Tautologies over implication with negative literals*

Hervé Fournier† Danièle Gardy† Antoine Genitrini† Marek Zaionc‡

September 14, 2009

Abstract

We consider logical expressions built on the single binary connector of implication and a finite number of literals (boolean variables and their negations). We prove that asymptotically, when the number of variables becomes large, all tautologies have the following simple structure: either a premise equal to the goal, or two premises which are opposite literals.

Keywords: Implicational expressions; Boolean formulae; Tautologies; Analytic combinatorics.

1 Introduction

We consider the logical system of Boolean expressions built on the single connector of implication and on literals from k variables. Our interest lays in tautologies, more precisely in the proportion of tautologies among all expressions of size n: our aim is to prove the existence of a limit of that fraction as n grows to infinity, and to compute its limit. This limit gives, in a way a measure of the "density" of truth for the logic of implication and literals over k variables. After isolating a special class of expressions called simple tautologies, for which we can compute explicitely the asymptotic density for large k, we prove that asymptotically all tautologies are simple.

The present work is part of a research in which the likelihood of truth is estimated for various propositional logics with a finite number of variables. A result correlated to our study was presented in [4] about the simpler logic equipped with implication and positive literals only. 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 by Moczurad, Tyszkiewicz and Zaionc [14]. The tautologies over purely implicational logic studied there correspond to inhabited types in simple lambda-calculus and the question of enumerating them had been raised earlier by Statman [16] – for more about motivations, we refer the reader to [14] and references therein. The classical logic of one variable and the two connectors of implication

*This work was partially supported by the PHC Polonium project "Mikolaj", by the ANR project "SADA" (Structures Aléatoires Discrètes et Algorithmes), and by grant number NN206 376137, Polish Ministry of Science and Higher Education.

†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.

‡Theoretical Computer Science, Jagiellonian University, Łojasiewicza 6, 30-348 Kraków, Poland. Email: zaionc@tcs.uj.edu.pl.

and negation was studied in Zaionc [19]; over the same language, the exact proportion between intuitionistic and classical logics has been determined by Kostrzycka and Zaionc [10]. Some variants involving expressions with other logical connectives have also been considered. Genitrini and Kozik have studied the influence of adding the connectors $\vee$ and $\wedge$ to implication [7], while Matecki [13] considered the case of the single equivalence connector. For two connectors again, the *and/or* case has already received much attention – see Lefmann and Savický [12], Chauvin, Flajolet, Gardy and Gittenberger [1], Gardy and Woods [6], Woods [18] and Kozik [11]. Let us also mention the survey [5] on the probability distributions on Boolean functions induced by random Boolean expressions; this survey deals with the whole set of Boolean functions on some finite number of variables, whereas the present work is an in-depth study of the expressions that compute the constant function *True* in a specific system for propositional logic.

The organization of the paper is as follows. We present our propositional system and the basic definitions in Section 2, then recall in Section 3 some necessary facts about generating functions and asymptotic methods that we shall use to obtain our main results. We prove asymptotic results for the logic equipped with the binary connector of implication and a finite number of positive and negative literals in Section 4. We show there our main result: *When the number of Boolean variables grows to infinity, asymptotically all the tautologies have one of their premises equal to their goal, or have two opposite literals among their premises.* Finally we present some concluding remarks in Section 5.

2 Formulae over implication and negative literals

Definition 1 *Let $\{x_1, x_2, \ldots, x_k\}$ be a set of Boolean propositional variables. We define $\mathcal{F}_k$ to be the set of all Boolean expressions (or formulae) over the literals $\{x_1, \bar{x}_1, \ldots, x_k, \bar{x}_k\}$ and the implication connective $\rightarrow$: Boolean expressions are defined recursively from Boolean literals and the implication connective by the following grammar:*

$$F := x_1 \mid \bar{x}_1 \mid \ \ldots \ \mid x_k \mid \bar{x}_k \mid (F \rightarrow F).$$

An expression of $\mathcal{F}_k$ naturally computes a Boolean function over the variables $\{x_1, \ldots, x_k\}$. Notice that we do not obtain all the functions expressible in the whole implicational-negational fragment of the logic: For example the function $\neg(x \rightarrow y)$ is not expressible within this logic.

Obviously the expressions can be represented by full (without unary nodes) binary planar trees, whose nodes have been suitably labelled: the internal nodes by the connector $\rightarrow$, the leaves by some Boolean literals. From now on, we shall use indifferently "tree" or "expression".

We define next a *canonical form of an expression.* Let T be an expression. It can be decomposed with respect to its right branch; hence it is of the form

$$A_1 \rightarrow (A_2 \rightarrow (\ldots \rightarrow (A_p \rightarrow \alpha) \ldots)),$$

where $A_i \in \mathcal{F}_k$ and $\alpha \in \{x_1, \bar{x}_1, \ldots, x_k, \bar{x}_k\}$. We shall write this as

$$T \equiv A_1, \ldots, A_p \rightarrow \alpha.$$

The expressions A_i are called the *premises* of T and the rightmost leaf of the tree α is called the *goal* of T. We shall denote the goal of a tree T by $r(T)$. The goals of the premises of T (i.e., $r(A_1), \ldots, r(A_p)$) will be called the *subgoals* of T. Of course the expression $T = A_1 \rightarrow$

$(A_2 \to (\ldots \to (A_p \to \alpha)\ldots))$ is logically equivalent to $\bar{A}_1 \vee \bar{A}_2 \vee \cdots \vee \bar{A}_p \vee \alpha$, where $\bar{A}_i$ stands for negation of A_i.

For a formula $A \in \mathcal{F}_k$, we denote by $\|A\|$ the *size* of A, which we define as the total number of occurrences of propositional variables in A (or leaves in the tree representation of this formula). Parentheses and the implication sign itself are not included in the size of a formula. Formally,

$$\|x_i\| = \|\bar{x}_i\| = 1 \text{ and } \|A \to B\| = \|A\| + \|B\|\,.$$

Throughout this paper we denote by $|X|$ the cardinal of any finite set X. For a subset $\mathcal{A} \subseteq \mathcal{F}_k$ we define the *density*[1] $\mu_k(\mathcal{A})$ as

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

if the limit exists. We immediately see that the density μ_k is finitely additive: if $\mathcal{A}$ and $\mathcal{B}$ are disjoint classes of expressions such that $\mu_k(\mathcal{A})$ and $\mu_k(\mathcal{B})$ both exist, then $\mu_k(\mathcal{A} \cup \mathcal{B})$ also exists and $\mu_k(\mathcal{A} \cup \mathcal{B}) = \mu_k(\mathcal{A}) + \mu_k(\mathcal{B})$. Not all subsets of $\mathcal{F}_k$ have a well-defined density; hence, we define

$$\mu_k^+(\mathcal{A}) = \limsup_{n\to\infty} \frac{|\{A \in \mathcal{A} : \|A\| = n\}|}{|\{A \in \mathcal{F}_k : \|A\| = n\}|}.$$

This quantity is well-defined for any family $\mathcal{A}$ of formulae, even when the density of $\mathcal{A}$ is not known to exist.

3 Generating functions

We shall investigate the ratio of expressions that are tautologies among all expressions of size n in 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 [17], or in Flajolet and Sedgewick [3]; see also Gardy [5, Section 5.2] for a systematic application of these techniques to the computation of probability distributions for Boolean functions.

Let $(a_0, a_1, a_2, \ldots)$ be a sequence of real numbers. The *ordinary generating series* for the sequence (a_n) is the formal power series $\sum_{n=0}^{\infty} a_n z^n$. 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 $f(z)$ in some open disc $\{z \in \mathbb{C} : |z| < R\}$ of maximal diameter, and $R \geqslant 0$ is called its radius of convergence. So, when $R > 0$, we can associate with the sequence (a_n) a complex function $f(z) = \sum_{n=0}^{\infty} a_n z^n$, called the *(ordinary) generating function* for (a_n), defined in a neighbourhood of 0. In the other way, as is well known from the theory of analytic functions, the expansion of a complex function $f(z)$, analytic in a neighbourhood of z_0, into a power series $\sum_{n=0}^{\infty} a_n (z - z_0)^n$ is unique. For a function $g(z)$ analytic in a neighbourhood of 0, we shall denote by $[z^n]g$ the coefficient of z^n in the series expansion of g in 0.

Many questions concerning the asymptotic behaviour of the sequence (a_n) can be efficiently resolved by analysing the behaviour of $\sum a_n z^n$ at the complex circle $|z| = R$. This is

[1] The term *density* is to be taken in some “physical” sense, not in a probabilistic sense: we are *not* dealing with probability densities.

the approach we take to determine the asymptotic fraction of tautologies and other classes of expressions among all expressions of a given size.

Going back to the density of a subset $\mathcal{A} \subseteq \mathcal{F}_k$, the computation of set cardinalities $|\{A \in \mathcal{A} : \|A\| = n\}|$ can be done through the (ordinary) generating function

$$\phi_{\mathcal{A}}(z) = \sum_{A \in \mathcal{A}} z^{\|A\|} = \sum_{n} |\{A \in \mathcal{A} : \|A\| = n\}|\, z^n.$$

Each of the sets of expressions that we shall consider 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 where the variable z marks 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 $\lim_{n\to\infty}[z^n]\phi(z)/[z^n]\phi_{\mathcal{F}_k}(z)$, where $\phi_{\mathcal{F}_k}(z)$ is the generating function for $\mathcal{F}_k$, the set of all expressions.

To compute the generating functions for the sets of expressions relevant to our problem requires us to know how basic constructions on sets translate on generating functions. We next recall three constructions on classes of combinatorial objects, and their effect on generating functions. Let $\mathcal{A}$ and $\mathcal{B}$ be two classes of combinatorial objects, with generating functions $\phi_{\mathcal{A}}(z)$ and $\phi_{\mathcal{B}}(z)$. The first construction, called *combinatorial sum*, captures the union of disjoint sets. The generating function of the combinatorial sum of $\mathcal{A}$ and $\mathcal{B}$ is $\phi_{\mathcal{A}}(z) + \phi_{\mathcal{B}}(z)$. The second construction, *cartesian product*, forms all possible ordered pairs of objects from $\mathcal{A}$ and $\mathcal{B}$ – the size of (A, B) being the sum of the sizes of A and B. The generating function enumerating this class is $\phi_{\mathcal{A}}(z)\ \phi_{\mathcal{B}}(z)$. Finally the *sequence* construction builds all finite sequences of objects from $\mathcal{A}$. Again the size of a sequence of objects is the sum of their sizes. The generating function enumerating this class is $1/(1 - \phi_{\mathcal{A}}(z))$.

We have seen in Section 2 that a Boolean expression can be represented as a labelled binary tree; hence we may expect that the enumeration of classes of such expressions will amount to enumeration of suitable classes of binary trees, and will involve the classical enumeration of binary trees by Catalan numbers. The Catalan number C_n is defined as the number of full binary trees (i.e. every vertex has either two children or no child) with n internal nodes and $n + 1$ leaves. Basic results about Catalan numbers and their generating function are summarised below; see e.g. [3].

Proposition 2 *Let $C(z)$ be the generating function enumerating full binary trees with respect to the number of leaves; it satisfies:*

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

and is equal to:

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

Its coefficients are

$$[z^{n+1}]C(z) = C_n = \frac{1}{n+1}\binom{2n}{n}.$$

As an example, we obtain the generating function $f_k(z)$ for the set of all the expressions built on k variables and their negations, and the implication connector.

Proposition 3 *The generating function enumerating the set $\mathcal{F}_k$ of all Boolean expressions over k variables satisfies the equation $f_k(z) = 2kz + f_k(z)^2$; it is equal to*

$$f_k(z) = C(2kz) = \frac{1 - \sqrt{1 - 8kz}}{2}.$$

It follows that the number of Boolean expressions of size n over k variables is $(2k)^n C_{n-1}$: such an expression is obtained by labelling the n leaves with any of the literals $x_1, \bar{x}_1, \ldots, x_k, \bar{x}_k$.

Finally, we shall make intensive use of the series expansion

$$[z^n]\sqrt{1 - 8kz} = -(8k)^n \frac{C_{n-1}}{2^{2n-1}}.$$

4 The simple structure of almost all tautologies

In this section, we study tautologies in the logic obtained from implication and (positive or negative) literals. We denote by $\mathcal{T}_k$ the set of all tautologies of $\mathcal{F}_k$, i.e., the set of formulae of $\mathcal{F}_k$ which evaluate to *True* under all assignments of the variables $\{x_1, \ldots, x_k\}$.

Lemma 4 *The density of tautologies $\mu_k(\mathcal{T}_k)$ exists.*

Proof: For a Boolean function f over the variables $\{x_1, \ldots, x_k\}$, consider the generating function ϕ_f of the set of expressions of $\mathcal{F}_k$ computing this function. By considering functions computed by the left and right subtrees of an expression, one can write a quadratic system satisfied by these generating functions. It is easily checked that this system satisfies the conditions of the Drmota-Lalley-Woods theorem, well presented e.g. in the book of Flajolet and Sedgewick [3, Chapter VII]. It follows that the density of the set of trees computing any given Boolean function is well-defined. This holds in particular for the function *True*. $\square$

We shall prove that, when the number of variables becomes large, most of the tautologies of $\mathcal{F}_k$ exhibit the very simple structure presented below.

Definition 5 *Simple tautologies of the first kind $\mathcal{S}_k^{(1)}$ are expressions $A_1, \ldots, A_p \to \alpha$ such that $A_i = \alpha$ for some $i \in \{1, \ldots, p\}$; they are all expressions of the form*

$$\ldots, \ell, \ldots \to \ell$$

for some literal ℓ.

Simple tautologies of the second kind $\mathcal{S}_k^{(2)}$ are expressions $A_1, \ldots, A_p \to \alpha$ such that A_i and A_j are opposite literals for some $i, j \in \{1, \ldots, p\}$; they are all expressions of the form

$$\ldots, \ell, \ldots, \bar{\ell}, \ldots \to \cdot$$

where $\bar{\ell}$ is the negation of ℓ (i.e. $\bar{\ell} = \bar{x}_i$ if $\ell = x_i$ and $\bar{\ell} = x_i$ if $\ell = \bar{x}_i$).

The set of simple tautologies is defined as $\mathcal{S}_k = \mathcal{S}_k^{(1)} \cup \mathcal{S}_k^{(2)}$.

We first evaluate the densities of simple tautologies of the first kind.

Lemma 6 *The density of simple tautologies of the first kind is asymptotically equal to*

$$\mu_k(\mathcal{S}_k^{(1)}) = \frac{1}{2k} + O\left(\frac{1}{k^2}\right).$$

Proof: Let $\mathcal{S}^{(1)}[\ell]$ be the set of tautologies of the first kind, relative to some literal ℓ: the $2k$ sets $\mathcal{S}^{(1)}[\ell]$ form a partition of $\mathcal{S}_k^{(1)}$. A tautology of $\mathcal{S}^{(1)}[\ell]$ is a sequence of premises differing from ℓ, followed by ℓ, then a sequence of arbitrary premises (which may now include ℓ), and finally the goal ℓ. We can write it, in the framework of formal languages (identifying a tree $A_1, \ldots, A_p \to \alpha$ with the word $A_1 \ldots A_p\alpha$ over the alphabet $\mathcal{F}_k$), as

$$\mathcal{S}^{(1)}[\ell] = (\mathcal{F}_k \setminus \{\ell\})^* \cdot \ell \cdot \mathcal{F}_k^* \cdot \ell.$$

We obtain the generating function for $\mathcal{S}^{(1)}[\ell]$ as the product of the generating functions for each of the terms:

$$\frac{1}{1-(f_k(z)-z)} \cdot z \cdot \frac{1}{1-f_k(z)} \cdot z.$$

Since there are $2k$ choices for the literal ℓ involved, the generating function for the whole set $\mathcal{S}_k^{(1)}$ of simple tautologies of the first kind is

$$\sigma_1(z) = 2k \cdot \frac{z}{1-(f_k(z)-z)} \cdot \frac{z}{1-f_k(z)}.$$

Our next step is to prove the existence of the density, and compute its asymptotic value for large k. We can rewrite $\sigma_1(z)$ as

$$\sigma_1(z) = \frac{1+z-4kz-(1+z)\sqrt{1-8kz}}{2(1+2k+z)}.$$

The singularities of σ_1 come from the algebraic singularity $1/(8k)$ of f_k, and from the cancellation of the denominator, which introduces a pole $-(1+2k)$. Obviously the absolute value of this pole is larger that $1/(8k)$; hence the dominant singularity of $\sigma_1(z)$ is $1/(8k)$. The asymptotic behaviour of the coefficients $[z^n]\sigma_1$ being determined by the dominant singularity, and transfer theorems [2, 3] allow to obtain their asymptotic values up to an error term: The function $-(1+z)/(2(1+2k+z))$ takes the value $-(8k+1)/(2(4k+1)^2)$ at the singularity $1/(8k)$, and

$$\begin{aligned} |\mathcal{S}_k^{(1)}(n)| = [z^n]\sigma_1(z) &= -\frac{8k+1}{2(4k+1)^2}\,[z^n]\{\sqrt{1-8kz}\}(1+O(1/n)) \\ &= \frac{8k+1}{2(4k+1)^2}\,(8k)^n\,\frac{C_{n-1}}{2^{2n-1}}(1+O(1/n)). \end{aligned}$$

We are now ready to prove the existence and compute the value of the density of $\mathcal{S}_k^{(1)}$. We have seen that the cardinality of the set of expressions of $\mathcal{F}_k$ of size n, denoted below as $\mathcal{F}_k(n)$, satisfies $|\mathcal{F}_k(n)| = (2k)^n\,C_{n-1}$; hence

$$\frac{|\mathcal{S}_k^{(1)}(n)|}{|\mathcal{F}_k(n)|} = \frac{8k+1}{(4k+1)^2}\,(1+O(1/n)).$$

Hence the density of $\mathcal{S}_k^{(1)}$ does exist, and we know its value; moreover we have an asymptotic expression of this density when k in turn becomes large:

$$\mu_k(\mathcal{S}_k^{(1)}) = \frac{8k+1}{(4k+1)^2} = \frac{1}{2k} + O\left(\frac{1}{k^2}\right).$$

□

We now turn to simple tautologies of the second kind. Let us first give some definitions. For a tree A, let $\mathcal{G}(A)$ be the multiset containing the labels of the goal and the subgoals of A; that is, $\mathcal{G}(A_1, \ldots, A_p \to \alpha) = \{\{r(A_1), \ldots, r(A_p), \alpha\}\}$. A variable $x \in \{x_1, \ldots, x_k\}$ is said to have r *repetitions* in a multiset of literals M if at least $r+1$ elements of M belong to $\{x, \bar{x}\}$. We define $\mathcal{R}$ to be the set of trees $A \in \mathcal{F}_k$ such that $\mathcal{G}(A)$ has two repetitions of a variable, or one repetition of two distinct variables.

Lemma 7 *It holds that* $\mu_k^+(\mathcal{R}) = O(1/k^2)$.

Proof: For two generating functions f and g, we shall write $f \prec g$ if $[z^n]f(z) \leqslant [z^n]g(z)$ for all $n \in \mathbb{N}$. For $x \in \{x_1, \ldots, x_k\}$, let $\mathcal{R}^x$ be the set of formulae $A \in \mathcal{F}_k$ such that x has two repetitions in $\mathcal{G}(A)$. By considering whether the repeated variable appears in three subgoals or two subgoals and the goal, the generating function $\phi_{\mathcal{R}^x}$ is easily seen to satisfy

$$\phi_{\mathcal{R}^x}(z) \prec \left(\frac{f_k(z)/k}{1-f_k(z)}\right)^3 \cdot f_k(z) + \left(\frac{f_k(z)/k}{1-f_k(z)}\right)^2 \frac{2z}{1-f_k(z)}.$$

Easy calculations yield $\mu^+(\mathcal{R}^x) = O(1/k^3)$. In the same way, for two distinct variables $x \neq y$, the set $\mathcal{R}^{x,y}$ of trees $A \in \mathcal{F}_k$ such that both x and y are repeated in $\mathcal{G}(A)$ satisfies $\mu^+(\mathcal{R}^{x,y}) = O(1/k^4)$. Since $\mathcal{R} = \bigcup_x \mathcal{R}^x \cup \bigcup_{x \neq y} \mathcal{R}^{x,y}$, it follows that $\mu^+(\mathcal{R}) \leqslant k\ O(1/k^3) + \binom{k}{2}\ O(1/k^4) = O(1/k^2)$. □

Lemma 8 *The density of simple tautologies of the second kind is asymptotically equal to*

$$\mu_k(\mathcal{S}_k^{(2)}) = \frac{3}{8k} + O\left(\frac{1}{k^2}\right).$$

Proof: For a tree A, we define $\mathcal{U}(A)$ to be the multiset of the goal of A and the labels of the premises of size 1 of A. That is,

$$\mathcal{U}(A_1, \ldots, A_p \to \alpha) = \{\{A_i \ : \ \|A_i\| = 1\}\} \cup \{\{\alpha\}\}.$$

Let $\tilde{\mathcal{S}}_k^{(2)}$ be the set of trees $A \in \mathcal{F}_k$ such that two premises are opposite literals, and no other repetition occurs in $\mathcal{U}(A)$. More precisely, $(A_1, \ldots, A_p \to \alpha) \in \tilde{\mathcal{S}}_k^{(2)}$ if there exists a literal ℓ and $1 \leqslant i < j \leqslant p$ such that $A_i = \ell$, $A_j = \bar{\ell}$, and there is no repetition in $\mathcal{U}(A) \setminus \{\{\bar{\ell}\}\}$. Notice that $\tilde{\mathcal{S}}_k^{(2)} \subseteq \mathcal{S}_k^{(2)} \setminus \mathcal{S}_k^{(1)}$.

Let $T_{\ell,\alpha,t}$ be the set of expressions $A \in \tilde{\mathcal{S}}_k^{(2)}$ satisfying the following conditions:

- ℓ and $\bar{\ell}$ appear exactly once as a premise in A, in this order;
- α is the goal of A;

- A has exactly $t+2$ premises of size 1.

The subset $T_{\ell,\alpha,t}[y_1,\dots,y_{t+2}]$ of $T_{\ell,\alpha,t}$ is defined as the set of expressions such that y_i is the ith premise among those equal to a literal (for all $1 \leqslant i \leqslant t+2$).

Let $\mathcal{E}_L$ be the set of all expressions differing from a literal, i.e. of size at least 2. The set $T_{\ell,\alpha,t}[y_1,\dots,y_{t+2}]$ is described by the following non-ambiguous expression:

$$\mathcal{E}_L^* \cdot y_1 \cdot \ldots \cdot \mathcal{E}_L^* \cdot y_{t+2} \cdot \mathcal{E}_L^* \cdot \alpha.$$

We can now enumerate this set. The generating function for $\mathcal{E}_L$ is $f_k(z) - 2kz$; using the equation that defines f_k (see Proposition 3), we can write it as $f_k(z)^2$. Consequently the generating function for $T_{\ell,\alpha,t}[y_1,\dots,y_{t+2}]$ is:

$$\sigma_2(z) = \left(\frac{z}{1-f_k(z)^2}\right)^{t+3}.$$

Since

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

we can rewrite $\sigma_2(z)$ as

$$\sigma_2(z) = \left(\frac{1}{8k+8k^2z}\right)^{t+3}\left(P(z) - \sqrt{1-8kz}\sum_{s=0}^{\lfloor\frac{t+2}{2}\rfloor}\binom{t+3}{2s+1}(1-8kz)^s(1+4kz)^{t-2s+2}\right),$$

where $P(z)$ is an adequate polynomial. The dominant singularity of σ_2 is $1/(8k)$. Using the transfer theorems [3, Chapter VI], the asymptotic behaviour of $[z^n]\sigma_2$ is given by the first term of the sum ($s=0$), and we conclude

$$[z^n]\sigma_2(z) = \frac{-(t+3)(3/2)^{t+2}}{(9k)^{t+3}}[z^n]\{\sqrt{1-8kz}\}\ (1+O(1/n)).$$

As in the proof of Lemma 6, we obtain that the density of $T_{\ell,\alpha,t}[y_1,\dots,y_{t+2}]$ is well-defined, and that its value is equal to $\mu_k(T_{\ell,\alpha,t}[y_1,\dots,y_{t+2}]) = (4/3)(t+3)/(6k)^{t+3} =: \theta_t$.

Since the sets $T_{\ell,\alpha,t}[y_1,\dots,y_{t+2}]$ (for all $0 \leqslant t \leqslant k-2$ and all $(\alpha,\ell,y_1,\dots,y_{t+2})$ satisfying the right conditions) form a partition of $\tilde{\mathcal{S}}_k^{(2)}$, the density of $\tilde{\mathcal{S}}_k^{(2)}$ exists and can be computed as follows. For a given $t \in \{0,\dots,k-2\}$, a class $T_{\ell,\alpha,t}[y_1,\dots,y_{t+2}]$ is obtained by choosing the literal ℓ and the positions of ℓ and $\bar{\ell}$ in the sequence of premises of size 1, then the $t+1$ literals of $\alpha, y_1,\dots,y_{t+2}$ different from ℓ and $\bar{\ell}$. Consequently the density of $\tilde{\mathcal{S}}_k^{(2)}$ is

$$\begin{aligned}\mu_k(\tilde{\mathcal{S}}_k^{(2)}) &= \sum_{t=0}^{k-2} 2k\binom{t+2}{2}\cdot(k-1)\ldots(k-t-1)\cdot 2^{t+1}\cdot\theta_t \\ &= \sum_{t=0}^{k-2}\frac{(t+3)(t+2)(t+1)}{3^{t+4}k}\cdot\frac{(k-1)\ldots(k-(t+1))}{k^{t+1}}.\end{aligned}$$

We need to estimate the asymptotic behaviour of the above sum. Using $(k-1)\ldots(k-t-1)/k^{t+1} \leqslant 1$ gives the upper bound $\mu_k(\tilde{\mathcal{S}}_k^{(2)}) \leqslant 3/(8k)$. On the other hand, $(k-1)\ldots(k-$

$t-1)/k^{t+1} \geqslant ((k-t-1)/k)^{t+1} = (1-(t+1)/k)^{t+1} \geqslant 1-(t+1)^2/k$ gives the lower bound $\mu_k(\tilde{\mathcal{S}}_k^{(2)}) \geqslant 3/(8k) + O(1/k^2)$. We conclude that

$$\mu_k(\tilde{\mathcal{S}}_k^{(2)}) = \frac{3}{8k} + O\left(\frac{1}{k^2}\right).$$

In the same way as we did for $\tilde{\mathcal{S}}_k^{(2)}$, we can partition $\mathcal{S}_k^{(2)} \setminus (\mathcal{S}_k^{(1)} \cup \tilde{\mathcal{S}}_k^{(2)})$ into a finite number of sets (depending on k), all of them having a well-defined density, by considering the sequence of premises of size 1 of a tree, truncated at some suitable position. Thus the set $\mathcal{S}_k^{(2)} \setminus (\mathcal{S}_k^{(1)} \cup \tilde{\mathcal{S}}_k^{(2)})$ also has a well-defined density. This shows that the density of $\mathcal{S}_k^{(2)} \setminus \mathcal{S}_k^{(1)}$ exists; with Lemma 6, it follows that $\mu_k(\mathcal{S}_k^{(2)})$ is well-defined.

In order to estimate the density of $\mathcal{S}_k^{(2)}$, we take a shorter route: Since $\mathcal{S}_k^{(2)} \setminus (\mathcal{S}_k^{(1)} \cup \tilde{\mathcal{S}}_k^{(2)}) \subseteq \mathcal{R}$, it follows from Lemma 7 that its density is equal to $O(1/k^2)$. This shows that $\mu_k(\mathcal{S}_k^{(2)}) = 3/(8k) + O(1/k^2)$. $\square$

It is now an easy matter to prove the existence of the density for the whole set of simple tautologies.

Lemma 9 *The density of simple tautologies is asymptotically equal to* $\mu_k(\mathcal{S}_k) = 7/(8k) + O(1/k^2)$.

Proof: It is shown in the proof of Lemma 8 that $\mathcal{S}_k^{(2)} \setminus \mathcal{S}_k^{(1)}$ has a well-defined density, equal to $3/(8k) + O(1/k^2)$. Together with Lemma 6, this shows the lemma. $\square$

The last step toward the main result is to estimate the density of non-simple tautologies.

Lemma 10 *The density of non-simple tautologies is asymptotically equal to*

$$\mu_k(\mathcal{T}_k \setminus \mathcal{S}_k) = O\left(\frac{1}{k^2}\right).$$

Proof: The densities of both $\mathcal{T}_k$ and $\mathcal{S}_k$ exist by Lemmas 4 and 9; since $\mathcal{S}_k \subseteq \mathcal{T}_k$, the density of $\mathcal{T}_k \setminus \mathcal{S}_k$ is also well-defined. We next turn to its asymptotic evaluation. Let $A_1, \ldots, A_p \to \ell$ be a tautology. Let $\alpha_i = r(A_i)$. Necessarily $\bar{\alpha}_1 \vee \ldots \vee \bar{\alpha}_p \vee \ell$ computes *True*. Thus, there exists i such that $\alpha_i = \ell$ or there exist $i \neq j$ such that $\alpha_i = \bar{\alpha}_j$, and we can assume this is the only repetition of variable in the multiset $\mathcal{G}(A) = \{\{\alpha_1, \ldots, \alpha_p, \ell\}\}$. Indeed, by Lemma 7, $\mu^+(\mathcal{R}) = O(1/k^2)$. Hence the lemma follows from the following two claims.

Claim 1. The set $\mathcal{N}_1$ of all expressions $A \in \mathcal{T}_k \setminus (\mathcal{S}_k \cup \mathcal{R})$ such that one subgoal of A is equal to $r(A)$ satisfies $\mu^+(\mathcal{N}_1) = O(1/k^2)$.

Claim 2. The set $\mathcal{N}_2$ of all expressions $A \in \mathcal{T}_k \setminus (\mathcal{S}_k \cup \mathcal{R})$ such that two subgoals of A are opposite literals satisfies $\mu^+(\mathcal{N}_2) = O(1/k^2)$.

Let us prove *Claim 1.* Consider an expression $(A_1, \ldots, A_p \to \ell) \in \mathcal{T}_k \setminus (\mathcal{S}_k \cup \mathcal{R})$. Let $\alpha_i = r(A_i)$. There is a unique i_0 such that $\alpha_{i_0} = \ell$; moreover, $A_{i_0} \neq \ell$. Let $B = A_{i_0}$. The tree B is of the form $B_1, \ldots, B_q \to \ell$ with $q \geqslant 1$. Let $C = B_1$; once again we can develop it and write $C = C_1, \ldots, C_s \to \gamma$. Let $\gamma = r(C)$ and $\gamma_i = r(C_i)$ – note that we do not assume

$s \geqslant 1$, i.e. C could be reduced to the leaf γ (corresponding to the case $s = 0$). As the tree A is a tautology, it computes *True*. Let us define some conditions on A. The tree is logically equivalent to $\bigvee_i \bar{A}_i \vee \ell = \bar{B} \vee \bigvee_{i \neq i_0} (\bigwedge_j A_{i,j} \wedge \bar{\alpha}_i) \vee \ell$. By expanding the disjunction, we get the necessary condition that

$$\bar{B} \vee \bigvee_{1 \leqslant i \leqslant p,\ i \neq i_0} \bar{\alpha}_i \vee \ell$$

must compute to *True*. Expanding the premise B, we obtain that

$$\bigvee_{1 \leqslant i \leqslant s} \bar{\gamma}_i \vee \gamma \vee \bigvee_{1 \leqslant i \leqslant p,\ i \neq i_0} \bar{\alpha}_i \vee \ell \tag{1}$$

must evaluate to *True*.

Let $\mathcal{N}'_1 \subseteq \mathcal{N}_1$ be the set of trees such that $\gamma \in \{\ell, \bar{\ell}\}$. It is easy to show that $\mu_k^+(\mathcal{N}'_1) = O(1/k^2)$. We now focus on $\mathcal{N}''_1 := \mathcal{N}_1 \setminus \mathcal{N}'_1$. We shall compute some formal power series $a(z)$ such that $|\{A \in \mathcal{N}''_1 \ :\ \|A\| = n\}| \leqslant [z^n]a(z)$ for all $n \in \mathbb{N}$, in order to get an upper bound on $\mu_k^+(\mathcal{N}''_1)$.

First let us fix p, s, ℓ and γ. In the trees of $\mathcal{N}''_1$, the number of possible subtrees C is bounded by the number of trees counted by the following generating function

$$c_{p,s}(z) = \left(\frac{1}{1 - (f_k(z) - z)} \right)^p \cdot z \cdot n_{p,s}$$

where $n_{p,s}$ is an integer to be defined precisely later. Now, the number of subtrees B is bounded by the generating function

$$b_{p,s}(z) = c_{p,s}(z) \cdot \frac{1}{1 - f_k(z)} \cdot z,$$

and the total number of trees A for some fixed p, s, ℓ, γ is bounded by

$$a_{p,s}(z) = p \cdot \left(\frac{1}{1 - (f_k(z) - z)} \right)^{p-1} \cdot z \cdot b_{p,s}(z).$$

From Expression (1) we get that (at least) one of the α_i or γ_i is equal to $\bar{\ell}$ or $\bar{\gamma}$, or (at least) one couple of literals among $\{\alpha_i \mid 1 \leqslant i \leqslant p, i \neq i_0\} \cup \{\gamma_i \mid 1 \leqslant i \leqslant s\}$ are equal. Thus we choose

$$n_{p,s} = 2(2k)^{p-2+s}(p - 1 + s) + 2k(2k)^{p-3+s} \binom{p-1+s}{2}.$$

It remains to define

$$a(z) = 2k(2k-2) \sum_{s=0}^{\infty} \sum_{p=1}^{\infty} a_{p,s}(z)$$

where $2k$ corresponds to the choice of ℓ and $2k - 2$ to the choice of γ. Hence the generating series of $\mathcal{N}''_1$ satisfies $\phi_{\mathcal{N}''_1} \prec a$. An easy computation on the power series $a(z)$ now shows that $\mu_k^+(\mathcal{N}''_1) = O(1/k^2)$. This ends the proof of Claim 1. Proof of Claim 2 is very similar and left to the reader. $\square$

Theorem 11 *In the model of trees over implication and both positive and negative literals, asymptotically (when the number of variables tends to infinity) all tautologies are simple, i.e.*

$$\lim_{k\to\infty} \frac{\mu_k(\mathcal{S}_k)}{\mu_k(\mathcal{T}_k)} = 1.$$

Moreover, the density of tautologies amongst all expressions is equal to $\mu_k(\mathcal{T}_k) = 7/(8k) + O(1/k^2)$.

Proof: The proof follows directly from Lemmas 9 and 10. □

5 Final remarks

We have shown that asymptotically, all tautologies over implication are simple, i.e. either one of the premises is equal to the goal, or two of their premises are opposite literals. Let us mention here that a similar result holds for expressions built from implication and positive literals only. In this restricted context, the set of simple tautologies is simply $\mathcal{S}_k^{(1)}$, the set of tautologies with one premise equal to the goal. This characterisation furthermore enabled us to prove that asymptotically, in the fragment of implication, all tautologies are intuitionistic ones [4].

Acknowledgements. We are grateful to Brigitte Chauvin for fruitful discussions about this problem.

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 A. M.Odlyzko. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, 3(2): 216–240, 1990.

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

[4] H. Fournier, D. Gardy, A. Genitrini and M. Zaionc. Classical and intuitionistic logic are asymptotically identical. In Proc. *Internat. Conf. on Computer Science Logic (CSL)*, LNCS Vol. 4646: 177–193, 2007.

[5] D. Gardy. Random Boolean expressions. *Colloquium on Computational Logic and Applications*, Chambéry (France), June 2005. Proceedings in *Discrete Mathematics and Theoretical Computer Science*, AF: 1–36, 2006.

[6] D. Gardy and A. Woods. And/Or tree probabilities of Boolean function. First Colloquium on the Analysis of Algorithms, Barcelona, 2005. Proceedings in *Discrete Mathematics and Theoretical Computer Science*, AD: 139–146, 2005.

[7] A. Genitrini and J. Kozik. Quantitative comparison of Intuitionistic and Classical logics – full propositional system. *International Symposium on Logical Foundations of Computer Science*, Florida (USA), Jan. 2009. Proceedings LNCS Vol. 5407: 280–294.

[8] A. Genitrini, J. Kozik and G. Matecki. On the density and the structure of the Peirce-like formulae. In Proc. *Fifth Colloquium on Mathematics and Computer Science*, Blaubeuren, Germany, September 2008. DMTCS Proceedings.

[9] A. Genitrini, J. Kozik and M. Zaionc. Intuitionistic vs classical tautologies, quantitative comparison. In Proc. *TYPES 2007*, LNCS Vol. 4941: 100–109. Springer, 2008.

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

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

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

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

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

[15] 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.

[16] R. Statman. On the Existence of Closed Terms in the Typed Lambda Calculus I. *To H. B. Curry: essays on combinatory logic, lambda calculus and formalism*: 511–534, Academic Press, 1980.

[17] H. Wilf. *Generatingfunctionology.* Second edition, Academic Press, Boston, 1994.

[18] A. Woods. On the probability of absolute truth for And/Or formulas. *The Bulletin of Symbolic Logic*, 12(3): 523, 2006.

[19] M. Zaionc. On the asymptotic density of tautologies in logic of implication and negation. *Reports on Mathematical Logic*, 39: 67–87, 2005.

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