Classical and intuitionistic logic are asymptotically identical

Hervé Fournier[*†] Danièle Gardy[*] Antoine Genitrini[*] Marek Zaionc[‡]

September 21, 2007

Abstract

This paper considers logical formulas built on the single binary connector of implication and a finite number of variables. When the number of variables becomes large, we prove the following quantitative results: *asymptotically, all classical tautologies are simple tautologies*. It follows that *asymptotically, all classical tautologies are intuitionistic*.

Keywords: Implicational formulas; Tautologies; Intuitionistic logic; Analytic combinatorics

1 Introduction

We investigate the proportion between the number of formulas of size n that are tautologies against the number of all formulas of size n for propositional formulas built on implication and k variables. Our interest lays in proving the existence and computing the limit of that fraction when n grows to infinity. It can be called the density of truth for the logic with k variables. After isolating the special class of formulas called simple tautologies, of density $1/k+O(1/k^2)$, we exhibit some families of non-tautologies whose cumulated density is $1-1/k-O(1/k^2)$. It follows that the fraction of tautologies, for large k, is very close to the lower bound determined by simple tautologies. A consequence is that classical and intuitionistic logics are close to each other when the number of propositional variables is large.

This work is a part of the research in which the likelihood of truth is estimated for the propositional logic with a restricted number of variables. We refer to Gardy [4] for a survey on probability distribution on Boolean functions induced by random Boolean expressions. For the purely implicational logic of one variable, and at the same time simple type systems, the exact value of the density of truth was computed in the paper of Moczurad, Tyszkiewicz and Zaionc [9]. The classical logic of one variable and the two connectors implication and negation was studied in Zaionc [12]. Over the same language, the exact proportion between intuitionistic and classical logics has been determined in Kostrzycka and Zaionc [6]. Some variants involving formulas with other logical connectives have also been considered. The case of and/or connectors received much attention – see Lefmann and Savický [7], Chauvin, Flajolet, Gardy and Gittenberger [1] and Gardy and Woods [5]. Matecki [8] considered the case of the equivalence connector.

[*]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`.

[†]Corresponding author.

[‡]Theoretical Computer Science, Jagiellonian University, Gronostajowa 3, 30-387 Kraków, Poland. Email: `zaionc@tcs.uj.edu.pl`.

We next give a couple of definitions. Section 2 briefly presents the use of enumeration via generating functions and analytic combinatorics, which constitutes the main tool we shall use. The different classes of formulas we consider are described in Section 3, while Section 4 is devoted to the enumeration of these classes and the computation of their densities.

Definition 1 *Let $\{x_1, x_2, \ldots, x_k\}$ a set of Boolean propositional variables. We define $\mathcal{F}_k$ to be the set of all Boolean expressions (or formulas) over these variables and the implication connector $\rightarrow$. Boolean expressions are defined recursively from Boolean variables and the implication connector by the following grammar: $F := x_1 \mid \ldots \mid x_k \mid (F \rightarrow F)$.*

Obviously the expressions can be represented by binary planar trees, suitably labeled: their internal nodes are labeled by the connector $\rightarrow$ and their leaves by some Boolean variables. By $\|\phi\|$ we mean the *size* of expression ϕ which we define as the total number of occurrences of propositional variables in the expression (or leaves in the tree representation of the expression). Parentheses which are sometimes necessary and the implication sign itself are not included in the size of expression. Formally,

$$\|x_i\| = 1 \text{ and } \|\phi \rightarrow \psi\| = \|\phi\| + \|\psi\| .$$

We denote by $\mathcal{F}_k^n$ the set of expressions of $\mathcal{F}_k$ of size n.

We can now define the *canonical form of an expression.* Let T be an expression. It can be decomposed with respect to its right branch – see Figure 1. Hence it is of the form

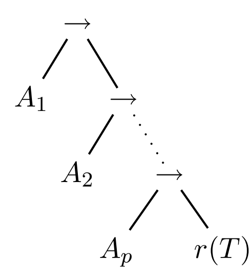

Figure 1: The canonical decomposition of a tree.

$A_1 \rightarrow (A_2 \rightarrow (\ldots \rightarrow (A_p \rightarrow r(T)))\ldots)$; we shall write it

$$T = A_1, \ldots, A_p \rightarrow r(T).$$

The formulas A_i are called the *premises* of T and $r(T)$, the rightmost leaf of the tree, is called the *goal* of T. Of course the expression $T = A_1 \rightarrow (A_2 \rightarrow (\ldots \rightarrow (A_p \rightarrow r(T)))\ldots)$ is logically equivalent with $\overline{A_1} \vee \overline{A_2} \vee \ldots \vee \overline{A_p} \vee r(T)$, where $\overline{A_i}$ stands for negation of A_i.

For a subset $\mathcal{A} \subseteq \mathcal{F}_k$ we define the *density* $\mu(\mathcal{A})$ as:

$$\mu(\mathcal{A}) = \lim_{n\to\infty} \frac{|\{t \in \mathcal{A} : \|t\| = n\}|}{|\{t \in \mathcal{F}_k : \|t\| = n\}|}$$

if the limit exists. The number $\mu(\mathcal{A})$ if it exists is an asymptotic probability of finding a formula from the class $\mathcal{A}$ among all formulas from $\mathcal{F}_k$; it can be interpreted as the asymptotic density of the set $\mathcal{A}$ in the set $\mathcal{F}_k$. It can be seen immediately that the density μ is finitely additive so if $\mathcal{A}$ and $\mathcal{B}$ are disjoint classes of formulas such that $\mu(\mathcal{A})$ and $\mu(\mathcal{B})$ exist then $\mu(\mathcal{A} \cup \mathcal{B})$ also exists and $\mu(\mathcal{A} \cup \mathcal{B}) = \mu(\mathcal{A}) + \mu(\mathcal{B})$.

2 Generating functions

In this paper we investigate the proportion between the number of formulas of size n that are tautologies against the number of all formulas of size n for propositional formulas of the language $\mathcal{F}_k$. Our interest lays in finding the limit of that fraction when n grows to infinity. For this purpose analytic combinatorics has developed an extremely powerful tool, in the form of generating series and generating functions. A nice exposition of the method can be found in Wilf [11], or in Flajolet, Sedgewick [2, 3]; see also Gardy [4, 5.2] for a systematic application of these technics to densities for Boolean functions. As the reader may now expect, while working with propositional logic we will be often concerned with complex analysis, analytic functions and their singularities.

Let $A = (A_0, A_1, A_2, \ldots)$ be a sequence of real numbers. The *ordinary generating series* for A is the formal power series $\sum_{n=0}^{\infty} A_n z^n$. And, of course, formal power series are in one-to-one correspondence to sequences. However, considering z as a complex variable, this series, as known from the theory of analytic functions, converges uniformly to a function $f_A(z)$ in some open disc $\{z \in \mathcal{C} : |z| < R\}$ of maximal diameter, and $R \geqslant 0$ is called its radius of convergence. So with the sequence A we can associate a complex function $f_A(z)$, called the *ordinary generating function* for A, defined in a neighborhood of 0. This correspondence is one-to-one again (unless $R = 0$), since, as it is well known from the theory of analytic functions, the expansion of a complex function $f(z)$, analytic in a neighborhood of z_0, into a power series $\sum_{n=0}^{\infty} A_n(z - z_0)^n$ is unique.

Definition 2 *Let F be a series in powers of z. Then by the symbol $[z^n]\{F\}$ we will mean the coefficient of z^n in the series expansion of F.*

Many questions concerning the asymptotic behavior of A can be efficiently resolved by analyzing the behavior of f_A at the complex circle $|z| = R$. This is the approach we take to determine the asymptotic fraction of tautologies and many other classes of formulas among all formulas of a given size.

Each set of expressions is defined recursively from simpler sets; we build the generating functions enumerating the elements of these sets by size (number of leaves), using univariate functions with the variable z marking the leaves, and obtain a generating function $\phi(z)$ for the set under consideration. We then extract the coefficient $[z^n]\phi(z)$ and obtain the density of the set under study as $\lim_{n\to\infty}[z^n]\phi(z)/[z^n]f(z)$, $f(z)$ being the generating function for the set of all expressions of $\mathcal{F}_k$.

The Catalan number C_n is defined as the number of complete binary trees with n internal nodes and $n+1$ leaves. Basic results about Catalan numbers and its generating function are summarized below.

Proposition 3 *Let $C(z)$ be the generating function enumerating Catalan numbers; it satisfies:*

$$C(z) = 1 + zC(z)^2,$$

and is equal to:

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

Its coefficients are

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

It follows that the number of Boolean expressions of size n over k variables is $k^n C_{n-1}$, since such an expression is obtained by labelling the n leaves with any of the variables $x_1, \ldots, x_k$.

As an example, in the rest of this section we show how we can obtain the generating function $f(z)$ for the set of all the expressions built on k variables and the implication connector, before defining several subsets of expressions in Section 3 and computing their generating functions in Section 4.

Proposition 4 *The generating function enumerating the set $\mathcal{F}_k$ of all Boolean expressions over k variables is*

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

Proof: Using the canonical form of an expression, we know that a tree is a (possibly empty) sequence of trees, followed by a leaf – see Figure 1. The function $f(z)$ thus satisfies

$$f(z) = \frac{kz}{1 - f(z)}, \text{ ie } f(z) = kz + f(z)^2.$$

Solving the equation and choosing between the two possibilities ($f(0) = 0$) gives the solution.
□

The last result gives another way to obtain the number of expressions of size n by extracting the coefficients from the generating function given in proposition 4.

3 Tautologies and non-tautologies

Let us now define several classes of expressions, all of them being special kinds of either tautologies or non-tautologies.

Definition 5 *We define the following subsets of $\mathcal{F}_k$:*

- Cl_k *is the set of all* classical tautologies *i.e. formulas which are true under any valuation.*

- Int_k *is the set of all* intuitionistic tautologies *i.e. formulas for which there are closed lambda terms (constructive proofs) of type identical with the formula.*

- $Pierce_k$ *is the set of all* Pierce expressions *i.e. classical tautologies which are not intuitionistic ones.*

- SN_k *is the set of* simple expressions which are not classical tautologies*, defined as*

  $$T = A_1, \ldots, A_p \to r(T),$$

  such that for all i, $r(A_i) \neq r(T)$.

- G_k *is the set of* simple tautologies *i.e. expressions that can be written as*

  $$T = A_1, \ldots, A_p \to r(T),$$

  where there exists i such that A_i is a variable equal to $r(T)$.

- LN_k *is the set of* less simple expressions that are not classical tautologies, *defined as the set of trees of the form*

$$T = B_1, \ldots, B_{i-1}, C, B_i, \ldots, B_p \rightarrow r(T),$$

such that

$$C = C_1, C_2, \ldots, C_q \rightarrow r(C),$$

where $r(C) = r(T)$, $q \geqslant 1$, *and*

$$C_1 = D_1, D_2, \ldots, D_r \rightarrow r(D),$$

where $r(D) \neq r(T)$, $r \geqslant 0$, *and the following holds: for all* j, $r(B_j) \notin \{r(T), r(D)\}$ *and* $r(D_j) \notin \{r(T), r(D)\}$.

Adding a superscript n *to the sets we have just defined means that we consider only expressions of size exactly* n *(the tree that represents the expression has* n *leaves).*

Note that simple tautologies are instuitionistic ones since one of the premises is equal to the goal. The obvious relations between classes above are the following.

$$\begin{array}{rcl} SN_k \cup LN_k & \subset & \mathcal{F}_k \setminus Cl_k \\ SN_k \cap LN_k & = & \emptyset \\ G_k & \subsetneq & Int_k \subsetneq \; Cl_k \subsetneq \; \mathcal{F}_k \setminus (SN_k \cup LN_k) \\ Pierce_k & = & Cl_k \setminus Int_k \end{array}$$

Our aim in the rest of this paper will be to compute the densities of these sets. Results are summed up in Figure 2; proofs are given in the following section. As a consequence, we obtain the following result, giving a positive answer to the conjecture of [9, page 593].

Theorem 6 *Asymptotically (for a large number* k *of Boolean variables), all tautologies are simple i.e.*

$$\lim_{k\rightarrow\infty} \frac{\mu(G_k)}{\mu(Cl_k)} = 1.$$

Proof: We know that for any k the density of classical logic with k propositional variables $\mu(Cl_k)$ exists. Such a result is obtained by standard technics in analysis of algorithms; we skip the details and refer the interested reader to Flajolet and Sedgewick [3] or to [4].

Since $G_k \subset Cl_k \subset \mathcal{F}_k \setminus (SN_k \cup LN_k)$, and from the densities obtained in propositions 8, 9 and 10, we have

$$\frac{4k+1}{(2k+1)^2} = \mu(G_k) \leqslant \mu(Cl_k) \leqslant 1 - \left(\frac{k(k-1)}{(k+1)^2} + \frac{2k(k-1)^2}{(k+2)^4} \right).$$

The upper and lower bounds are asymptotically identical, equal to $1/k + O(1/k^2)$. $\square$

Using the very same argument we can also obtain a result relating the asymptotic behavior of classical versus intuitionistic logics.

Corollary 7 *Asymptotically (for a large number k of Boolean variables), classical tautologies are intuitionistic i.e.*

$$\lim_{k\to\infty} \frac{\mu^-(Int_k)}{\mu(Cl_k)} = 1$$

where $\mu^-(Int_k) = \liminf_{n\to\infty} \frac{|Int_k^n|}{|\mathcal{F}_k^n|}$.

Proof: From the fact that $G_k \subset Int_k \subset Cl_k$, we have

$$\mu(G_k) = \lim_{n\to\infty} \frac{|G_k^n|}{|\mathcal{F}_k^n|} \leqslant \liminf_{n\to\infty} \frac{|Int_k^n|}{|\mathcal{F}_k^n|} \leqslant \limsup_{n\to\infty} \frac{|Int_k^n|}{|\mathcal{F}_k^n|} \leqslant \lim_{n\to\infty} \frac{|Cl_k^n|}{|\mathcal{F}_k^n|} = \mu(Cl_k).$$

The result follows from the fact that both $\mu(G_k)$ and $\mu(Cl_k)$ are equal to $1/k + O(1/k^2)$. □

This result also allows to estimate the size of the difference between classical and intuitionistic logics (so called Pierce formulas). Details are given in section 4.4.

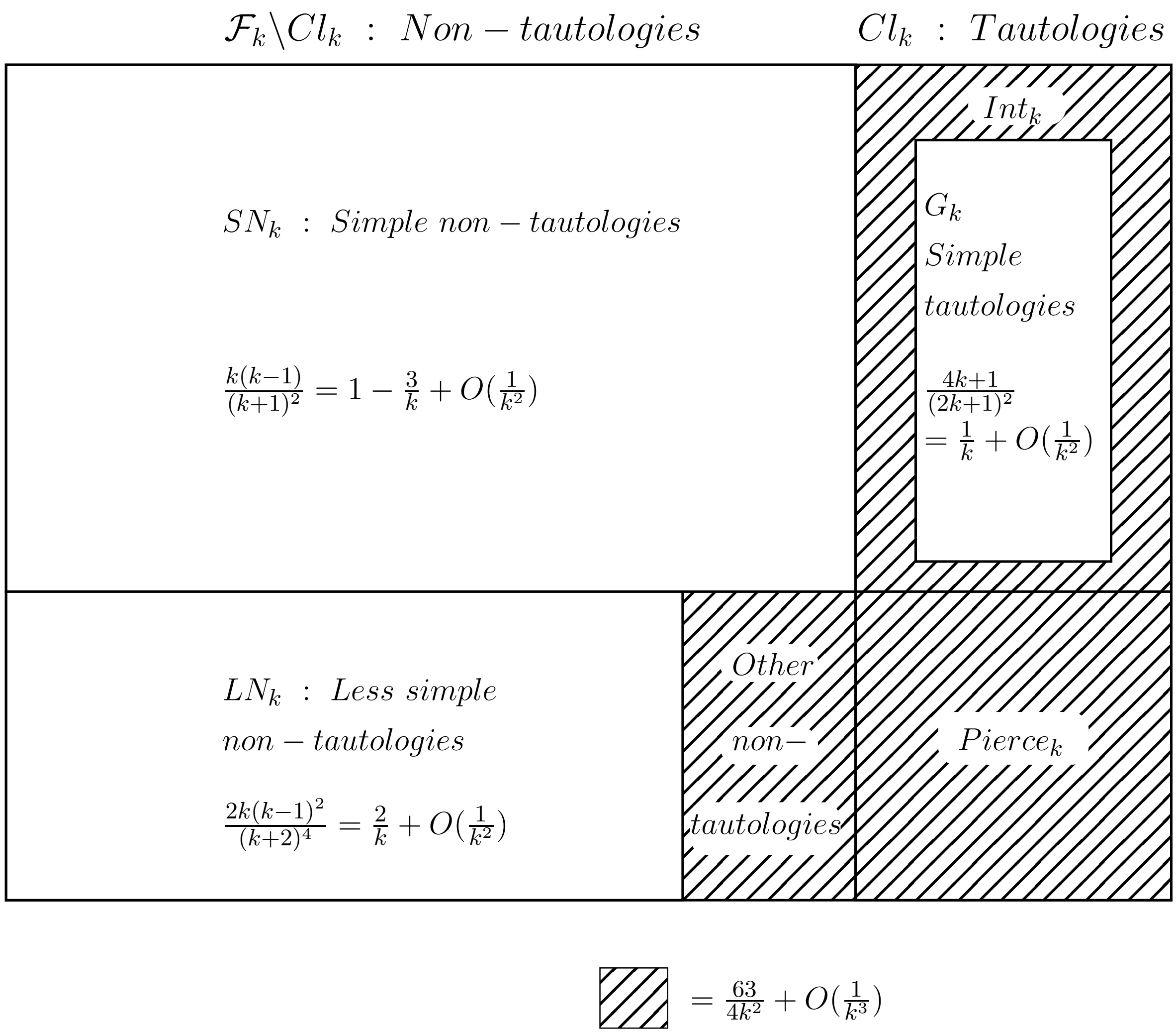

Figure 2: Densities of simple tautologies, simple and less simple non-tautologies.

4 Enumeration of classes

We now compute the densities of the three sets SN_k, G_k and LN_k. The computation of these densities is done in a systematic way. First each set of expressions is defined recursively from simpler sets; this allows to build the generating functions enumerating the elements of these sets by their size (the number of leaves), and to obtain a generating function ϕ for the considered class. Then we extract the coefficient $[z^n]\phi(z)$ and obtain the density of the set under study as $\lim_{n\to\infty}[z^n]\phi(z)/[z^n]f(z)$.

The last part deals with Pierce formulas. Although we don't know if this set of formulas has a density, we give some bounds and show that their order is $\Theta(1/k^2)$.

4.1 Simple non-tautologies

We first consider the set SN_k of simple expressions that are non-tautologies. If $T \in SN_k$, then T is of the kind

$$T = A_1, \dots, A_p \to r(T),$$

such that for all i, $r(A_i) \neq r(T)$. We first check that this is indeed not a tautology. Just consider the following valuation of propositional variables. Define $r(T)$ as *false* and all $r(A_i)$ as *true*. Under this valuation the whole expression is *false*. Let us next compute the generating function $SN(z)$ associated to SN_k.

First fix a Boolean variable α and consider all trees with $r(T) = \alpha$. Such a tree is a simple non-tautology if and only if all its premises A_i satisfy $r(A_i) \neq \alpha$. The generating function of all possibles premises is $\frac{k-1}{k}f(z)$. As a simple non-tautology with goal α is a sequence of such premises followed by the leaf α, the generating function SN^α of simple non-tautologies with goal α is equal to

$$SN^\alpha(z) = \frac{z}{1 - \frac{k-1}{k}f(z)}.$$

Since α can be chosen arbitrarily among the k litterals, we have $SN(z) = k \cdot SN^\alpha(z)$, which gives

$$SN(z) = \frac{kz}{1 - \frac{k-1}{k}f(z)}.$$

Proposition 8 *The density of simple non-tautologies exists and is equal to*

$$\mu(SN_k) = \frac{k(k-1)}{(k+1)^2}.$$

For large k, this density is $1 - 3/k + O(1/k^2)$.

Proof: This result was already given in the paper [9, page 586], with a different proof. We give an alternative proof here. If it exists, the density is given by the following formula:

$$\mu(SN_k) \quad = \quad \lim_{n\to\infty} \frac{|SN_k^n|}{|\mathcal{F}_k^n|} = \lim_{n\to\infty} \frac{[z^n]SN(z)}{[z^n]f(z)}.$$

After modifying the denominator of the generating function $SN(z)$, we obtain :

$$SN(z) = \frac{P(z) + kz(1-k)\sqrt{1-4kz}}{2(1 + z(k-1)^2)},$$

where $P(z)$ is a suitable polynomial. The denominator of the rational fraction $SN(z)$ has a unique zero $\rho = -1/(k-1)^2$. However this value also cancels the numerator of the expression, and is not an actual pole. Hence the only singularity that matters asymptotically is $z = 1/4k$. Putting aside the error term, we obtain

$$\begin{aligned}
[z^n]SN(z) &= -\frac{2k^2(k-1)}{(k+1)^2}[z^{n-1}]\sqrt{1-4kz} \\
&= -\frac{2k^2(k-1)}{(k+1)^2}(4k)^{n-1}[z^{n-1}]\sqrt{1-z} \\
&= -\frac{2k(k-1)}{(k+1)^2}k^n \cdot (-2)C_{n-2} \\
&= \frac{4k(k-1)}{(k+1)^2}k^n C_{n-2}.
\end{aligned}$$

This gives

$$\mu(SN_k) = \lim_{n\to\infty} \frac{|SN_k^n|}{|\mathcal{F}_k^n|} = \frac{4k(k-1)}{(k+1)^2} \lim_{n\to\infty} \frac{C_{n-2}}{C_{n-1}} = \frac{k(k-1)}{(k+1)^2},$$

hence the density of SN_k exists and is equal to $k(k-1)/(k+1)^2$. $\square$

4.2 Simple tautologies

If T is a simple tautology, then T can be written as

$$T = A_1, \ldots, A_p \to r(T),$$

with one of the A_i equal to $r(T)$. It is straightforward to check that T is indeed a tautology since it is logically equivalent with

$$T \sim \overline{A_1} \vee \ldots \vee \overline{r(T)} \vee \ldots \vee \overline{A_p} \vee r(T).$$

which obviously evaluates to *true*.

Let us now compute the generating function of simple tautologies. A tree T is not a simple tautology if and only if all its premises are different from $r(T)$ – see figure 3. The generating function for $\mathcal{F}_k \setminus G_k$ is therefore equal to $kz/(1-(f(z)-z))$. It follows that the generating function of G_k is

$$G(z) = f(z) - \frac{kz}{1+z-f(z)}.$$

Proposition 9 *The limit density of simple tautologies on k variables exists and is equal to*

$$\mu(G_k) = \frac{4k+1}{(2k+1)^2}.$$

For large k, this density is asymptotically equal to $1/k - 3/4k^2 + O(1/k^3)$.

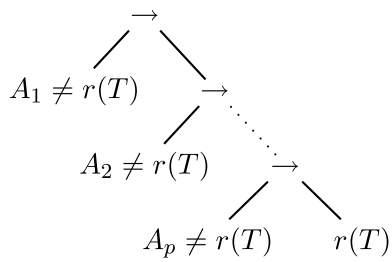

Figure 3: Trees that are not simple tautologies.

Proof: Another, earlier proof of this result is given in the paper [9, page 584]. We give here an alternative proof. The generating function $G(z)$ can be written as

$$G(z) = \frac{P(z) - (1+z)\sqrt{1-4kz}}{2(1+k+z)},$$

with P a suitable polynomial. Let ρ be its pole; $\rho = -1-k$. But ρ is larger that the algebraic singularity $1/(4k)$; hence $1/(4k)$ is the dominant singularity of $G(z)$. Finally we obtain (up to the error term)

$$\begin{aligned}
[z^n]G(z) &= -\frac{2k}{(2k+1)^2}[z^n]\sqrt{1-4kz} - \frac{2k}{(2k+1)^2}[z^{n-1}]\sqrt{1-4kz} \\
&= -\frac{2k}{(2k+1)^2}4^n k^n [z^n]\sqrt{1-z} - \frac{2k}{(2k+1)^2}4^{n-1}k^{n-1}[z^{n-1}]\sqrt{1-z} \\
&= \frac{4k}{(2k+1)^2}k^n C_{n-1} + \frac{4}{(2k+1)^2}k^n C_{n-2}.
\end{aligned}$$

Let us prove the existence and compute the value of the density of G_k^n.

$$\begin{aligned}
\mu(G_k) &= \lim_{n\to\infty}\frac{|G_k^n|}{|F_k^n|} = \lim_{n\to\infty}\left(\frac{4k}{(2k+1)^2}k^n C_{n-1} + \frac{4}{(2k+1)^2}k^n C_{n-2}\right)\cdot\frac{1}{k^n C_{n-1}} \\
&= \frac{4k}{(2k+1)^2} + \frac{4}{(2k+1)^2}\cdot\lim_{n\to\infty}\frac{C_{n-2}}{C_{n-1}} \\
&= \frac{4k}{(2k+1)^2} + \frac{4}{(2k+1)^2}\cdot\frac{1}{4} \\
&= \frac{4k+1}{(2k+1)^2}.
\end{aligned}$$

This density does exist, and is equal to : $(4k+1)/(2k+1)^2$. □

4.3 Less simple non-tautologies

In the family SN_k of simple non-tautologies, we did not allow any premise to have a rightmost leaf equal to $r(T)$. But here we will consider trees with exactly one such premise.

We recall that a tree T defines a less simple non-tautology if it is of the kind $T = B_1, \ldots, B_{i-1}, C, B_i, \ldots, B_p \to r(T)$, where $C = C_1, \ldots, C_q \to r(C)$, with $r(C) = r(T)$, $q \geqslant 1$, and $C_1 = D_1, D_2, \ldots, D_r \to r(D)$ is such that $r(D) \neq r(T)$, $r \geqslant 0$, and the following holds: for all j, $r(B_j) \notin \{r(T), r(D)\}$ and $r(D_j) \notin \{r(T), r(D)\}$. See figure 4 for the general form of the tree and figure 5 for the subtree C; in these figures, if a subtree A is underlined, it means that it is subject to the constraint $r(A) \notin \{r(T), r(D)\}$.

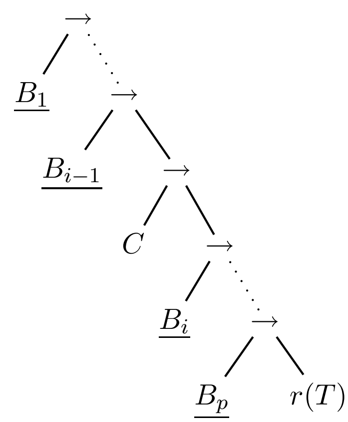

Figure 4: Less simple non-tautologies.

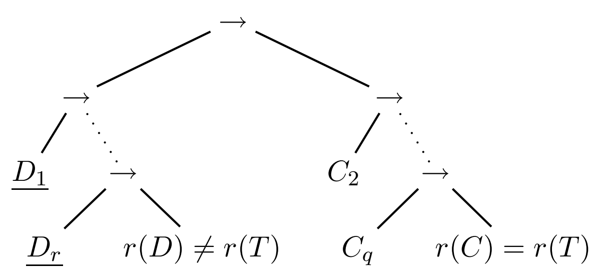

Figure 5: Subtree C of a less simple non-tautology.

Let us first prove that such a tree is not a tautology. For this, consider the assignement where all the variables are *true*, except $r(T)$ and $r(D)$ which are *false*; under this assignement, the whole expression evaluates to *false* – to check this, just notice that the function computed by such a tree can be developped into a conjuction of terms, one of them being $\bigvee_i \overline{r(B_i)} \vee r(T) \vee \bigvee_i \overline{r(D_i)} \vee r(D)$.

We shall now compute the generating function of LN_k. Let us fix α and β two distinct literals. We shall first compute $\psi(z)$ the generating functions of all trees $LN_k^{\alpha,\beta}$ from LN_k such that $r(T) = \alpha$ and $r(D) = \beta$. By symmetry, $\psi(z)$ is independent of the choice of α and β.

Let $b(z)$ be the generating function of all trees $T \in \mathcal{F}_k$ such that $r(T) \notin \{\alpha, \beta\}$. Of course $b(z) = (k-2)/k \cdot f(z)$. This generating function enumerate the possible subtrees B_j but also the possible subtrees D_j. Thus, the generating function of all possible trees for D is $d(z) = z/(1-b(z))$, since it is a sequence of trees D_j such that $r(D_j) \notin \{\alpha, \beta\}$, followed by the leaf β. In the same way, the generating function for the subtree C is $c(z) = d(z) \cdot 1/(1-f(z)) \cdot z$. Note that a tree of $LN_k^{\alpha,\beta}$ is built as a sequence of trees B_j with $r(B_j) \notin \{\alpha, \beta\}$, then a subtree

C as described as above, another sequence of trees B_j with $r(B_j) \notin \{\alpha, \beta\}$, followed by the leaf α. Moreover, this decomposition is unique. The generating function for $LN_k^{\alpha,\beta}$ is thus equal to

$$\psi(z) = \frac{1}{1-b(z)} c(z) \frac{1}{1-b(z)} z.$$

Now it can be easily seen that LN_k is the disjoint union of the $LN_k^{\alpha,\beta}$. Indeed, given a tree $T \in LN_k$, then α is equal to $r(T)$ and the premise C of T is uniquely determined because it is the only premise of T with goal $r(T)$. Thus, β is uniquely determined as well since it is the goal of the first premise of C. It follows that $\phi(z) = k(k-1)\psi(z)$.

Proposition 10 *The density of less simple non-tautologies is equal to*

$$\mu(LN_k) = \frac{2k(k-1)^2}{(k+2)^4}.$$

For large k it is equal to $2/k + O(1/k^2)$.

Proof: After modifying the denominator of the generating function $\phi(z)$, we obtain :

$$\phi(z) = \frac{P(z) + k(k-1)(-k^2 + (2k^3 - 6k^2 + 8)z)z^2\sqrt{1-4kz}}{2(2+(k-2)^2 z)^3},$$

where $P(z)$ is a suitable polynomial. The denominator of the rational fraction $\phi(z)$ has a zero $\rho = -2/(k-2)^2$. However this value also cancels the numerator (and its first two derivatives) of the expression, and is not an actual pole of ϕ. Hence the only singularity that matters asymptotically is $z = 1/4k$. Putting aside the error term, we obtain :

$$\begin{aligned}
[z^n]LN(z) &= -\frac{k^3(k-1)}{2(2+\frac{(k-2)^2}{4k})^3}[z^{n-2}]\sqrt{1-4kz} \\
&\quad + \frac{k(k-1)(2k^3-6k^2+8)}{2(2+\frac{(k-2)^2}{4k})^3}[z^{n-3}]\sqrt{1-4kz} \\
&= -\frac{k^3(k-1)}{2(2+\frac{(k-2)^2}{4k})^3}4^{n-2}k^{n-2}[z^{n-2}]\sqrt{1-z} \\
&\quad + \frac{k(k-1)(2k^3-6k^2+8)}{2(2+\frac{(k-2)^2}{4k})^3}4^{n-3}k^{n-3}[z^{n-3}]\sqrt{1-z} \\
&= \frac{k^{n+1}(k-1)}{(2+\frac{(k-2)^2}{4k})^3}C_{n-3} - \frac{k^{n-2}(k-1)(2k^3-6k^2+8)}{(2+\frac{(k-2)^2}{4k})^3}C_{n-4}.
\end{aligned}$$

Let us prove the existence and compute the value of the density of LN_k^n :

$$\begin{aligned}
\mu(LN) &= \lim_{n\to\infty} \frac{|LN_k^n|}{|F_k^n|} \\
&= \lim_{n\to\infty} \Big(\frac{k^{n+1}(k-1)}{(2+\frac{(k-2)^2}{4k})^3}\frac{C_{n-3}}{k^n C_{n-1}} - \frac{k^{n-2}(k-1)(2k^3-6k^2+8)}{(2+\frac{(k-2)^2}{4k})^3}\frac{C_{n-4}}{k^n C_{n-1}} \Big) \\
&= \frac{64k^4(k-1)}{(k+2)^6} \cdot \lim_{n\to\infty}\frac{C_{n-3}}{C_{n-1}} - \frac{64k(k-1)(2k^3-6k^2+8)}{(k+2)^6} \cdot \lim_{n\to\infty}\frac{C_{n-4}}{C_{n-1}} \\
&= \frac{4k^4(k-1) - k(k-1)(2k^3-6k^2+8)}{(k+2)^6} = \frac{2k(k-1)^2}{(k+2)^4}
\end{aligned}$$

This density does exist, and is equal to:

$$2k(k-1)^2/((k+2)^4).$$

For large k this is asymptotically equal to $2/k + O(1/k^2)$. $\square$

4.4 Pierce formulas

We are ready to estimate the number of Pierce formulas. Although we don't know if the set of Pierce formulas has a density, we shall give bounds on $\limsup_{n\to\infty} \frac{|Pierce_k^n|}{|\mathcal{F}_k^n|}$ and $\liminf_{n\to\infty} \frac{|Pierce_k^n|}{|\mathcal{F}_k^n|}$. A simple upper bound on $Pierce_k$ can be obtained from

$$Pierce_k = Cl_k \setminus Int_k \subset F_k \setminus (SN_k \cup LN_k \cup G_k).$$

Since SN_k, LN_k and G_k are disjoint we have a simple upper estimation based on propositions 8, 9 and 10:

$$\limsup_{n\to\infty} \frac{|Pierce_k^n|}{|\mathcal{F}_k^n|} \leqslant 1 - \frac{k(k-1)}{(k+1)^2} - \frac{2k(k-1)^2}{(k+2)^4} - \frac{4k+1}{(2k+1)^2} = \frac{63}{4k^2} + O(\frac{1}{k^3}).$$

However, we can obtain a sharper bound on the number of Pierce formulas. For this, we next bound the density of tautologies which are not simple – this density exists since we already know that both the density of all tautologies and the density of simple tautologies exist. Note that this result gives an alternative proof for Theorem 6.

Lemma 11 *The density of non simple tautologies T such that exactly one premise has a goal equal to $r(T)$ is bounded from above by $5/k^2 + O(1/k^3)$.*

Proof: Let A be a non simple tautology with goal $r(A) = \alpha$. Let p be the number of premises of A. We call B the premise of A whose goal is $r(A)$ and $\alpha_1, \ldots, \alpha_{p-1}$ the goal of the premises other than B. By hypothesis, $\alpha_i \neq \alpha$ for all $i \in \{1, \ldots, p-1\}$. Of course B cannot be reduced to a leaf (otherwise A would be a simple tautology). Let us decompose $B = (B_1, \ldots, B_m, \alpha)$, with $m \geqslant 1$. As $\overline{B} = B_1 \wedge \ldots \wedge B_m \wedge \overline{\alpha}$, by developping the expression A, we obtain that necessarily, for all $j \in \{1, \ldots, m\}$,

$$B_j \vee \overline{\alpha}_1 \ldots \vee \overline{\alpha}_{p-1} \vee \alpha$$

computes *true*. Let us denote $\mathcal{C}_{(\alpha_1,\ldots,\alpha_{p-1},\alpha)}$ the set of trees such that

$$C \vee \overline{\alpha}_1 \ldots \vee \overline{\alpha}_{p-1} \vee \alpha$$

computes *true*. Let $C \in \mathcal{C}_{(\alpha_1,\ldots,\alpha_{p-1},\alpha)}$.

- If C is reduced to a leaf γ then necessarily $\gamma \in \{\alpha_1, \ldots, \alpha_{p-1}\}$.
- Otherwise, let us decompose $C = (C_1, \ldots, C_s, \gamma)$ with $s \geqslant 1$. Let $\gamma_i = r(C_i)$. Then

$$\overline{\gamma_1} \vee \ldots \vee \overline{\gamma_s} \vee \gamma \vee \overline{\alpha_1} \ldots \vee \overline{\alpha_{p-1}} \vee \alpha$$

  has to evaluate to *true*. It follows that $\alpha \in \{\gamma_1, \ldots, \gamma_s\}$ or $\gamma \in \{\gamma_1, \ldots, \gamma_s, \alpha_1, \ldots, \alpha_{p-1}\}$.

We shall now compute a generating function $c_{(\alpha_1,\ldots,\alpha_{p-1},\alpha)}$ giving an upper bound on the number of trees of $\mathcal{C}_{(\alpha_1,\ldots,\alpha_{p-1},\alpha)}$. Let us define

$$c_{(\alpha_1,\ldots,\alpha_{p-1},\alpha)}(z) = (p-1)z + \frac{1}{1-((k-1)/k)f(z)} \cdot \frac{f(z)}{k} \cdot \frac{1}{1-f(z)} \cdot kz + \sum_{s=1}^{\infty} f(z)^s \cdot (s+p-1)z$$

the first term corresponding to the first point above, the second term corresponding to the case $\alpha \in \{\gamma_1,\ldots,\gamma_s\}$ and the third term to the case $\gamma \in \{\gamma_1,\ldots,\gamma_s,\alpha_1,\ldots,\alpha_{p-1}\}$. This generating function depends only on p; thus we shall now denote it by c_p. Let us now define

$$b_p(z) = \frac{c_p(z)}{1-c_p(z)} \cdot z.$$

This function gives an upper bound on the number of trees B (for $p \geqslant 1$ and $\alpha, \alpha_1, \ldots, \alpha_{p-1}$ fixed) such that

$$B \vee \overline{\alpha}_1 \ldots \vee \overline{\alpha}_{p-1} \vee \alpha$$

computes *true.* Of course

$$b_p(z) \leqslant \tilde{b}_p(z) := c_p(z) + \frac{(c_p(z))^2}{1-f(z)}.$$

We now define

$$a_p(z) = p \cdot ((k-1)/k \cdot f(z))^{p-1} \cdot \tilde{b}_p(z) \cdot z \cdot k.$$

The generating function a_p gives an upper bound on the number of non simple tautologies A with p premises, exactly one of them having a goal equal to $r(A)$. Indeed, z corresponds to $r(A) = \alpha$, k corresponds to the choice of α among the litterals and p corresponds to the position of the unique premise with goal α.

We now define $a(z) = \sum_{p=1}^{\infty} a_p(z)$. This function bounds the number of non simple tautologies A with only on premise with goal $r(A)$. The computation based on the generating function defined above leads to an asymptotic density $5/k^2 + O(1/k^3)$. □

Lemma 12 *The density of non simple tautologies T such that exactly two premises have a goal equal to $r(T)$ is $O(1/k^3)$.*

Proof: Let us consider a non simple tautology A with exactly two premises B_1 and B_2 having a goal equal to $r(A)$. Let $\alpha_1, \ldots, \alpha_{p-2}$ the goals of the other premises. Since A is not simple, both B_1 and B_2 are not reduced to a leaf. Let C be the first premise of B_1, and D be the first premise of B_2. Let γ be the goal of C and $\gamma_1, \ldots, \gamma_s$ the goals of its premises (with $s \geqslant 0$). We define $\delta, \delta_1, \ldots, \delta_t$ the corresponding litterals for the tree D. Since A is a tautology we can argue as in the previous lemma and we obtain that necessarily

$$\overline{\gamma_1} \vee \ldots \vee \overline{\gamma_s} \vee \gamma \vee \overline{\delta_1} \vee \ldots \vee \overline{\delta_t} \vee \delta \vee \overline{\alpha_1} \ldots \vee \overline{\alpha_{p-2}} \vee \alpha$$

evaluates to *true.* The same method as in the previous lemma (not detailed here) leads to a density $O(1/k^3)$. □

Lemma 13 *The asymptotic density of trees T such that at least three premises have a goal equal to $r(T)$ is $O(1/k^3)$.*

Proof: The generating function of this family of trees is equal to

$$\left(\frac{1}{1-(k/(k-1))f(z)} \cdot \frac{f(z)}{k}\right)^3 \cdot \frac{1}{1-f(z)} \cdot kz.$$

We obtain a density $O(1/k^3)$. □

Proposition 14 *The asymptotic density of non simple tautologies is bounded from above by $5/k^2 + O(1/k^3)$.*

Proof: A tautology is not reduced to a leaf. Moreover, a tautology T has (at least) a premise with goal $r(T)$: otherwise, it would be a simple non-tautology. The density of non simple tautologies is thus bounded from above by the sum of the three densities obtained in lemmas 11, 12 and 13. Hence it is bounded above by $5/k^2 + O(1/k^3)$. □

We can obtain a lower bound for Pierce formulas by the following argument. Consider special formulas from $\mathcal{F}_k$ of the form $((a \to T) \to a) \to a$ where $T = A_1, \ldots, A_p \to r(T)$ is a simple non-tautology taken from $\mathcal{F}_k$ (see section 4.1) and variable a differs from $r(T)$. We observe that $((a \to T) \to a) \to a$ must be a Pierce formula. It is obviously a classical tautology. Suppose $((a \to T) \to a) \to a$ is also an intuitionistic tautology. It means that there must exist a closed term of the type $((a \to T) \to a) \to a$. The long normal form of this term has the form $\lambda p_{(a \to T) \to a}.p(\lambda q_a.t)$ where t is a term of type T with only free variables p and q. Consider a closed term $\lambda p_{(a \to T) \to a} \lambda q_a.t$. The type of this term is the implicational formula

$$((a \to T) \to a) \to (a \to T).$$

But this type is again a simple non-tautology since the variables a and $r(T)$ are different. So the formula is unprovable classically and therefore intuitionistically too; contradiction. For more details about relation between intuitionistic logic and lambda calculus consult for example Sørensen, Urzyczyn [10].

Now we have to count this family. The number of such formulas is $(k-1) \cdot |SN_k^{n-3}|$. Thus the density of this special set of Pierce formulas exists and is equal to

$$\lim_{n\to\infty} \frac{(k-1) \cdot |SN_k^{n-3}|}{|\mathcal{F}_k^n|} = \lim_{n\to\infty} \frac{(k-1) \cdot |SN_k^{n-3}|}{|\mathcal{F}_k^{n-3}|} \cdot \frac{|\mathcal{F}_k^{n-3}|}{|\mathcal{F}_k^n|} = \frac{1}{64k^2} \frac{(k-1)^2}{(k+1)^2}$$

since $\lim_{n\to\infty} |\mathcal{F}_k^{n-3}|/|\mathcal{F}_k^n| = 1/(4k)^3$.

Proposition 15 *We have the following bounds on the number of Pierce formulas:*

$$\frac{1}{64k^2} - O\left(\frac{1}{k^3}\right) \leqslant \liminf_{n\to\infty} \frac{|Pierce_k^n|}{|\mathcal{F}_k^n|} \leqslant \limsup_{n\to\infty} \frac{|Pierce_k^n|}{|\mathcal{F}_k^n|} \leqslant \frac{5}{k^2} + O\left(\frac{1}{k^3}\right).$$

Proof: The lower bound comes from the previous discussion. Since Pierce formulas are non simple tautologies, the upper bound is a consequence of proposition 14. □

5 Final remarks

We have shown that asymptotically, all tautologies over implication are simple, i.e. one of the premises is equal to the goal. The method developped in this paper extends to the logic of implication with both positive and negative litterals. In this new setting again, we can prove that most of the tautologies, when the number of variables becomes large, exhibit a very simple structure; more precisely, most of the tautologies have one of their premises equal to the goal (as before), or have two of their premises which are opposite litterals.

Some questions remain about the set of Pierce formulas. We conjecture that for any k, the densities $\mu(Int_k)$ and $\mu(Pierce_k)$ exist. If it is the case, it would be interesting to evaluate the asymptotic densities of these sets.

References

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

[2] P. Flajolet and R. Sedgewick. *Analytic combinatorics: functional equations, rational and algebraic functions,* INRIA, Number 4103, 2001.

[3] P. Flajolet and R. Sedgewick. *Analytic combinatorics.* Book in preparation, available at `http://algo.inria.fr/flajolet/Publications/books.html`, 2007.

[4] D. Gardy. Random Boolean expressions, *Colloquium on Computational Logic and Applications*, Chambery (France), June 2005. Proceedings in DMTCS, pp 1-36, 2006.

[5] D. Gardy and A. Woods. And/or tree probabilities of Boolean function, *Discrete Mathematics and Theoretical Computer Science*, pp 139-146, 2005.

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

[7] H. Lefmann and P. Savický. Some typical properties of large And/Or Boolean formulas, *Random Structures and Algorithms*, vol 10, pp 337-351, 1997.

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

[9] M. Moczurad, J. Tyszkiewicz and M. Zaionc. Statistical properties of simple types, *Mathematical Structures in Computer Science,* 10(5):575-594, 2000.

[10] M. Sørensen and P. Urzyczyn. *Lectures on the Curry-Howard Isomorphism*, volume 149 of *Studies in Logic and the Foundations of Mathematics*. Elsevier Science, 2006.

[11] H. Wilf. *Generatingfunctionology*, second edition. Academic Press, Boston, 1994.

[12] M. Zaionc. On the asymptotic density of tautologies in logic of implication and negation, *Reports on Mathematical Logic*, vol 39, pp 67-87, 2005.

[13] M. Zaionc. Probability distribution for simple tautologies, *Theoretical Computer Science*, 355(2):243-260, 2006.