equation like this, altogether forming a system of functional equations.

We can apply the Drmota-Lalley-Woods theorem (DLW) to this system. We will not go into details here but refer the reader to [5, Chapter VII] or [3, Sec. 2.2.5].

DLW states that, under certain technical conditions, all functions in the system have the same dominant singularity η and admit, locally around η, an expansion of the form

$$P_f(z) = g_f(z) - h_f(z)\sqrt{1-\frac{z}{\eta}} \tag{1}$$

where $g_f(z)$ and $h_f(z)$ are analytic at $z = \eta$. The system above fulfills all assumptions of DLW. We refer to [6, Sec. 3] where an application of DLW in a very similar context is carried out in detail.

By the transfer lemma (see [4, Corollary 2]) we can further extract asymptotic coefficients from an expansion like (1) and obtain

$$[z^m]P_f(z) \sim a_n(f)n^{-3/2}\eta^{-n} \text{as } m \to \infty, \tag{2}$$

where a_n is a positive constant. Note that $P(z) = \sum_f P_f(z)$ and $P(z)$ admits a singular expansion of the same type, it has the same singularity and therefore has coefficients of the same asymptotic behaviour. Since the limiting distribution $\mathbb{P}_n$ is determined by the limits $\lim_{m\to\infty}[z^m]P_f(z)/[z^m]P(z)$, (2) immediately implies the existence of the limiting distribution.

4 Simple tautologies

The proof of our main result relies on the expansion techniques introduced in the next section. These are essentially sets of rules how to attach certain trees to a given tree. For

instance, adding a tautology as a premise never changes the function. Thus we want to count the number of trees obtained by expansions and in particular the trees representing a tautology. Exact counting of tautologies is not easy, but fortunately only tautologies of a simple shape are asymptotically relevant. The three families that will be presented in the next definition have been introduced in the paper [7] in the case where premises cannot commute.

Definition 8. A **simple tautology** is a general implication tree in which exactly one subtree of the root is a leaf labelled by the same label as the root itself (c.f. Figure 3). If this label is x, we say that the simple tautology is **realised by** x.

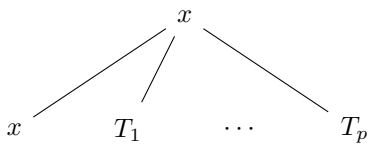


Figure 3: A simple tautology realised by the variable x.

A **simple non-tautology** is a general implication tree such that the label of the root is different from the labels of all its children (c.f. Figure 4).

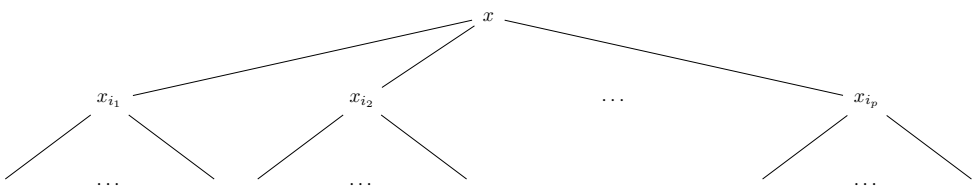


Figure 4: A simple non-tautology: $x \notin \{x_{i_1}, \ldots, x_{i_p}\}$.

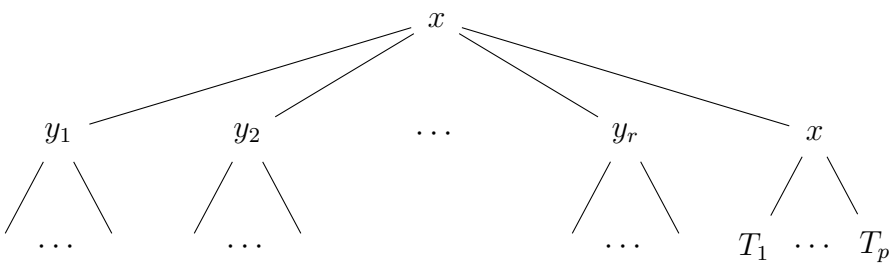
A **less simple non-tautology** is a general implication tree such that (c.f. Figure 5):

- exactly one of the children of the root, say s, has the same label x as the root itself,
- the subtree rooted at s has more than one vertex,
- every child of s has a root labelled by a label different from all the labels appearing on the first level of the tree, and all the grand-children of s have root-labels different from x and from their parent's label.

In order to proceed, we will show that asymptotically almost every expression is a simple tautology, a simple non-tautology or a less simple non-tautology.

Definition 9. Let $\mathcal{B}_n \subseteq \mathcal{A}_n$ and $\mathcal{B}_{n,m} = \{t \in \mathcal{B}_n \mid t \text{ has } m \text{ vertices}\}$. We define the **limiting ratio** of $\mathcal{B}_n$ by

$$\mu_n(\mathcal{B}_n) = \lim_{m\to\infty} \frac{|\mathcal{B}_{n,m}|}{|\mathcal{A}_{n,m}|}.$$



The $(T_i)_{i=1\ldots p}$ are trees rooted by $w_i \notin \{x, y_1, \ldots, y_r\}$ and whose sons are labelled by labels $z_{ij} \notin \{x, w_i\}$:

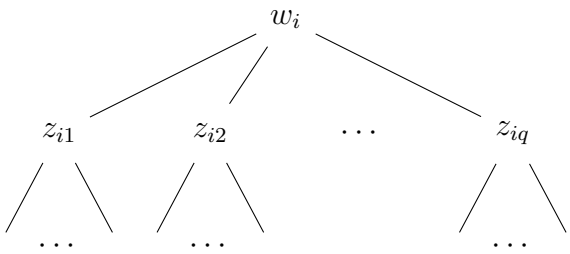


Figure 5: A less simple non-tautology

Remark 10. Note that we can easily construct sets $\mathcal{B}_n$ such that $\mu_n(\mathcal{B}_n)$ does not exist. But the considerations of the previous section imply that for all $f \in \mathcal{F}_n$ the set $\mathcal{A}_{n,f}$ of all general implication trees which represent f has a limiting ratio and even more we have $\mu_n(\mathcal{A}_{n,f}) = \mathbb{P}_n(f)$.

Let $\mathcal{T}$ denote the set of tautologies, $\mathcal{T}_s$ the set of simple tautologies, $\mathcal{N}_s$ and $\mathcal{N}_\ell$ the set of simple and less simple non-tautologies, respectively.

Theorem 2. *Asymptotically almost every tautology is simple, i.e.* $\mu_n(\mathcal{T}_s) \sim \mu_n(\mathcal{T})$ *as* $n \to \infty$.

In order to prove this theorem we will show that $\mu_n(\mathcal{T}_s \cup \mathcal{N}_s \cup \mathcal{N}_\ell) = 1 - \mathcal{O}\left(\frac{1}{n^2}\right)$ which implies $\mu_n(\mathcal{T} \setminus \mathcal{T}_s) = \mathcal{O}\left(\frac{1}{n^2}\right)$. On the other hand, the set of simple tautologies is much larger, as the following proposition shows.

Proposition 11. *The limiting ratio of simple tautologies satisfies* $\mu_n(\mathcal{T}_s) = \frac{1}{en} + \mathcal{O}\left(\frac{1}{n^2}\right)$, *as* $n \to \infty$.

Proof. A simple tautology is constructed by $\mathcal{L} \times \mathcal{Z} \times \texttt{multiset}(\mathcal{P} \setminus \mathcal{Z})$ where $\mathcal{P}$ and $\mathcal{L}$ are the set of general implication trees and the set of variables, respectively, and $\mathcal{Z}$ is the set of a single variable. The $\mathcal{Z}$ corresponds to the distinguished leaf in the first level determined by the label of the root. Hence the generating function of simple tautologies

is

$$S(z) = nz^2 \exp\left(\sum_{i\geqslant 1} \frac{P(z^i) - z^i}{i}\right) = zP(z)\exp\left(-\sum_{i\geqslant 1}\frac{z^i}{i}\right) = z(1-z)P(z).$$

We know that $S(z)$ and $P(z)$ have the same singularity η and singular expansions of the kind (1) around them. It follows that (cf. [10, Lemma 3.4])

$$\mu_n(\mathcal{T}_s) = \lim_{m\to\infty}\frac{S_m}{P_m} = \lim_{z\to\eta}\frac{S'(z)}{P'(z)}.$$

We have

$$\frac{S'(z)}{P'(z)} = \frac{(1-z)P(z)}{P'(z)} - \frac{zP(z)}{P'(z)} + \frac{z(1-z)P'(z)}{P'(z)}.$$

From (1) we can deduce that $P'(z)$ tends to infinity as z tends to η, while $P(\eta)$ is bounded. Therefore, the first two summands become negligible compared to the third one if z tends to η, and

$$\lim_{z\to\eta}\frac{S'(z)}{P'(z)} = \eta(1-\eta).$$

We now have to determine an asymptotic expansion of η, as $n\to\infty$. By an application of the implicit function theorem (cf. [3, Theorem 2.19]), $(\eta, P(\eta))$ is a solution of the system

$$\begin{cases} P = nze^P Q \\ 1 = nze^P Q \end{cases}$$

where $Q = \exp\left(\sum_{i\geqslant 2}\frac{P(z^i)}{i}\right)$ is analytic in η. Therefore, $P(\eta) = 1$ and $\eta = \frac{1}{neQ(\eta)}$. Further, as $P(z) = nz + \mathcal{O}(z^2)$, we know that $Q(z) = 1 + nz^2 + \mathcal{O}(z^3)$, which implies $\eta = \mathcal{O}\left(\frac{1}{n}\right)$ and $Q(\eta) = 1 + \mathcal{O}\left(\frac{1}{n}\right)$. We can thus conclude that $\eta = \frac{1}{en} + \mathcal{O}\left(\frac{1}{n^2}\right)$. By further bootstrapping it is possible to show

$$\eta = \frac{1}{en} - \frac{1}{2e^2n^2} + \mathcal{O}\left(\frac{1}{n^3}\right). \tag{3}$$

We finally get $\lim_{z\to\eta}\frac{S'(z)}{P'(z)} = \eta(1-\eta) = \frac{1}{en} + \mathcal{O}\left(\frac{1}{n^2}\right)$. □

Proposition 12. *The limiting ratio of simple non-tautologies satisfies* $\mu_n(\mathcal{N}_s) = 1 - \frac{2}{n} + \mathcal{O}\left(\frac{1}{n^2}\right)$.

Proof. The set of simple non-tautology satisfies $\mathcal{N}_s = \mathcal{L} \times \texttt{multiset}((\mathcal{L}\setminus\mathcal{Z}) \times \texttt{multiset}(\mathcal{P}))$. Since $(\mathcal{L}\setminus\mathcal{Z})\times\texttt{multiset}(\mathcal{P})$ gives

$$(n-1)z\exp\left(\sum_{i\geqslant 1}\frac{P(z^i)}{i}\right) = \frac{n-1}{n}P(z),$$

the generating function of simple non-tautologies is given by the equation

$$N_s(z) = nz\exp\left(\frac{n-1}{n}\sum_{i\geqslant 1}\frac{P(z^i)}{i}\right) = nz\left(\frac{P(z)}{nz}\right)^{\frac{n-1}{n}} = (nz)^{1/n}P(z)^{\frac{n-1}{n}}.$$

The limiting ratio of simple non-tautologies is $\mu_n(\mathcal{N}_s) = \lim_{m\to\infty} \frac{[z^m]N_s(z)}{P_m} = \lim_{z\to\eta} \frac{N_s'(z)}{P'(z)}$, and

$$N_s'(z) = \frac{1}{n}(nz)^{\frac{1-n}{n}} P(z)^{\frac{n-1}{n}} + \frac{n-1}{n}(nz)^{1/n} P'(z) P(z)^{\frac{n-1-n}{n}}.$$

If we divide by $P'(z)$ and consider the limit when z tends to η, the first term of the sum becomes negligible compared to the second. Eventually, $P(\eta) = 1$ implies

$$\lim_{z\to\eta} \frac{N_s'(z)}{P'(z)} = \frac{n-1}{n}\left(\frac{1}{e}\right)^{1/n} + \mathcal{O}\left(\frac{1}{n^2}\right) = \left(1 - \frac{1}{n}\right)^2 + \mathcal{O}\left(\frac{1}{n^2}\right) = 1 - \frac{2}{n} + \mathcal{O}\left(\frac{1}{n^2}\right),$$

which completes the proof. □

Proposition 13. *We have* $\mu_n(\mathcal{N}_\ell) = \left(2 - \frac{1}{e}\right) \cdot \frac{1}{n} + \mathcal{O}\left(\frac{1}{n^2}\right)$.

Proof. First, we may assume that $y_1, \ldots, y_r$ (c.f. Figure 5) are pairwise different. Any dependency introduces an additional factor $1/n$ and therefore the set of such trees is negligibly small.

Let us fix an integer $r \geqslant 0$ and variables $\{x, y_1, \ldots, y_r\}$. Moreover, let $T(z)$ be the generating function of the T_i, $i = 1 \ldots p$, as in Figure 5. Since for the choice of the root-label $r+1$ possible labels are forbidden, and for the labels in the first generation two possible labels are ineligible, we obtain

$$T(z) = (n-r-1)z \exp\left(\sum_{i\geqslant 0} \frac{n-2}{n} \frac{P(z^i)}{i}\right) = (n-r-1)z \left(\frac{P(z)}{nz}\right)^{\frac{n-2}{n}}. \tag{4}$$

To construct $\mathcal{N}_\ell$ we start with a pair of equally labelled vertices (the root and one of its children). Then we have $\binom{n-1}{r}$ possibilities to choose $y_1, \ldots, y_r$. The resulting subtrees are all different, hence this is no multiset but an ordinary set of trees. Finally, from the multiset comprising the T_i's (c.f. Figure 5) we must exclude the empty set, since x is not a leaf. These considerations yield the formal relation

$$\mathcal{N}_\ell = \mathcal{L} \times \mathcal{Z} \times \bigcup_{r\geqslant 0} \binom{n-1}{r} (\mathcal{Z} \times \texttt{multiset}(\mathcal{P}))^r \times (\texttt{multiset}(\mathcal{T}_r) \setminus \{\emptyset\}),$$

where $\mathcal{T}_r$ is the set of trees with the generating function given in (4). Therefore, the generating function $N_\ell(z)$ of less simple non-tautologies is given by

$$\begin{aligned} N_\ell(z) &= nz^2 \sum_{r\geqslant 0} \binom{n-1}{r} \left(\frac{P(z)}{n}\right)^r \left(\exp\left(\sum_{i\geqslant 1} \frac{T(z^i)}{i}\right) - 1\right) \\ &= nz^2 \sum_{r\geqslant 0} \binom{n-1}{r} \left(\frac{P(z)}{n}\right)^r \left(\exp\left(\sum_{i\geqslant 1} \frac{(n-r-1)z^i}{i} \left(\frac{P(z^i)}{nz^i}\right)^{\frac{n-2}{n}}\right) - 1\right). \end{aligned}$$

As before, the limiting ratio of less simple non-tautologies is $\mu_n(\mathcal{N}_\ell) = \lim_{z\to\eta} \frac{N'_\ell(z)}{P'(z)}$. Let $Q(z) = \exp\left(\sum_{i\geqslant 2} \frac{(n-r-1)z^i}{i}\left(\frac{P(z^i)}{nz^i}\right)^{\frac{n-2}{n}}\right)$. Then

$$\lim_{z\to\eta} \frac{N'_\ell(z)}{P'(z)} \sim n\eta^2 \sum_{r\geqslant 0} \binom{n-1}{r} \frac{P(\eta)^{r-1}}{n^r} \times \left(\left(\frac{n-2}{n}(n-r-1)\eta\left(\frac{P(z)}{n\eta}\right)^{\frac{n-2}{n}} + r\right) Q(\eta) e^{(n-r-1)\eta\left(\frac{P(\eta)}{n\eta}\right)^{\frac{n-2}{n}}} - r\right).$$

We know that $P(\eta) = 1$ and $\eta \sim \frac{1}{en}$. Therefore $Q(\eta) \sim 1$ and hence

$$\lim_{z\to\eta} \frac{N'_\ell(z)}{P'(z)} \sim \frac{1}{ne^2} \sum_{r\geqslant 0} \frac{(n-1)_r}{r!n^r} \left(\left(\frac{n-2}{n}(n-r-1)\frac{1/ne}{(1/e)^{\frac{n-2}{n}}} + r\right) e^{(n-r-1)\frac{1}{en}e^{\frac{n-2}{n}}} - r\right)$$

Truncating the sums at $n^{1/3}$ causes a negligibly small error, moreover we can use the estimate $(n-1)_r/n^r \sim 1$ and obtain

$$\lim_{z\to\eta} \frac{N'_\ell(z)}{P'(z)} \sim \frac{1}{ne^2} \sum_{r=0}^{\lceil \frac{n}{3} \rceil} \frac{1}{r!}\left((r+1)e - r\right) \sim \frac{2}{n} - \frac{1}{en}.$$

□

It is interesting to note that, asymptotically, these three families of trees play the same rôle in the model without commutativity of the premises [7] and in our model.

5 Probability of a general Boolean function

To show Theorem 1 we follow the ideas used in [6] and define different kinds of expansions: tautology expansions, premise expansions and goal expansions.

Definition 14. Given a general implication tree, the tree obtained by adding a new subtree t_e to a node ν of t is called an expansion of t.

- If the expanded tree represents the same function as t, we call the expansion a **valid expansion**.
- If t_e is a simple tautology, the expansion is called a **tautology expansion** of t. For any choice of ν, the expanded tree represents the same function as t and the expansion is valid.
- If t_e has exactly one subtree being a leaf labelled by x, then the expansion is an **x-premise expansion**. If ν has an ancestor labelled by x, or is labelled by x, then the expansion is valid(c.f. Figure 6).
- If the root of t_e is labelled by x, then the expansion is called **x-goal expansion**. If ν is the parent of a leaf labelled by x or if ν has an ancestor which is a brother of a leaf labelled by x, then the expansion is valid (c.f. Figure 7).

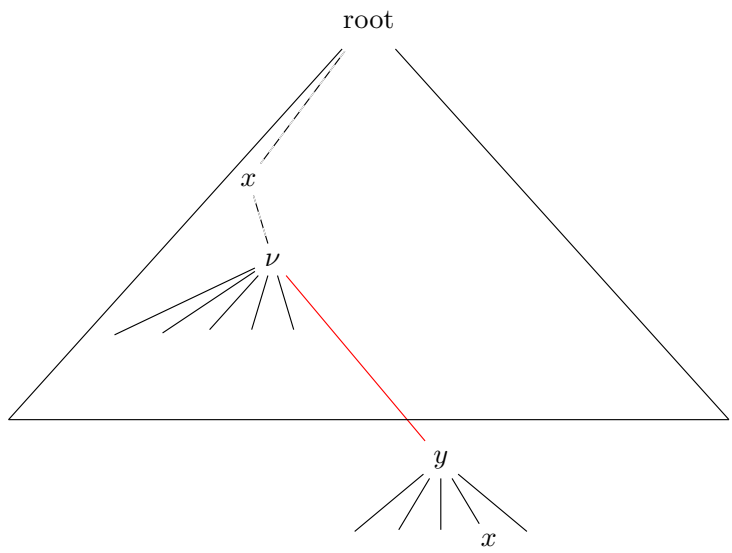


Figure 6: A valid x-premise expansion in ν (y can be any label in $\{x_1, \ldots, x_k\}$).

Fix a Boolean function $f \in \mathcal{F}_{n_0}$. We denote by $E_t(\mathcal{M}_f)$ (resp. $E_p(\mathcal{M}_f)$, resp. $E_g(\mathcal{M}_f)$) the set of trees obtained by a valid tautology (resp. premise, resp. goal) expansion of a minimal tree of f, and by $E(\mathcal{M}_f)$ the union of these three sets.

Lemma 15. *The limiting ratio of the family $E(\mathcal{M}_f)$ verifies, asymptotically when n tends to infinity,*

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

with λ_f being a constant depending only on f.

Proof. Let $\mathcal{E}_x$ denote the set of trees with exactly one child of the root being a leaf labelled by x and by $\mathcal{G}_x$ the set of trees with root labelled by x. Every x-premise (resp. x-goal) expansion of a tree t is realised by attaching a tree from $\mathcal{E}_x$ (resp. $\mathcal{G}_x$) to some node of t.

We already know (cf. Section 4) that $\mu_n(\mathcal{T}) \sim 1/en$, as $n \to \infty$. Using again the symbolic method, we obtain the generating function of the family $\mathcal{E}_x$:

$$E_x(z) = nz^2 \exp \sum_{i \geqslant 0} \frac{P(z^i) - z^i}{i} = z(1-z)P(z) = S(z),$$

and therefore $\mu_n(\mathcal{E}_x) = \frac{1}{en} + \mathcal{O}(\frac{1}{n^2})$. The generating function of the family $\mathcal{G}_x$ is given by $G_x(z) = P(z)/n$ and therefore $\mu_n(\mathcal{G}_x) = 1/n$.

Let t be a tree and define

$$p(t) = \sum_{x \in \{x_1, \ldots, x_n\}} \#\text{vertices of } t \text{ where a valid } x\text{-premise expansion is possible},$$

$$g(t) = \sum_{x \in \{x_1, \ldots, x_n\}} \#\text{vertices of } t \text{ where a valid } x\text{-goal expansion is possible}.$$

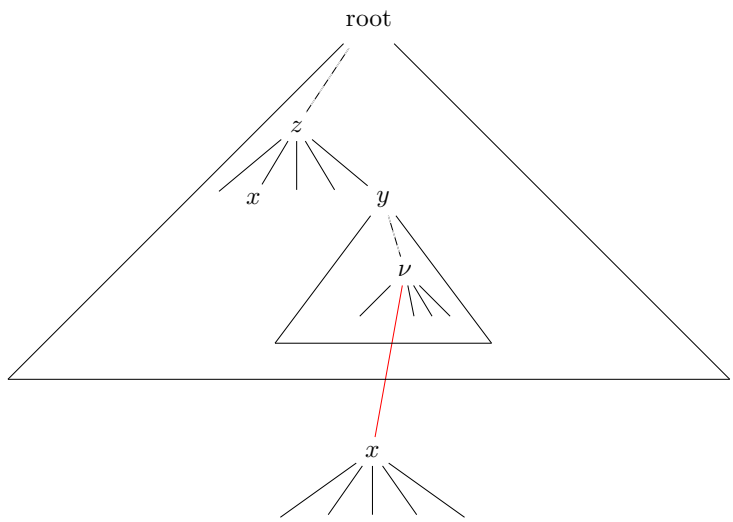


Figure 7: A valid x-goal expansion in ν (y and z can be any label in $\{x_1, \ldots, x_k\}$).

Note that a tautology expansion is always valid. Hence the limiting ratio of $E(\mathcal{M}_f)$ is given by

$$\begin{aligned}\mu_n(E(\mathcal{M}_f)) &= \lim_{m\to\infty} \sum_{t\in\mathcal{M}_f} \left(p(t)\frac{[z^{m-L(f)}]E_x(z)}{[z^m]P(z)} + g(t)\frac{[z^{m-L(f)}]G_x(z)}{[z^m]P(z)} + L(f)\frac{[z^{m-L(f)}]S(z)}{[z^m]P(z)} \right) \\ &= \frac{1}{(en)^{L(f)}} \left(\lambda_p \frac{1}{en} + \lambda_g \frac{1}{n} + L(f)\frac{1}{en} \right),\end{aligned}$$

where $\lambda_p = \sum_{t\in\mathcal{M}_f} p(t)$ and $\lambda_g = \sum_{t\in\mathcal{M}_f} g(t)$. Set

$$\lambda_f = \frac{1}{e^{L(f)}} \left(\frac{\lambda_p}{e} + \lambda_g + \frac{L(f)}{e} \right).$$

Thus,

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

Note that trees obtained by at least two distinct kinds of expansions have been counted several times here. These trees have two constraints though: either two premises with specific labels, or a premise and the goal with fixed labels. As λ_p, λ_g and $L(f)$ are independent of n, such double expansions have a limiting ratio $\mathcal{O}(1/n^{L(f)+2})$. □

Lemma 16. *Further expansions yield an asymptotically negligible contribution to the limiting ratio. Precisely, we have*

$$\mu_n\left(\bigcup_{k\geqslant 2} E^k(\mathcal{M}_f) \setminus E(\mathcal{M}_f) \right) = \mathcal{O}\left(\frac{1}{n^{L(f)+2}} \right), \text{ when } n \to \infty.$$

Proof. First, consider a minimal tree t and expand it once by a tree t', such that the resulting tree t'' is in $E_g(\mathcal{M}_f)$. Then any tree obtained from t'' by doing a second expansion at a node of t' lies in $E(\mathcal{M}_f)$.

Second, consider a minimal tree t and expand it once (by attaching a tree t') such that the resulting tree t'' lies in $E_p(\mathcal{M}_f) \cup E_t(\mathcal{M}_f)$. If we expand a second time, then we obtain a tree not in $E(\mathcal{M}_f)$ if and only if the second expansion is done at a node of t or the expansion is done at the (unique) child of the root of t' which is a leaf labelled by the variable realising the expansion (or the child labelled in the same way than the root in tautology expansions, respectively).

Let us calculate the limiting ratio of the set of premise expansion trees where the expansion is done at the child of the root which realises the expansion: the generating function of this family is

$$N(z) = E_x(z)(E_x(z) + E_g(z) + S(z)).$$

Hence, its limiting ratio is of order $\frac{1}{n^2}$ and the trees obtained by such a nested expansion of a minimal tree have limiting ratio $\mathcal{O}\left(\frac{1}{n^{L(f)+2}}\right)$.

We obtain an analogous result for the set of trees obtained by a second expansion in the specific node related to the first tautology expansion.

The considerations above imply that the only possibly non-negligible family of expanded trees in $\bigcup_{k\geqslant 2} E^k(\mathcal{M}_f) \setminus E(\mathcal{M}_f)$ is the set of trees obtained by k expansions of some minimal tree t, all of which have been done at nodes of t. The limiting ratio of the family of trees obtained by k expansions of a minimal tree, done in k of its initial nodes (counted with possible repetitions), is

$$G_k(z) = \frac{L(f)^k}{k!}(E_t(z) + E_p(z) + E_g(z))^k m_f z^{L(f)},$$

where m_f is the number of minimal trees of f. Its limiting ratio is $\mathcal{O}(\frac{1}{n^{L(f)+k}})$ which is negligible compared to $\frac{1}{n^{L(f)+1}}$ for all $k \geqslant 2$. □

Note that expanding a tree is obviously a reversible procedure. Thus valid expansions have valid reductions as their counterparts. In [6] the authors presented an example of a tree which cannot be reduced though it is not minimal.

Definition 17. We pick a tree representing f and simplify it by removing every subtree which is a tautology, a premise or a goal expansion. If the tree cannot be simplified further and is not minimal, we call the obtained tree an **irreducible** tree of f. The set of such trees is denoted by $\mathcal{I}_f$.

In order to complete the proof of Theorem 1, we have to show that the set of trees obtained by sucessive expansions of all irreducible trees has a negligibly small limiting ratio. The reader will observe that these reductions are not confluent, consequently by expanding irreducible trees we will obtain some trees several times. The following results will state that this double counting is negligible. We will adopt some ideas of the corresponding proof in [6] for binary plane implication trees. In fact, we only have to prove the analogue of [6, Corollary 25]:

Lemma 18. *Fix a subset $\Gamma \subseteq \{x_1, \ldots, x_n\}$. Let $\mathcal{A}_q^p$ be the set of trees having at least p nodes labelled by a variable in Γ at depth less than or equal to q. Moreover, set $E^*(\mathcal{A}_q^p) = \bigcup_{k\geqslant 0} E^k(\mathcal{A}_q^p)$. Then*

$$\mu_n(E^*(\mathcal{A}_q^p) = \mathcal{O}\left(\frac{1}{n^p}\right).$$

Proof. Let us consider the set $\mathcal{B}_q^p$ of trees of height less than or equal to q and with at most pq vertices, at least p of which have labels chosen from Γ. Let $X(\mathcal{B}_q^p)$ denote the set of trees obtained by adding an arbitrary number of general implication trees at nodes of $\mathcal{B}_q^p$. Note that $\mathcal{A}_q^p \subseteq X(\mathcal{B}_q^p)$. If we denote by $\Phi_{q,p}(z)$ the generating function of $\mathcal{B}_q^p$, then

$$[z^\ell]\Phi_{q,p}(z) \leqslant c_\ell \binom{\ell}{p} \gamma^p n^{\ell-p},$$

with γ being the cardinality of Γ and c_ℓ a constant. Adding a multiset of trees at a node z of $\mathcal{B}_{q,p}$ gives an additional $\exp(\sum_{i\geqslant 1} {}^{P(z^i)}/_i) = {}^{P(z)}/_{nz}$. Therefore, the generating function $\Phi_{X_{q,p}}(z)$ of $X(\mathcal{B}_{q,p})$ verifies

$$[z^m]\Phi_{X_{q,p}}(z) \leqslant [z^m] \sum_{\ell\leqslant pq} c_\ell \binom{\ell}{p} \gamma^p n^{l-p} \frac{P(z)^\ell}{n^\ell} = \frac{1}{n^p} \sum_{\ell\leqslant pq} [z^m] C_\ell P(z)^\ell,$$

where C_ℓ is a constant. Thanks to the asymptotic coefficients derived from the singular expansion of type (1) by a transfer lemma (see [4]), we know that $\frac{[z^m]P(z)^\ell}{[z^m]P(z)}$ tends to a constant as $m \to \infty$, since $P(z)$ and $P^\ell(z)$ both have a square-root singularity at η. Therefore, there are constants K_ℓ such that

$$\frac{[z^m]\Phi_{X_{q,p}}(z)}{[z^m]P(z)} \leqslant \frac{1}{n^p} \sum_{\ell\leqslant pq} K_\ell,$$

which implies that the limiting ratio of the set $\mathcal{A}_q^p$ is of order $\mathcal{O}\left(\frac{1}{n^p}\right)$. □

Thanks to this lemma the proofs of [6] can easily be adapted to our new model: The idea is to define a set $\mathcal{N}$ which is the union of certain sets of the form $\mathcal{A}_q^p$, all of them satisfying $p > L(f) + 1$ which implies $\mu_n(\mathcal{N}) = \mathcal{O}\left({}^1/_{n^{L(f)+2}}\right)$. Then the set of irreducible trees $\mathcal{I}_f$ is partitioned according to the tree size and the number of essential and inessential variables of f that occur among the labels. The final step is to show that the sets in this decomposition are either empty or subsets of $\mathcal{N}$. All those arguments are independent of whether the premises are considered permutable or not. Thus the following lemma holds in our model.

Lemma 19. *The limiting ratio of the set $E^*(\mathcal{I}_f)$ verifies*

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

Proof. See [6, Sec. 5]. □

Proof of Theorem 1. The set of implication trees representing a fixed Boolean function f is the union of the following sets: the set $\mathcal{M}_f$ of minimal trees of f, the set $E(\mathcal{M}_f)$ of minimal trees of f expanded once, the set $E^{\geqslant 2}(\mathcal{M}_f)$ of minimal trees of f expanded at least twice, and the set $E^*(\mathcal{I})$ of (iterated) expansions of irreducible trees representing f.

When n tends to infinity, thanks to Lemmas 16 and 19, the set $E(\mathcal{M}_f)$ is the only one to contribute to the asymptotic probability of f. This contribution is given in Lemma 15. □

6 Disproving the existence of the Shannon effect

To show that our model does not exhibit the Shannon effect, we will follow the proof presented in [9] for binary plane implication trees: we consider a family Π of trees with less than $g(n)$ vertices where $g(n)$ is a suitable function of sufficiently fast but subexponential growth in n. If we show that the limiting ratio of all expansions of all trees of Π is non-zero, then there is a set of functions of subexponential complexity (actually bounded by $g(n)$) with positive probability, which contradicts the assertion of the Shannon effect.

Some care is required since we want to avoid double-counting. Thus we restrict the expansions to those where, given an expanded tree, the location where the expansion took place can be identified. Of course, doing so we will not get all trees of $E(\Pi)$ but alone all those representing a function f with $L(f) < g(n)$. Since we are only interested in lower bounds, it is sufficient to get enough trees to guarantee a positive probability.

Given two variables x, y, we denote by $\mathcal{H}_{x,y}$ the family of trees with exactly two children of the root being leaves labelled by x and y, respectively, and with no siblings x and $\tilde{y}$, where $\tilde{y} \neq y$, in the first $g(n)$ levels. Let $H_{x,y}(z)$ be the generating function of this family. Moreover, consider the family $\Pi_{x,y}$ of trees with root label x and in which a leaf labelled by x and a leaf labelled by y cannot be siblings. The generating function of this family is denoted by $\Pi_{x,y}(z)$. We generate trees by expanding a tree $t \in \Pi_{x,y}$ with trees only from $\mathcal{H}_{x,y}$. Hence, an expansion can easily be identified: the forbidden pattern of t appears in every tree of $\mathcal{H}_{x,y}$ among the children of the root. So, the topmost occurence (from the root) of siblings x and y exhibits the location where the expansion took place. The second condition (forbidden siblings x and $\tilde{y}$) enables us to identify y. Finally, note that every such expansion is an x-premise expansion of t.

In order to derive an expression for $H_{x,y}(z)$ let us introduce the following operator: Let $Q(z)$ be a generating function and set

$$\Psi(Q) = nz \exp\left(\sum_{i\geqslant 1} \frac{Q(z^i) - nz^i}{i}\right)((1-z)^{-n} - (n-1)z^2).$$

If we assume that the combinatorial structure $\mathcal{Q}$ corresponding to $Q(z)$ is a tree class related to our model, then the exponential term corresponds to a multiset of non-leaves. The last factor corresponds to a multiset of leaves with a restriction on the multisets of size two. Indeed, we subtract the $n-1$ forbidden multisets $\{x, \tilde{y}\}$. Therefore the combinatorial structure which corresponds to $\Psi(Q)$ has the following property: Either

the number of leaves in the first level is different from two or the set of labels of the two leaves is different from any set $\{x, \tilde{y}\}$ with $\tilde{y} \neq y$. Then we obtain

$$H_{x,y}(z) = nz^3(1-z)\exp\left(\sum_{i\geqslant 1}\frac{\tilde{P}(z^i)}{i}\right)$$

where $\tilde{P}(z) = \Psi^{g(n)}(P)$. Here Ψ^l denotes l consecutive applications of Ψ. Observe that $\tilde{P}(z)$ has the same dominant singularity η as $P(z)$.

To count the number of trees of the family $\Pi_{x,y}$ let us first count the number N_k of trees where the root is unlabelled, it has exactly k children, all those children are leaves, and the set of the labels of these leaves does not contain the set $\{x, y\}$ as a subset. This is exactly the number of k-multisets of labels which contain at most one of the labels x and y. The number of such trees without this restriction is given by $\mathcal{Z}_k(n, n, \ldots, n) = \mathcal{Z}_k(\mathbf{n}) = \binom{n+k-1}{k}$ where $\mathcal{Z}_k$ denotes the cycle index of the symmetric group S_k. Moreover, the number of such trees with at least one leaf labelled by x and one leaf labelled by y is

$$\begin{aligned}&\sum_{l_x,l_y\geqslant 1}^{l_x+l_y\leqslant k}\mathcal{Z}_{k-l_x-l_y}(n-2, n-2, \ldots, n-2)\\ &\qquad = \sum_{l=2}^{k}(l-1)\mathcal{Z}_{k-l}(n-2, n-2, \ldots, n-2) = \sum_{l=2}^{k}(l-1)\mathcal{Z}_{k-l}(\mathbf{n}-2),\end{aligned}$$

which gives $N_k = \mathcal{Z}_k(\mathbf{n}) - \sum_{l=2}^{k}(l-1)\mathcal{Z}_{k-l}(\mathbf{n}-2)$.

Lemma 20. *For $k \geqslant 0$ we have $N_k = \binom{n+k-1}{k} - (k-1)\binom{n+k-4}{k-2} + (n-2)\binom{n+k-4}{k-3}$.*

Proof. The well-known identity (see for example [12, page 174])

$$\sum_{m=0}^{M} m\binom{x+m}{m} = (x+1)\sum_{m=1}^{M}\binom{x+1+m-1}{m-1} = (x+1)\binom{x+1+M}{M-1}$$

yields

$$\begin{aligned}N_k &= \binom{n+k-1}{k} - (k-1)\sum_{l=0}^{k-2}\binom{n-3+l}{l} + \sum_{l=0}^{k-2} l\binom{n-3+l}{l}\\ &= \binom{n+k-1}{k} - (k-1)\binom{n+k-4}{k-2} + (n-2)\binom{n+k-4}{k-3}.\end{aligned}$$

□

Let $D_{x,y}(z)$ be the generating function associated to the class $\mathcal{D}_{x,y}$ of trees in which a leaf labelled by x and a leaf labelled by y cannot be siblings. Clearly, $D_{x,y}(z) = n\Pi_{x,y}(z)$. Decomposing $\mathcal{D}_{x,y}$ according to the number of leaves in the first level we obtain

$$\begin{aligned}D_{x,y}(z) &= nz\exp\left(\sum_{i\geqslant 1}\frac{D_{x,y}(z^i) - nz^i}{i}\right)\sum_{k\geqslant 0} z^k N_k\\ &= nze^{D_{x,y}(z)}Q(z)\end{aligned}$$

where

$$Q(z) = \exp\left(\sum_{i\geqslant 2} \frac{D_{x,y}(z^i) - nz^i}{i}\right) \sum_{k\geqslant 0} z^k N_k$$
$$= (1-z)^n \exp\left(\sum_{i\geqslant 2} \frac{D_{x,y}(z^i)}{i}\right) \sum_{k\geqslant 0} z^k N_k \tag{5}$$

Proposition 21. *The singularity ρ of $D_{x,y}$ verifies $\frac{\rho}{\eta} = 1 + \frac{d}{n^2} + \mathcal{O}(n^{-3})$, as $n \to \infty$, where d is a non-negative constant. Moreover, the function $Q(z)$ given in (5) is analytic in the domain $|z| \leqslant \rho$.*

Proof. The second assertion is an immediate consequence of the first one, since the first one implies $\rho \sim 1/en$.

Note that $(z, D) = (\rho, D_{x,y}(\rho))$ is a solution of the system

$$\begin{cases} D = nze^D Q \\ 1 = nze^D Q \end{cases} \tag{6}$$

where $Q = Q(z)$ is the function given in (5). Moreover, $Q(0) = 1$ and $D_{x,y}(0) = 0$ immediately imply $\rho \sim 1/en$ when $n \to \infty$. For further bootstrapping set $\rho \sim \frac{1}{en}\left(1 + \frac{\alpha}{n}\right)$. First we wish to get a more accurate asymptotic expansion of $Q(\rho)$, as $n \to \infty$. Therefore, let us study the three factors in (5) separately. Observe that

$$(1-z)^n = \left(1 - \frac{1}{en}\left(1 + \frac{\alpha}{n}\right)\right)^n$$
$$= \exp\left(-n \ln \frac{1}{1 - \frac{1}{en}\left(1 + \frac{\alpha}{n}\right)}\right)$$
$$= \exp\left(-\frac{1}{e}\left(1 + \frac{\alpha}{n}\right) - \frac{1}{2e^2 n}\left(1 + \frac{\alpha}{n}\right)^2 + \mathcal{O}\left(\frac{1}{n^2}\right)\right)$$
$$= \exp\left(-\frac{1}{e} - \frac{\alpha}{en} - \frac{1}{2e^2 n} + \mathcal{O}\left(\frac{1}{n^2}\right)\right).$$

Examination of the second factor yields

$$\exp\left(\sum_{i\geqslant 2} \frac{D_{x,y}(z^i)}{i}\right) = \exp\left(\frac{n\rho^2}{2} + \mathcal{O}\left(\frac{1}{n^2}\right)\right) = 1 + \frac{1}{2e^2 n} + \mathcal{O}\left(\frac{1}{n^2}\right).$$

Turning to the last factor, we obtain

$$1 + n\rho + \frac{(n-1)(n+2)}{2}\rho^2 + \sum_{k\geqslant 3}\left[\binom{n+k-1}{k} z^k + \mathcal{O}\left(n^{k-2}\rho^k\right)\right]$$
$$= \sum_{k\geqslant 0} \frac{1}{k!}\left(1 + \frac{\alpha}{n}\right)^k \left(\frac{1}{e}\right)^k + \frac{1}{2}\sum_{k\geqslant 2} \frac{n^{-1}}{e^k (k-2)!}\left(1 + \frac{\alpha}{n}\right)^k + \mathcal{O}\left(\frac{1}{n^2}\right)$$
$$= \exp\left(\frac{1}{e} + \frac{\alpha}{en}\right)\left(1 + \frac{1}{2e^2 n} + \mathcal{O}\left(\frac{1}{n^2}\right)\right).$$

Finally, collecting the contributions of the three factors gives $Q(\rho) = 1 + \frac{1}{2e^2 n} + \mathcal{O}\left(1/n^2\right)$, then the second equation of (6) implies $\rho = \frac{1}{en}\left(1 - \frac{1}{2e^2 n} + \mathcal{O}\left(\frac{1}{n^2}\right)\right)$. Now, (3) implies the assertion, after all. $\square$

Now we are ready to turn to the Shannon effect. We add all the contributions coming from expansions of a tree $t \in \Pi_{x,y}$, say of size r, by attaching a tree $t' \in \mathcal{H}_{x,y}$. Note that there are r positions where t' can be attached and that $\Pi_{x,y}(z) = {}^{D_{x,y}(z)}/_{n}$. Finally, recall that $[z^m]R(z) \sim C\eta^{-m}m^{-3/2}$ for $R(z) = P(z)$ as well as $R(z) = H_{x,y}(z)$ (possibly with different constants C) and $[z^m]D_{x,y}(z) \sim C\rho^{-m}m^{-3/2}$, as $m \to \infty$. Hence we obtain

$$\begin{aligned}
\sum_{f:L(f)\leqslant g(n)} \mathbb{P}_n(f) &\geqslant \lim_{m\to\infty} \sum_{r\leqslant g(n)} \sum_{x\neq y} r \frac{[z^r]D_{x,y}(z)}{n} \frac{[z^{m-r}]H_{x,y}(z)}{[z^m]P(z)} \\
&= \lim_{m\to\infty} \sum_{x\neq y} \sum_{r\leqslant L(f)} c r^{-1/2} \eta^{r+3} \rho^{-r} \\
&= \lim_{m\to\infty} c\eta^3 n(n-1) \sum_{r\leqslant g(n)} r^{-1/2} u^r \qquad (7)
\end{aligned}$$

where $u = \frac{\eta}{\rho} = 1 - \frac{d}{n^2} + \mathcal{O}\left(\frac{1}{n^3}\right)$ and c is a positive constant. Therefore,

$$\begin{aligned}
\sum_{L(f)\leqslant g(n)} \mathbb{P}_n(f) &\geqslant \frac{c}{n} \int_0^{g(n)} \frac{u^r}{\sqrt{r}} dr \\
&= \frac{c}{n}\left(\frac{\sqrt{\pi}}{\sqrt{-\ln u}} - \int_{g(n)}^{\infty} \frac{u^r}{\sqrt{r}} dr\right) \\
&= \frac{c}{n}\left(\frac{\sqrt{\pi}}{\sqrt{-\ln u}} + \mathcal{O}\left(\frac{u^{g(n)}}{-\ln u\sqrt{g(n)}}\right)\right).
\end{aligned}$$

We know that $u = 1 - \mathcal{O}\left(\frac{1}{n^2}\right)$, and thanks to (7), we can conclude that $u = 1 - \Theta\left(\frac{1}{n^2}\right)$, because otherwise the right-hand side of (7) would tend to infinity while the left-hand side is a probability bounded by 1. Thus,

$$\begin{aligned}
\frac{u^{g(n)}}{-\ln u\sqrt{g(n)}} = \frac{n^2}{d\sqrt{g(n)}}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right)\left(1 - \frac{d}{n^2} + \mathcal{O}\left(\frac{1}{n^3}\right)\right)^{g(n)} &= \Omega(n) \text{ if } g(n) = \mathcal{O}(n^2). \\
&= o(n) \text{ if } \frac{n^2}{g(n)} = o(1).
\end{aligned}$$

Therefore, taking $g(n)$ such that $n^2 = o(g(n))$ gives, as $n \to \infty$,

$$\sum_{L(f)\leqslant g(n)} \mathbb{P}_n(f) \geqslant \frac{c\sqrt{\pi}}{n} \frac{1}{\sqrt{-\ln u}} + o(1) = \frac{c\sqrt{\pi}}{d} + o(1)$$

with $\frac{c\sqrt{\pi}}{d} \neq 0$. We thus have shown the following result.

Theorem 3. *If $g(n)$ is a function in n growing faster than n^2, then $\lim_{n\to\infty} \mathbb{P}_n(\{f | L(f) \leqslant g(n)\}) \geqslant \alpha > 0$. Consequently, there is no Shannon effect in our model.*

7 Conclusion

It is interesting to note here that the ideas developed in [7] and in [6] in the implication model without commutativity of the premises can be directly adapted to the present model with commutativity. However, the singularity of the generating function of general trees, which is the fundamental key point of the analysis, cannot be explicitely determined. This makes all computations more intricate in this model, and whereas in the previous model we got exact formulas, we must deal with approximations here. Finally, although both distributions on Boolean functions (with or without commutativity of the premises) are distinct, they are still quite similar: in fact the order of magnitude of the probability of each fixed function is the same for both distributions (cf. Theorem 1). Moreover in both distributions there is no Shannon effect.

We have already noticed that in fact the probability distributions with or without associativity or commutativity in AND/OR trees are almost identical (see [10]). Would it be possible to prove a meta-theorem that would give a relation between the probability distributions induced by logical models taking some properties of the connectives into account, or not?

In this model of general implicational trees the connectives naturally disappear in the tree representation and consequently the size is the number of occurrences of variables, or the total number of nodes in the tree. When we studied commutativity and associativity in AND/OR trees [10] we defined the size to be the number of leaves, which is equal to the number of variables. In the non-associative case, this choice is equivalent to counting every node, because the number of internal nodes is equal to the number of leaves minus one. But in general trees, there is no link between the number of leaves and the number of internal nodes and one could imagine different definitions of the size. Since in [10] we aimed to compare the distribution induced on Boolean functions of this model with the model without the properties of commutativity and associativity (see [16]), the complexity notion and therefore the size notion had to be kept.

But if we take into account the storage of the trees, the natural notion of size in general AND/OR trees is the total number of nodes, i.e. the number of connectives plus the number of occurrences of variables. We are currently working on this different notion size ([11]), which induces a change of the notion of complexity: We expect that the induced distribution will be biased to functions of larger complexity in this model.

References

[1] 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.

[2] William F. Dowling and Jean H. Gallier. Linear-time algorithms for testing the satisfiability of propositional Horn formulae. *J. Logic Programming*, 1(3):267–284, 1984.

[3] M. Drmota. *Random Trees*. Springer, Vienna-New York, 2009.

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

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

[6] H. Fournier, D. Gardy, A. Genitrini, and B. Gittenberger. The fraction of large random trees representing a given Boolean function in implicational logic. *Random Structures and Algorithms*, 40(3):317–349, 2012.

[7] H. Fournier, D. Gardy, A. Genitrini, and M. Zaionc. Classical and intuitionistic logic are asymptotically identical. In Springer-Verlag, editor, *Annual Conference on Computer Science Logic (CSL)*, pages 177–193, Lausanne, Swiss, 2007.

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

[9] A. Genitrini and B. Gittenberger. No Shannon effect on probability distributions on Boolean functions induced by random expressions. In *21st International Meeting on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms*, Vienna, Austria, july 2010. DMTCS Proceedings.

[10] A. Genitrini, B. Gittenberger, V. Kraus, and C. Mailler. Associative and commutative tree representations for Boolean functions. 2011. submitted.

[11] A. Genitrini, B. Gittenberger, V. Kraus, and C. Mailler. A new notion of complexity for random Boolean expressions., 2012. Work in progress.

[12] R.L. Graham, D.E. Knuth, and O. Patashnik. *Concrete Mathematics: a Foundation for Computer Science*, volume 2. Addison-Wesley Reading, MA, 1994.

[13] Z. Kostrzycka. On the density of truth in Grzegorczyk's modal logic. *Bull. Sect. Logic Univ. Łódź*, 33(2):107–120, 2004.

[14] Z. Kostrzycka. On the density of truth in modal logics. In *Mathematics and Computer Science*, volume AG, pages 161–170, Nancy (France), September 2006. DMTCS Proceedings.

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

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

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

[18] G. Matecki. Asymptotic density for equivalence. *Electronic Notes in Theoretical Computer Science*, 140:81–91, 2005.

[19] 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.