In the full propositional logic, 5/8 of classical tautologies are intuitionistically valid.*

Antoine Genitrini[†] Jakub Kozik[‡]

October 9, 2012

Abstract

We present a quantitative comparison of classical and intuitionistic logics, based on the notion of density, within the framework of several propositional languages. In the most general case - the language of the "full propositional system" - we prove that the fraction of intuitionistic tautologies among classical tautologies of size n tends to 5/8 when n goes to infinity. We apply two approaches, one with a bounded number of variables, and another, in which formulae are considered "up to the names of variables". In both cases, we obtain the same results. Our results for both approaches are derived in a unified way based on structural properties of formulae. As a by-product of these considerations we present a characterization of the structures of almost all random tautologies.

Keywords: Propositional formulae, Tautologies, Asymptotic density, Intuitionistic logic, Analytic combinatorics.

1 Introduction

One of the first papers to address the quantitative aspects of intuitionistic logic was [8]. According to the authors their work was partially motivated by the short note in some paper of Statman saying: "*It is a good bet but not a sure thing, that ρ (type) contains a closed term*". The exact meaning of this sentence is not obvious. The set of types is countable, and it is impossible to have an uniformly distributed probabilistic measure on it. It is a standard approach to use the notion of the asymptotic density in a situation like this. The general idea is to consider the subsets of elements of bounded size, and observe the uniform measure of one subset in the other when the maximal allowed size tends to infinity. This approach requires that the number of elements of bounded size is finite. This assumption is not easy to be satisfied for propositional logic formulae, since we usually assume that the number of variables is infinite. We analyse two approaches, which for implicational formulae

*Research described in this paper was partially supported by the A.N.R. projects *SADA* and *BOOLE* (contract reference ANR-09-BLAN0011), by POLONIUM grant *Quantitative research in logic and functional languages*, cooperation between Jagiellonian University of Krakow, L' École Normale Supérieure de Lyon and Université de Versailles Saint-Quentin (contract number 7087/R07/R08), and by grant number N206 376137, Polish Ministry of Science and Higher Education.

[†]Laboratoire LIP6, CNRS UMR 7606 and Université Pierre et Marie Curie, 4 place Jussieu, 75252 Paris cedex 05, France. Email: antoine.genitrini@lip6.fr

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

have been considered in [3] and [5], obtaining the same results for both of them - "In the full propositional logic, 5/8 of classical tautologies are intuitionistically valid. "

The results presented in [8] were formulated in terms of inhabitation of types in simple λ - calculus. The authors of [8] considered calculus with a finite number of ground types and only functional types. In terms of logical formulae it corresponds to the situation when the number of different variables in a formula is bounded by some constant, and an only connective $\Rightarrow$ is used. Although formulated in terms of type inhabitation, these results can be translated to the language of propositional logic by Curry-Howard isomorphism (see e.g. [10]). The authors proved that, for any finite, fixed number of available variables, at least 1/3 of classical tautologies are intuitionistic and gave some lower and upper bounds (dependent on the number of allowed variables) for the density of intuitionistic tautologies among all the formulae. They also stated a conjecture saying that among the formulae with a number of different variables bounded by any constant the probability that a classical tautology of size n chosen uniformly at random is intuitionistic tends to one, when n goes to infinity. The conjecture turned out to be false, nevertheless its slight reformulation has been proved to be true in [3]. The authors of [3] showed that the lower bound for the density of intuitionistic logic in the classical one tends to 1, when the number of allowed variables tends to infinity. It can be argued that the approach using a bounded number of variables is not appropriate - we do not expect that "typical" formula of large size have a small number of variables. In the paper [5] authors suggested another approach, in which formulae were considered up to the names of variables (i.e. two formulae which differ only in the naming of variables were assumed to be the same). In that case one can deal with formulae with an unbounded number of variables, while preserving the property that there is only a finite number of formulae of bounded size. In that setup, using methods similar as in [3] the authors obtained an analogous result - the density is equal to 1. We want to emphasize at this point that the fact that both results coincide is, in our opinion, no less surprising that the fact that the density tends to 1.

The work presented in this paper is a continuation of this research, considering other languages of propositional formulae. Among them the most interesting is the language which admits all the usual connectives $\Rightarrow, \wedge, \vee$, and the constant $\perp$. We prove that in this case the coherence of the results in both approaches is preserved, even though the limit is no longer equal to 1, but to 5/8. We obtain these results by defining large families of intuitionistic (resp. classical) tautologies with simple structures, which together have asymptotic density 1 in the set of all intuitionistic (resp. classical) tautologies.

Similar research has been done for the random generation of so called *And/Or trees* (see [1] and [4] for a detailed survey). Especially the method of "subcritical pattern languages" presented in [7] is very close to the development we preset in this paper.

The paper is organized as follows. Section 2 recalls some definitions on intuitionistic logic and states the main results of this paper. Section 3 is devoted to the analysis of the structures of some tautologies. In Section 4 we prove that almost all tautologies in both models (with a bounded and an unbounded number of variables) have these structures. Then we prove our main results for the language of the full propositional system. Finally, in the last section we summarize briefly our results for other propositional languages.

2 Prerequisites and results

For any set A and $n \in \mathbb{N}$ we denote by $A(n)$ the number of elements of set A with size n (when the size is well defined for the elements of A).

2.1 Formulae and logics

Let $\mathrm{Var} = \{x_1, x_2, x_3, \ldots\}$ be a countable set of variables, $\perp$ be a constant, and $\mathcal{C} = \{\Rightarrow, \vee, \wedge\}$ be a set of binary connectives. A term in our system is a binary complete tree with internal nodes labelled by the elements of $\mathcal{C}$ and leaves by the elements of $\mathrm{Var} \cup \{\perp\}$ (precisely the tree is rooted and plane i.e. the order of descendants matters). For every $k \in \mathbb{N}$, let $\mathcal{F}_k$ denote the set of terms in which all variables belong to the set $\mathrm{Var}_k = \{x_1, \ldots, x_k\}$. The set of all terms is denoted by Term. The *size* of a term is its number of leaves.

Two terms are *α-equivalent* if they differ only in the naming of variables, i.e. $(\varphi, \psi) \in \alpha$ if there exists an injective relabelling function $r : \mathrm{Var} \to \mathrm{Var}$, such that we obtain ψ after relabelling variables from φ according to r. Clearly, α is an equivalence relation on Term. We denote Term/α by $\mathcal{F}_\infty$. We use the name formula both for terms and for elements from $\mathcal{F}_\infty$.

2.1.1 Intuitionistic logic

For the general reference about intuitionistic logic we suggest [10]. In order to keep the exposition as elementary as possible we recall here some simple (but quite specific) definitions.

Let $\mathcal{O}(\mathbb{R})$ denote the set of the open subsets of $\mathbb{R}$ with respect to the euclidean topology. A valuation in $\mathcal{O}(\mathbb{R})$ is any function $v : \mathrm{Var} \to \mathcal{O}(\mathbb{R})$. We call the valuation of the type $v : \mathrm{Var} \to \mathcal{O}(\mathbb{R})$ *an intuitionistic valuation*, for short. For every intuitionistic valuation we define the function $[[\cdot]]^I_v : \mathrm{Term} \to \mathcal{O}(\mathbb{R})$ called intuitionistic interpretation in $\mathcal{O}(\mathbb{R})$ recursively as follows

- $[[x]]^I_v = v(x)$, if x is a variable,
- $[[\perp]]^I_v = \emptyset$,
- $[[\psi_1 \vee \psi_2]]^I_v = [[\psi_1]]^I_v \cup [[\psi_2]]^I_v$,
- $[[\psi_1 \wedge \psi_2]]^I_v = [[\psi_1]]^I_v \cap [[\psi_2]]^I_v$,
- $[[\psi_1 \Rightarrow \psi_2]]^I_v = \mathit{Interior}((\mathbb{R} \setminus [[\psi_1]]^I_v) \cup [[\psi_2]]^I_v)$.

If the interpretation of some formula does not depend on the valuation of the variables, we omit the subscript v (e.g. we write $[[\perp \wedge \perp]]^I = \emptyset$).

Definition 1 *Formula $\varphi \in \mathrm{Term}$ is an intuitionistic tautology if and only if*

$$[[\varphi]]^I_v = \mathbb{R}$$

for every valuation $v : \mathrm{Var} \to \mathcal{O}(\mathbb{R})$.

The definition above can be naturally extended to the set $\mathcal{F}_\infty$: an element $\varphi \in \mathcal{F}_\infty$ is an intuitionistic tautology if all (or equivalently some of) its representatives are.

Observation 1 *The interpretation of the implication is an inflation with respect to the second variable. I.e. for every valuation v we have:*

$$[[\psi_1 \Rightarrow \psi_2]]^I_v \supset [[\psi_2]]^I_v.$$

A *Boolean valuation* is any function $v : \mathrm{Var} \to \{True, False\}$. The Boolean interpretation for a Boolean valuation is a function $[[\cdot]]^C_v : \mathrm{Term} \to \{True, False\}$; its definition is straightforward ($\bot$ is interpreted as $False$). Let us note the fact which belongs to folklore (and can be easily derived from the presented definitions):

Observation 2 *Every intuitionistic tautology is classical.*

It is also a classical result that the converse is not true. The famous example of a tautology which is classical but not intuitionistic is $((p \Rightarrow q) \Rightarrow p) \Rightarrow p$ and is known as the Peirce's law. Indeed, choosing an intuitionistic valuation v such that $v(p) = \mathbb{R} \setminus \{0\}$ and $v(q) = \emptyset$ gives $[[((p \Rightarrow q) \Rightarrow p) \Rightarrow p]]^I_v = \mathbb{R} \setminus \{0\}$. The same valuation can be used to show that the law of excluded middle $p \vee (p \Rightarrow \bot)$ is not an intuitionistic tautology.

2.2 Main results

Let $Cl, Int \subset \mathrm{Term}$ denote the sets of terms which are respectively classical and intuitionistic tautologies. For every $k \in \mathbb{N}$ we put

$$Cl_k = Cl \cap \mathrm{Term}_k, \qquad Int_k = Int \cap \mathrm{Term}_k$$

and

$$Cl_\infty = Cl/\alpha, \qquad Int_\infty = Int/\alpha.$$

Let a sequence $(d_k(n))_{n \in \mathbb{N} \setminus 0,1}$ be defined as follows:

$$d_k(n) = \frac{Int_k(n)}{Cl_k(n)}.$$

Each fraction $d_k(n)$ equals the probability that a formula chosen uniformly at random among the set Cl_k of size n is an intuitionistic tautology. Note, that the sequence is well-defined, since there are classical tautologies of any size not smaller than 2. If the sequence converges, its limit is denoted by D_k and is called the relative density of Int_k in Cl_k. We do not address the problem of the existence of D_k. We use following bounds instead:

$$D_k^- = \liminf_{n \to \infty} d_k(n), \qquad D_k^+ = \limsup_{n \to \infty} d_k(n).$$

The first of our main results says that

$$\lim_{k \to \infty} D_k^- = \lim_{k \to \infty} D_k^+ = \frac{5}{8}.$$

This is analogous to the approach taken in [3] for the implicational fragment. In that case the limit was 1.

Considering the formulae "up to the names of variables" enables an arbitrary number of different variables in formula, while preserving the property that there is only a finite number of formulae with bounded size. In this approach we consider the sequence $(d_\infty(n))_{n\in\mathbb{N}}$ defined as follows:

$$d_\infty(n) = \frac{Int_\infty(n)}{Cl_\infty(n)}.$$

The second of our main results says that

$$\lim_{n\to\infty} d_\infty(n) = \frac{5}{8}.$$

We could give an informal interpretation that "about $\frac{5}{8}$ of classical tautologies are intuitionistic". It was proved in [5] that the analogous approach for the implicational fragments gives the density 1.

2.3 Structure and labelling

For a term φ, *the structure of* φ is a binary tree constructed from φ by forgetting about the labelling of its leaves (e.g. by changing it so that each leaf is labelled by $\bullet$). The definition can be naturally extended to the formulae from $\mathcal{F}_\infty$, since all the terms in each equivalence class have the same structure. The set of structures in our system is denoted by $\mathcal{T}$. It is the set of binary, complete trees with internal nodes labelled by $\Rightarrow, \wedge$ or $\vee$ and all leaves labelled by $\bullet$.

We say that a node is an $\Rightarrow$-node if the node is labelled with $\Rightarrow$. We use an analogous convention for the other connectives.

For a formula $\varphi \in \mathcal{F}_k$ with n leaves, a leaf labelling of φ is a function $f : \{1, \ldots, n\} \to \mathrm{Var}_k \cup \{\bot\}$ such that $f(i)$ coincides with the label at the i-th leaf of φ. We call such a function a k-labelling of size n.

For a formula $[\varphi] \in \mathcal{F}_\infty$ with n leaves, a leaf labelling of $[\varphi]$ is the equivalence relation R on the set $\{0, 1, \ldots, n\}$ consisting of all the pairs of numbers of leaves which are labelled by the same symbol (variable or $\bot$) and all the pairs $(0, j), (j, 0)$ for each leaf j labelled with $\bot$. Note that the relation R does not depend on the chosen representative of the equivalence class $[\varphi]$. It contains information about which leaves are labelled by the same variable (but not by which variable), and which leaves are labelled with $\bot$. We call such a relation the ∞-labelling of size n.

As usual, the size of a structure is the number of its leaves, and we denote by $\mathcal{T}(n)$ the number of structures from $\mathcal{T}$ of size n.

In all considered cases (bounded for every $k \in \mathbb{N}$ and unbounded) we have a one-to-one correspondence between the structure-labelling pairs of the size n and the formulae of that size. This fact is reflected in simple expressions for the numbers of formulae of size n. We have

$$\mathcal{F}_k(n) = \mathcal{T}(n) \cdot (k+1)^n, \qquad \mathcal{F}_\infty(n) = \mathcal{T}(n) \cdot B(n+1), \tag{1}$$

where $B(n+1)$ is the number of equivalence relations on the set $\{0, 1, \ldots, n\}$ known as Bell number (see e.g. [6]).

2.4 Generating functions

Within this paper we make an extensive use of the theory of generating functions and analytic combinatorics (see [2]). All the generating functions in this paper are ordinary.

We use a notation which always exposes the formal parameters of a generating function. E.g. we write $g(z)$ instead g for some generating function $\sum_{n\in\mathbb{N}} g_n z^n$. Although the notation may be a little bit misleading, it provides a convenient way of expressing substitutions for formal parameters (e.g. we have $g(y^2) = \sum_{n\in\mathbb{N}} g_n y^{2n}$). It is a standard convention to denote by $[z^n]g(z)$ the coefficient g_n (for the function $g(z)$ defined as above).

One of the most basic generating functions in this paper, is the one enumerating all the structures: $t(z) = \sum_{n\in\mathbb{N}} \mathcal{T}(n)z^n$. By a standard constructions we get an algebraic equation for $t(z)$:

$$t(z) = z + 3t(z)^2.$$

This equation reflects the fact that a structure is either a leaf (this case corresponds to the term z) or a tree with exactly two subtrees and root labelled by one of three connectives (term $3t(z)^2$). Solving this equation (and choosing the proper solution) we get

$$t(z) = \frac{1-\sqrt{1-12z}}{6}.$$

The radius of convergence of $t(z)$ is $\rho = 1/12$, $t(z)$ is bounded within its circle of convergence, and $t(\rho) = \lim_{z\to_{\mathbb{R}}\rho^-} t(z) = \frac{1}{6}$.

We use the following technical lemma.

Lemma 1 *Let $f, g \in \mathbb{Z}[[z]]$ be algebraic generating functions, having a common unique dominating singularity at $\varrho \in \mathbb{R}_+$. Suppose, that these functions have Puiseux expansions around ϱ of the form*

$$f(z) = c_f + d_f(z-\varrho)^{\frac{1}{2}} + o((z-\varrho)^{\frac{1}{2}})$$
$$g(z) = c_g + d_g(z-\varrho)^{\frac{1}{2}} + o((z-\varrho)^{\frac{1}{2}}),$$

with both d_f, d_g being nonzero. Then

$$\lim_{n\to\infty} \frac{[z^n]f(z)}{[z^n]g(z)} = \lim_{z\to_{\mathbb{R}}\varrho^-} \frac{f'(z)}{g'(z)}.$$

By singularity analysis for algebraic generating functions (see e.g. Theorem VII.8 from [2]) we obtain that:

$$\lim_{n\to\infty} \frac{[z^n]f(z)}{[z^n]g(z)} = \frac{d_f}{d_g}.$$

On the other hand it can be easily checked that $\lim_{z\to_{\mathbb{R}}\varrho^-} \frac{f'(z)}{g'(z)} = \frac{d_f}{d_g}$.

3 Structural properties of tautologies

In this section we analyse structural properties of tautologies. In order to obtain results independent of the kind of labelling, we use $\mathcal{F}$ to denote the set of formulae under consideration, and the function $Lab : \mathbb{N} \to \mathbb{N}$, whose value for n is the number of different labellings of the structures of size n. In particular we get results for the unbounded approach by setting $\mathcal{F}$ equal to $\mathcal{F}_\infty$ and $Lab(n)$ to $B(n+1)$. In an analogous way the results are translated to the bounded case for every fixed number of variables k by substituting $\mathcal{F}$ with $\mathcal{F}_k$ and $Lab(n)$ with $(k+1)^n$. E.g. in this convention equations (1) are formulated as

$$\mathcal{F}(n) = \mathcal{T}(n) \cdot Lab(n).$$

3.1 Pointed structures

An m-pointed structure is a pair (t, s) of a structure t and a sequence s of m different leaves of t. Usually we use a pointed structure to encode some constraints on the allowed labellings. For example let A denote some set of 1-pointed structures and consider the set of formulae $\mathcal{F}_A$, which can be constructed from the elements of A by labellings which assign $\perp$ to the pointed leaf. For every structure $a \in A$ of size n we are free to label all the remaining leaves. Therefore, there are $Lab(n-1)$ labellings which, together with the structure a, give a formula from $\mathcal{F}_A$. Therefore $\mathcal{F}_A(n) \leqslant A(n) \cdot Lab(n-1)$.

3.2 Tree decomposition

We say that a node v in a tree $t \in \mathcal{T}$ is *k-shallow* if the path from the root to v goes at most k times to the left from a node labelled with $\Rightarrow$ (i.e. it goes into the left subtree). A node is a *k-layer* node if it is k-shallow but not $(k-1)$-shallow.

To obtain an upper bound for the number of tautologies we focus on 3-shallow leaves.

Let us consider the set of trees $P \subset \mathcal{T}$ such that every left subtree of every node labelled with the connective $\Rightarrow$ is a leaf (i.e. all 1-layer nodes are leaves). Let $p(t, u)$ be the generating function for such trees with t marking leaves which are left sons of an $\Rightarrow$-node, and u marking the remaining leaves (t denotes a formal parameter, not the generating function for all structures, which we denote by $t(z)$). The generating function is given implicitly by the equation

$$p(t, u) = t \cdot p(t, u) + 2 \cdot p(t, u)^2 + u, \tag{2}$$

which, by standard combinatorial constructions ([2]), reflects the fact that every such a tree is

- either an implication with its left subtree being an 1-layer leaf and its right subtree belonging to P,
- or a conjunction or a disjunction with both subtrees belonging to P,
- or a leaf (which is in fact a 0-shallow leaf).

Clearly, $p(t(z), uz)$ is the generating function of all structures, with z marking the size and u marking 0-shallow leaves. We define a sequence of generating functions:

$$p_{\leqslant 0}(t, u) = t \qquad p_{\leqslant (n+1)}(t, u) = p(p_{\leqslant n}(t, u), u).$$

Each function $p_{\leqslant (n+1)}(t, u)$ is the generating function of the set of structures in which all $(n+1)$-layer nodes are leaves, with u marking n-shallow leaves, and t marking leaves which are left sons of n-layer $\Rightarrow$-nodes (i.e. all $(n+1)$-layer leaves). Since every node in every tree is an i-layer node for exactly one i, we get for every $n \in \mathbb{N}$

$$t(z) = p_{\leqslant n}(t(z), z).$$

Proposition 1 *For $s, m \in \mathbb{N}$ let $\mathcal{T}_{\leqslant s}^{(m)}$ denote the set of m-pointed structures with all pointed leaves being s-shallow (we call them s-shallow m-pointed structures). There exists a positive constant $c_{s,m} \in \mathbb{R}$ such that*

$$\lim_{n \to \infty} \frac{\mathcal{T}_{\leqslant s}^{(m)}(n)}{\mathcal{T}(n)} = c_{s,m}.$$

In fact we need this property only for the sets $\mathcal{T}^{(2)}_{\leqslant 3}, \mathcal{T}^{(3)}_{\leqslant 3}, \mathcal{T}^{(4)}_{\leqslant 3}$, and the results for these sets can be easily established by explicit calculations of their generating functions. Instead, we present a proof of the general case, which we believe is shorter and more interesting than algebraic computations.

Proof 1 *Solving the equation (2) and using the fact that* $p(0,0) = 0$ *we get*

$$p(t,u) = \frac{1}{4}(1 - t - \sqrt{(1-t)^2 - 8u}).$$

It shows that the function $p(t,u)$ *is analytic in the set* $D_\varepsilon = \{(t,u) \in \mathbb{C}^2 : |t| \leqslant \frac{1}{6} + \varepsilon, |u| \leqslant \frac{1}{12} + \varepsilon\}$ *for a sufficiently small* $\varepsilon \in \mathbb{R}$ *(note that* $t(\rho) = \frac{1}{6}$ *and* $\rho = \frac{1}{12}$*). By non-negativity of the coefficients of the expansion of* $p(t,u)$ *at 0 we get that* $\max_{(t,u)\in D_0} |p(t,u)| = p(\frac{1}{6}, \frac{1}{12}) = \frac{1}{6}$*. Therefore each* $p_{\leqslant s}(t,u)$ *is analytic in* D_ϵ *(for a sufficiently small, positive* $\epsilon \in \mathbb{R}$*) and so are all its partial derivatives, in particular* $u^m \frac{\partial^m p_{\leqslant s}(t,u)}{(\partial u)^m}$*. Let us observe, that since differentiation of generating function corresponds to pointing (see [2]), the latter function is exactly the generating function of* s*-shallow* m*-pointed structures in which all* $(m+1)$*-layer nodes are leaves (marked with variable* t*). It remains to substitute the generating function of all structures for* t *to obtain the generating function for all* s*-shallow* m*-pointed structures. We substitute* u *with* z *so that the variable* z *marks all leaves (after pointing, we are no longer interested in* s*-shallow leaves). As a result we obtain the following function*

$$p_{m,s}(z) = \left(u^m \frac{\partial^m p_{\leqslant s}(t,u)}{(\partial u)^m} \right) |_{u:=z, t:=t(z)},$$

which is the generating function of the set of all s*-shallow* m*-pointed structures. Let* $\widehat{D_\epsilon}$ *denote the set* $D_\epsilon \setminus [\rho, \infty]$*. Then the function* $t(z)$ *is analytically continuable to the set* $\widehat{D_\epsilon}$*, and since the outer function is analytic in* $\widehat{D_\epsilon}$ *we know that the function* $p_{m,s}(z)$ *is analytically continuable to that set. On the other hand the combinatorial interpretation shows that* $p_{m,s}(z)$ *must have a singularity in* ρ*. Therefore we know that* $p_{m,s}(z)$ *has a unique dominating singularity in* ρ*. In fact we know also that* $\lim_{z \to_{\mathbb{R}} \rho^-} p_{m,s}(z) < \infty$*, therefore the singularity is not a pole. Since* $p_{m,s}(z)$ *is algebraic, the singularity must be a branching point. By the fact that* $t(\rho - v^2)$ *is analytic at* ρ *we get that* $p_{m,s}(\rho - v^2)$ *is analytic as well, which shows that the branching type of* $p_{m,s}(z)$ *at* ρ *is 2 (we excluded the existence of pole). Finally, the fact that* $\lim_{z \to_{\mathbb{R}} \rho^-} p'_{m,s}(z) = \infty$ *shows that the singularity is of the square root type. A straightforward application of the Lemma 1 proves the result.*

3.3 Shallow repetitions

For every formula φ and set of its leaves L we say that φ has r repetitions among the leaves from L, if r equals the difference between the cardinality of L and the number of different variables assigned to the leaves from L. If the set L consists of all k-shallow leaves we say that φ has r k-shallow repetitions. Note that the occurrence of the constant is treated as repetition e.g. formula $(y \Rightarrow x) \Rightarrow (x \Rightarrow \bot)$ has two repetitions among all its leaves.

Proposition 2 *Within the set of elements of* $\mathcal{F}$ *of size* n*, the fraction of formulae with at least two 3-shallow repetitions is asymptotically bounded by* $c\frac{Lab(n-2)}{Lab(n)}$*. Formally, let* $\mathcal{F}^{[\geqslant 2]}_{\leqslant 3}$

denote the set of formulae with at least two 3-shallow repetitions, we have

$$\frac{\mathcal{F}^{[\geqslant 2]}_{\leqslant 3}(n)}{\mathcal{F}(n)} \lesssim c \cdot \frac{Lab(n-2)}{Lab(n)}$$

for some positive $c \in \mathbb{R}$.

Proof 2 *Every formula $\varphi \in \mathcal{F}^{[\geqslant 2]}_{\leqslant 3}$ satisfies at least one of the following properties:*

A *φ contains two 3-shallow leaves labelled with $\perp$,*

B *φ contains one 3-shallow leaf labelled with $\perp$ and two 3-shallow leaves labelled by the same variable,*

C *φ contains three 3-shallow leaves labelled by the same variable,*

D *two variables occur at least twice among 3-shallow leaves of φ.*

Let $\mathcal{F}^A, \mathcal{F}^B, \mathcal{F}^C, \mathcal{F}^D$ denote the sets of formulae with the previous properties. Clearly

$$\mathcal{F}^{[\geqslant 2]}_{\leqslant 3}(n) \leqslant \mathcal{F}^A(n) + \mathcal{F}^B(n) + \mathcal{F}^C(n) + \mathcal{F}^D(n),$$

and since the sets are not disjoint, when n is large enough, the inequality is usually strict.

Every formula from $\mathcal{F}^A$ contains at least two 3-shallow leaves labelled with $\perp$. Therefore all these formulae can be constructed from 3-shallow 2-pointed structures by labellings which assigns $\perp$ to the pointed leaves. Hence

$$\mathcal{F}^A(n) \leqslant \mathcal{T}^{(2)}_{\leqslant 3}(n) \cdot Lab(n-2).$$

An analogous reasoning for the other sets gives

$$\mathcal{F}^B(n) + \mathcal{F}^C(n) \leqslant 2 \cdot \mathcal{T}^{(3)}_{\leqslant 3}(n) \cdot Lab(n-2),$$

and

$$\mathcal{F}^D(n) \leqslant \mathcal{T}^{(4)}_{\leqslant 3}(n) \cdot Lab(n-2).$$

Using these equations and Proposition 1 we obtain

$$\frac{\mathcal{F}^{[2]}_{\leqslant 3}(n)}{\mathcal{F}(n)} \leqslant \frac{(\mathcal{T}^{(2)}_{\leqslant 3}(n) + 2\mathcal{T}^{(3)}_{\leqslant 3}(n) + \mathcal{T}^{(4)}_{\leqslant 3}(n))}{\mathcal{T}(n)} \cdot \frac{Lab(n-2)}{Lab(n)} \sim (c_{2,3} + 2 \cdot c_{3,3} + c_{4,3}) \cdot \frac{Lab(n-2)}{Lab(n)}$$

We use Proposition 2 to show that we can neglect all formulae with at least two 3-shallow repetitions, since, as we will prove in Section 4, the number of all such formulae is essentially smaller than the number of tautologies.

3.4 Shallow repetitions in Classical Tautologies

For a formula φ let a Boolean valuation v^1_φ assign $True$ only to those variables that have occurrences on the first layer, and let $v^{1,3}_\varphi$ assign $True$ only to those that have occurrences on the first or the third layer. The following proposition is a consequence of the fact that if there are no 1-shallow repetitions in φ, then the formula is valuated to $False$ by v^1_φ and to $True$ by the opposite valuation.

Proposition 3 *If a formula φ does not contain at least one 1-shallow repetition it does not define a constant function.*

3.4.1 Positive/Negative leaves

Definition 2 *A* positive path *in a formula (tree) is a path from the root to some node, which never crosses an $\wedge$-node, and never goes to the left subtree from an $\Rightarrow$-node. A node is called positive if there exists a positive path to it (see Figure 1).*

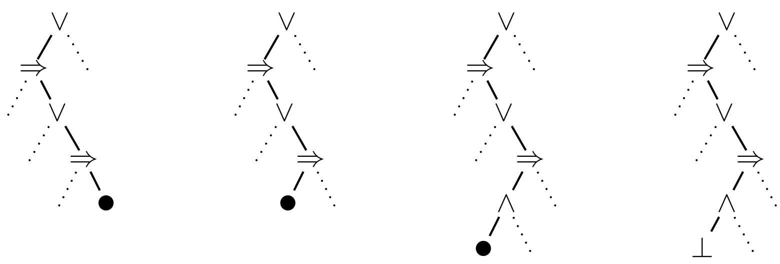

Figure 1: From left to right: a positive path, not a positive path, a negative path, a tree with a negative leaf labelled with $\perp$.

It is easy to observe that for every formula it is enough to valuate one of its positive nodes to $True$, to ensure that the valuation of the whole formula is $True$.

Definition 3 *A* negative path *in a formula (tree) is a path from the root, which contains a positive $\Rightarrow$-node h, such that the path is going to the left subtree from h and then follows only $\wedge$-nodes (if any). A node is called negative if there exists a negative path to it (see Figure 1).*

The motivation for the negative path is also straightforward - whenever some Boolean valuation assigns $False$ to a negative node, then the whole formula evaluates to $True$ (the last positive node on the negative path is of the form $\phi_l \Rightarrow \phi_r$ and we have a $\wedge$-path in ϕ_l to a node valuated to $False$, therefore ϕ_l is valuated to $False$ and hence $\phi_l \Rightarrow \phi_r$ to $True$, which is propagated along the positive path to the root).

Those two definitions give rise to two large families of classical tautologies.

Observation 3 *All the formulae in which some negative leaf is labelled with $\perp$ are* classical tautologies*. The set of these formulae is denoted by $S_\perp$ (see Figure 1).*

Observation 4 *All the formulae in which some positive leaf is labelled by the same variable as some negative leaf are classical tautologies. We denote this family by S_R.*

Proof 3 *The proof of the first observation is straightforward. For the proof of the second one, suppose that φ is a formula with a positive leaf and a negative leaf labelled with the same variable. For every Boolean valuation we have either a negative leaf valuated to $False$ or a positive leaf valuated to $True$. In both cases the whole formula φ is valuated to $True$, which proves that it is a classical tautology.*

We call the formulae from the set $S_R \cup S_\perp$, *simple tautologies.* We focus on the formulae with exactly one 1-shallow repetition and at the same time exactly one 3-shallow repetition (it means that taking into consideration also layers 2 and 3 does not increase the number of repetitions). The set of such formulae is denoted by $\mathcal{H}$. In the next two propositions we show that all tautologies belonging to $\mathcal{H}$ are simple.

Proposition 4 *If a formula $\varphi \in \mathcal{H} \setminus S_\perp$ contains a 3-shallow leaf l labelled with $\perp$, then it is not a tautology.*

Proof 4 *If the leaf l is not 1-shallow then there are no 1-shallow repetitions and the Boolean function defined by the formula is not constant (Proposition 3). If l is 0-shallow then we can use the valuation v^1_φ which valuates all the 0-shallow leaves to $False$ and all the 1-layer leaves to $True$. In that case the formula is valuated by v^1_φ to $False$.*

In the remaining case l is 1-layer leaf but is not negative. Let s be the last $\vee$-node or $\Rightarrow$-node on the path from the root to l. The node s is an 1-layer node, because l is not negative.

Suppose that s is labelled by $\vee$. One of its subtrees does not contain l. In that subtree all the 0-shallow leaves are valuated by $v^{1,3}_\varphi$ to $True$ (because they are all 1-layer leaves in φ) therefore the whole subtree with root s is valuated to $True$ by $v^{1,3}_\varphi$.

If s is labelled by $\Rightarrow$ then let s_2 be its left son. Clearly s_2 is a 2-layer node. Since we have only one 3-shallow repetition and it is realized by a 1-shallow node labelled with $\perp$, all the labels of 2-layer and 3-layer leaves are not repeated among 3-shallow leaves. Therefore the valuation $v^{1,3}_\varphi$ assigns $False$ to all the 2-layer leaves, and $True$ to all the 3-layer leaves. Consequently, every 2-layer node is valuated to $False$. It means that also s_2 is valuated to $False$, but then s is valuated to $True$.

In both cases the only 1-layer nodes which are valuated by $v^{1,3}_\varphi$ to $False$ are below the node s, which is valuated to $True$ anyway. Hence every 1-layer node which is a left son of a 0-shallow node is valuated to $True$. But then all 0-shallow nodes are valuated to $False$, which proves that φ is not a classical tautology.

Proposition 5 *If a formula $\varphi \in \mathcal{H} \setminus S_R$ contains a 3-shallow variable repetition, then it is not a tautology.*

Proof 5 *If φ does not contain any 1-shallow repetition, then according to the Proposition 3, the formula does not define a constant function. If both leaves with the repeated variable are on the same level, then the valuation v^1_φ valuates all the 0-shallow leaves to $False$ and all 1-layer leaves to $True$, and the formula is valuated by v^1_φ to $False$.*

Let l_1, l_2 be the 3-shallow leaves labelled with the same variable. We can assume that l_1 is 0-shallow and l_2 is a 1-layer leaf. If l_1 is not positive then there exists a node s on the path from the root to l_1, which is labelled with $\wedge$. In that case the only 0-shallow nodes which can be valuated to $True$ by v^1_φ are below s. But s is valuated to $False$, because it is a $\wedge$-node and one of its subtrees is valuated by v^1_φ to $False$ (the one which does not contain l_1).

In the remaining case we have two leaves l_1, l_2 labelled with the same variable, such that l_1 is positive (and hence 0-shallow), l_2 is not negative but is a 1-layer leaf. In this case we use the Boolean valuation b which assigns $False$ only to those variables which have occurrences among 0-shallow or 2-layer leaves. Then the leaf l_2 is valuated to $False$ and we can use the same reasoning as in the case when some not negative 1-layer leaves is labelled with $\perp$, to prove that φ is not a tautology.

Observation 5 *We have*

$$S_\perp(n) + S_R(n) - \mathcal{F}^{[\geqslant 2]}_{\leqslant 3}(n) \leqslant Cl(n) \leqslant S_\perp(n) + S_R(n) + \mathcal{F}^{[\geqslant 2]}_{\leqslant 3}(n) \tag{3}$$

The lower bound comes from the fact that every formula which belongs to $S_\perp \cap S_R$ has at least two 3-shallow repetitions. The upper bound is a consequence of Propositions 4 and 5, which together say that all tautologies which are not simple belong to $\mathcal{F}^{[\geqslant 2]}_{\leqslant 3}$.

3.5 Simple Intuitionistic Tautologies

It is easy to show that all the formulae from $S_\perp$ are intuitionistic tautologies. This is not true for S_R, and a simple counterexample is $x \vee (x \Rightarrow y)$ (consider valuation such that $v(x) = \mathbb{R} \setminus \{0\}$ and $v(y) = \emptyset$).

Proposition 6 *A formula from $S_R \cap \mathcal{H}$ is an intuitionistic tautology if and only if the positive prefix of the path leading to the negative leaf with the repeated variable is a prefix of the path leading to the positive leaf with the repeated variable. The set of those formulae is denoted by S_{RI} (see Figure 2).*

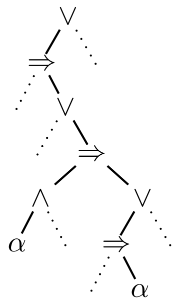

Figure 2: A tree with a negative path and a positive path with the same prefix.

Proof 6 *Let $\varphi \in S_R \cap \mathcal{H}$ be a formula such that the positive prefix of the path leading to the negative leaf with the repeated variable is a prefix of the path leading to the positive leaf with the repeated variable. Let s be the last common node of the positive and negative paths to the leaves with the repeated variable, v be any valuation in $\mathcal{O}(\mathbb{R})$, and $X \in \mathcal{O}(\mathbb{R})$ be the value assigned by v to the repeated variable. We know that the node s is labelled with $\Rightarrow$. Let φ_L and φ_R be its left and right subtrees. By the definition of the negative path we have an $\wedge$-path in φ_L to the leaf valuated to X. Since $\wedge$ is interpreted as an intersection we know that $[[\varphi_L]]^I_v \subset X$. In the similar way since $\vee$ is interpreted as an union and by the Observation 1 we get $[[\varphi_R]]^I_v \supset X$. Therefore $[[\varphi_L \Rightarrow \varphi_R]]^I_v = \mathbb{R}$ and the value is propagated to the root.*

For the other directions, let us take $\varphi \in S_R$ such that the last positive leaf on the negative path to the leaf with the repeated variable does not belong to the positive path to the other leaf with the repeated variable . Let l_1 and l_2 be the pair of leaves with repeated variable and s be the last common node of the paths to them. Obviously s is labelled with $\vee$. We use a valuation v which assigns the set $\mathbb{R} \setminus \{0\}$ to the variable occurring at the leaves l_1, l_2 and $\mathbb{R}$ to the other variables which have occurrences among 1-layer leaves, and $\emptyset$ to the remaining variables. Let $\varphi_L \vee \varphi_R$ be the subtree of φ rooted in s. The easy structural induction shows that $[[\varphi_R]]^I_v = \mathbb{R} \setminus \{0\}$ and $[[\varphi_L]]^I_v = \emptyset$ (the left subformula of the last positive node s_2 on the path to the negative node l_2 is valuated to $\mathbb{R} \setminus \{0\}$, it gives $\emptyset$ at the node s_2). Then we have $[[\varphi_R \vee \varphi_L]]^I_v = \mathbb{R} \setminus \{0\}$ and since all the 0-shallow nodes, which are not below or above s, are valuated by v to $\emptyset$, we get $[[\varphi]]^I_v = \mathbb{R} \setminus \{0\}$, which proves that φ is not an intuitionistic tautology.

Analogously to the inequality (3) we get

$$S_\perp(n) + S_{RI}(n) - \mathcal{F}^{[\geqslant 2]}_{\leqslant 3}(n) \leqslant Int(n) \leqslant S_\perp(n) + S_{RI}(n) + \mathcal{F}^{[\geqslant 2]}_{\leqslant 3}(n). \tag{4}$$

4 Counting simple families of tautologies

Within this section we denote by $\mathcal{T}_{\leq 3}^{(2,3,3,4)}(n)$ the value $\mathcal{T}_{\leqslant 3}^{(2)}(n) + 2 \cdot \mathcal{T}_{\leqslant 3}^{(3)}(n) + \mathcal{T}_{\leqslant 3}^{(4)}(n)$.

For any $i \in \mathbb{N}$ an i-positive-pointed structure is an i-pointed structure, whose pointed leaves are all positive (note that positivity of leaves depends only on the structure). Negative-pointed structures are defined analogously. We use the following sets of structures:

- $\mathcal{T}_N$ - the set of 1-negative-pointed structures,
- $\mathcal{T}_{PN}$- the set of 2-pointed structures such that the first pointed leaf is positive and the second is negative,
- $\mathcal{T}_{\widehat{PN}}$- the subset of $\mathcal{T}_{PN}$ consisting of all the structures for which the positive prefix of the path to the negative pointed leaf is a prefix of the (positive) path to the positive pointed leaf.

In the following propositions we give bounds on the number of elements of $S_\perp$ and S_R of size n.

Proposition 7

$$\mathcal{T}_N(n) \cdot Lab(n-1) - \mathcal{T}_{\leq 3}^{(2,3,3,4)}(n) \cdot Lab(n-2) \leqslant S_\perp(n) \leqslant \mathcal{T}_N(n) \cdot Lab(n-1)$$

Proof 7 *From every 1-negative-pointed structure we can construct a formula from $S_\perp$ by a labelling which assigns $\perp$ to the pointed leaf. If the pointed structure has n leaves we have exactly $Lab(n-1)$ such labellings. Since every formula from $S_\perp$ can be constructed in this way we get:*

$$S_\perp(n) \leqslant \mathcal{T}_N(n) \cdot Lab(n-1).$$

The inequality is usually strict, since some formulae can be generated with more than one structure-labelling pairs of considered type. These are exactly formulae, which have at least two negative leaves labelled with $\perp$ (hence they have at least two 3-shallow repetitions). The number of pairs which generate formulae with that property is smaller than the number of pairs which generate all the formulae with at least two 3-shallow repetition. We get (just as in the proof of the Proposition 2)

$$\mathcal{T}_N(n) Lab(n-1) - \mathcal{T}_{\leq 3}^{(2,3,3,4)}(n) \cdot Lab(n-2) \leqslant S_\perp(n).$$

Proposition 8

$$\mathcal{T}_{PN}(n) \cdot Lab(n-1) - \mathcal{T}_{\leq 3}^{(2,3,3,4)}(n) \cdot Lab(n-2) \leqslant S_R(n) \leqslant \mathcal{T}_{PN}(n) \cdot Lab(n-1)$$

Proof 8 *The upper bounds comes from the number of structure-labelling pairs, such that the structure has two pointed leaves, a first positive and a second negative, and the labelling is such that it labels both pointed leaves with the same variable. Clearly, the set of formulae constructed in this way equals S_R. Therefore*

$$S_R(n) \leqslant \mathcal{T}_{PN}(n) \cdot Lab(n-1).$$

Just like in the proof of the lower bound on the family $S_\perp$ (Proposition 7), each formula from S_R which is constructed by at least two structure-labelling pairs of considered type, has at least two 3-shallow repetitions. Hence,

$$\mathcal{T}_{PN}(n) \cdot Lab(n-1) - \mathcal{T}_{\leq 3}^{(2,3,3,4)}(n) \cdot Lab(n-2) \leqslant S_R(n).$$

Corollary 1 *Applying the same reasoning for S_{RI} as in the Proposition 8 we get the following inequalities*

$$\mathcal{T}_{\widehat{PN}}(n) \cdot Lab(n-1) - \mathcal{T}_{\leq 3}^{(2,3,3,4)}(n) \cdot Lab(n-2) \leqslant S_{RI}(n) \leqslant \mathcal{T}_{\widehat{PN}}(n) \cdot Lab(n-1).$$

4.1 Structural limits

To prove our main results we need to calculate the following three "structural limits":

$$D_N = \lim_{n\to\infty} \frac{\mathcal{T}_N(n)}{\mathcal{T}(n)}, \qquad D_{PN} = \lim_{n\to\infty} \frac{\mathcal{T}_{PN}(n)}{\mathcal{T}(n)}, \qquad D_{\widehat{PN}} = \lim_{n\to\infty} \frac{\mathcal{T}_{\widehat{PN}}(n)}{\mathcal{T}(n)}.$$

(It is not even obvious, that such limits exist.)

Proposition 9

$$D_N = \lim_{n\to\infty} \frac{\mathcal{T}_N(n)}{\mathcal{T}(n)} = \frac{5}{8}$$

Proof 9 *Let $g_N(y,z)$ be the generating function for all structures, with z marking the size and y marking leaves which can be obtained from the root by paths containing only $\wedge$-nodes. It satisfies:*

$$g_N(y,z) = 2 \cdot T(z)^2 + g_N(y,z)^2 + yz.$$

Let $f_N(y,z)$ be the generating function for all structures with z marking size and with negative leaves marked with y. We have

$$f_N(y,z) = f_N(y,z)^2 + g_N(y,z) \cdot f_N(y,z) + T(z)^2 + z. \tag{5}$$

The first term corresponds to the situation when the root of the tree is labelled by $\vee$. The second one, to the situation when the root is labelled with $\Rightarrow$ (the left subtree can add some negative paths when all the following nodes are labelled by $\wedge$ and the right subtree extends the positive part of eventually negative paths). The third term corresponds to the situation when the root is labelled by $\wedge$, such trees do not contain any negative paths. Finally, a single leaf gives the term z (it is 0-shallow leaf, therefore no negative path ends in it).

By the classical construction (pointing corresponds to differentiation), to obtain the generating function for 1-negative-pointed structures $SN(z)$ it is enough to differentiate $f_N(y,z)$ with respect to the variable y, multiply by y, and then substitute y by 1 (we no longer need bivariate function). Therefore

$$SN(z) = y \cdot \frac{\partial f_N(y,z)}{\partial y}|_{y:=1}.$$

After algebraic calculations and application of the Lemma 1 we get:

$$\lim_{n\to\infty} \frac{\mathcal{T}_N(n)}{\mathcal{T}(n)} = \lim_{n\to\infty} \frac{[z^n]SN(z)}{[z^n]T(z)} = \lim_{z\to_{\mathbb{R}} \frac{1}{12}^-} \frac{SN'(z)}{T'(z)} = \frac{5}{8}.$$

Proposition 10

$$D_{PN} = \lim_{n\to\infty} \frac{\mathcal{T}_{PN}(n)}{\mathcal{T}(n)} = \frac{11}{8}$$

Proof 10 *We proceed as in the previous proposition. We use the generating function $f_{PN}(x,y,z)$ enumerating all structures where x marks positive leaves, y marks negative leaves and z marks the size. This generating function satisfies modified equation (5):*

$$f_{PN}(x,y,z) = f_{PN}(x,y,z)^2 + g_N(y,z) \cdot f_{PN}(x,y,z) + T(z)^2 + xz.$$

By differentiation with respect to variables x and y, and then multiplication by $x \cdot y$, we get the generating function for the set $\mathcal{T}_{PN}$. Just as in the previous case we can substitute 1 for x and y, to obtain a univariate generating function $SPN(z)$.

$$SPN(z) = x \cdot y \cdot \frac{\partial^2 f_{PN}(x,y,z)}{\partial x \partial y}|_{x:=1,y:=1}$$

By algebraic computations and Lemma 1 we get

$$\lim_{n\to\infty} \frac{[z^n]SPN(z)}{[z^n]T(z)} = \lim_{z\to_{\mathbb{R}} \frac{1}{12}^-} \frac{SPN'(z)}{T'(z)} = \frac{11}{8}.$$

Proposition 11

$$\lim_{n\to\infty} \frac{\mathcal{T}_{\widehat{PN}}(n)}{\mathcal{T}(n)} = \frac{5}{8}$$

Proof 11 *The set $\mathcal{T}_{\widehat{PN}}$ consists of elements of $\mathcal{T}_{PN}$, for which the positive prefix of the path to the negative pointed leaf is a prefix of the path to the positive pointed leaf. Every such a structure can be uniquely decomposed into three 1-pointed structures. Let s be the last positive node on the negative path to the pointed leaf. The first structure is obtained by substituting s by a leaf, and pointing to this leaf (it is an 1-positive-pointed structure). The second one is the right subtree of s - it is an 1-positive-pointed structure (pointing is inherited). The last structure is the left subtree of s, it inherits one pointed leaf, this time it is an 1-pointed structure such that all the nodes on the path to the pointed leaf are labelled by $\wedge$. It is easy to observe that such a decomposition is unique, and from a pair of positive-pointed structures and one "$\wedge$-pointed" structure we can construct a structure belonging to $\mathcal{T}_{\widehat{PN}}$. The size of the constructed structure is smaller by 1 than the sum of the sizes of its component. Therefore the generating function for the elements of $\mathcal{T}_{\widehat{PN}}$ is*

$$SPNI(z) = \frac{1}{z}\left(x \cdot y \cdot \frac{\partial f_{PN}(x,y,z)}{\partial x}|_{x:=1,y:=1}\right)^2 \left(y \cdot \frac{\partial g_N(y,z)}{\partial y}|_{y:=1}\right).$$

Just like in the previous cases the functions are algebraic and after algebraic computations the application of Lemma 1 yields:

$$\lim_{n\to\infty} \frac{\mathcal{T}_{\widehat{PN}}(n)}{\mathcal{T}(n)} = \lim_{n\to\infty} \frac{[z^n]SPNI(z)}{[z^n]T(z)} = \lim_{z\to_{\mathbb{R}} \frac{1}{12}^-} \frac{SPNI'(z)}{T'(z)} = \frac{5}{8}.$$

Using the bounds from Proposition 7 and the "structural limits", that we have just computed, we get:

$$\frac{S_\perp(n)}{\mathcal{F}(n)} \leqslant \frac{\mathcal{T}_N(n)}{\mathcal{T}(n)} \cdot \frac{Lab(n-1)}{Lab(n)} \sim \frac{5}{8}\frac{Lab(n-1)}{Lab(n)} \tag{6}$$

and

$$\frac{S_\perp(n)}{\mathcal{F}(n)} \geqslant \frac{\mathcal{T}_N(n)}{\mathcal{T}(n)}\frac{Lab(n-1)}{Lab(n)} - \frac{\mathcal{T}_{\leq 3}^{(2,3,3,4)}(n)}{\mathcal{T}(n)}\frac{Lab(n-2)}{Lab(n)} \sim \frac{5}{8}\frac{Lab(n-1)}{Lab(n)} - C_\perp \frac{Lab(n-2)}{Lab(n)}, \quad (7)$$

for some $C_\perp \in \mathbb{R}$.

In the similar way, using Proposition 8, for S_R we get

$$\frac{S_R(n)}{\mathcal{F}(n)} \leqslant \frac{\mathcal{T}_{PN}(n)}{\mathcal{T}(n)} \cdot \frac{Lab(n-1)}{Lab(n)} \sim \frac{11}{8}\frac{Lab(n-1)}{Lab(n)} \tag{8}$$

and

$$\frac{S_R(n)}{\mathcal{F}(n)} \geqslant \frac{\mathcal{T}_{PN}(n)}{\mathcal{T}(n)} \cdot \frac{Lab(n-1)}{Lab(n)} - \frac{\mathcal{T}_{\leq 3}^{(2,3,3,4)}(n)}{\mathcal{T}(n)}\frac{Lab(n-2)}{Lab(n)} \sim \frac{11}{8}\frac{Lab(n-1)}{Lab(n)} - C_{PN}\frac{Lab(n-2)}{Lab(n)}, \tag{9}$$

for some $C_{PN} \in \mathbb{R}$.

Finally, from Corollary 1, we obtain

$$\frac{S_{RI}(n)}{\mathcal{F}(n)} \leqslant \frac{\mathcal{T}_{\widehat{PN}}(n)}{\mathcal{T}(n)} \cdot \frac{Lab(n-1)}{Lab(n)} \sim \frac{5}{8}\frac{Lab(n-1)}{Lab(n)} \tag{10}$$

and

$$\frac{S_{RI}(n)}{\mathcal{F}(n)} \geqslant \frac{\mathcal{T}_{\widehat{PN}}(n)}{\mathcal{T}(n)} \cdot \frac{Lab(n-1)}{Lab(n)} - \frac{\mathcal{T}_{\leq 3}^{(2,3,3,4)}(n)}{\mathcal{T}(n)}\frac{Lab(n-2)}{Lab(n)} \sim \frac{5}{8}\frac{Lab(n-1)}{Lab(n)} - C_{\widehat{PN}}\frac{Lab(n-2)}{Lab(n)}, \tag{11}$$

for some $C_{\widehat{PN}} \in \mathbb{R}$.

4.2 Main result - bounded case

We specialize now to the case with the number of variables bounded by k. In that case we have $Lab(n) = (k+1)^n$. For the clarity of the exposition we keep using $\mathcal{F}$ instead of $\mathcal{F}_k$. From inequalities 6 and 7, we get

$$\limsup_{n\to\infty} \frac{S_\perp(n)}{\mathcal{F}(n)} \leqslant \frac{5}{8k} \tag{12}$$

and

$$\liminf_{n\to\infty} \frac{S_\perp(n)}{\mathcal{F}(n)} \geqslant \frac{5}{8k} - \frac{C}{k^2}. \tag{13}$$

Analogous reasoning for families S_R and S_{RI}, using inequalities 9, 8, 11, 10 gives

$$\limsup_{n\to\infty} \frac{S_R(n)}{\mathcal{F}(n)} \leqslant \frac{11}{8k} \qquad \liminf_{n\to\infty} \frac{S_R(n)}{\mathcal{F}(n)} \geqslant \frac{11}{8k} - \frac{C}{k^2}, \tag{14}$$

and

$$\limsup_{n\to\infty} \frac{S_{RI}(n)}{\mathcal{F}(n)} \leqslant \frac{5}{8k} \qquad \liminf_{n\to\infty} \frac{S_{RI}(n)}{\mathcal{F}(n)} \geqslant \frac{5}{8k} - \frac{C}{k^2}. \tag{15}$$

We want to estimate $D_k^+ = \limsup_{n\to\infty} \frac{Int_k(n)}{Cl_k(n)}$. Clearly, we have

$$D_k^+ = \limsup_{n\to\infty} \frac{\mathcal{F}(n)^{-1} Int_k(n)}{\mathcal{F}(n)^{-1} Cl_k(n)}.$$

Applying upper bound from 4 and lower bound from 3 we get

$$D_k^+ \leqslant \limsup_{n\to\infty} \frac{\mathcal{F}(n)^{-1}(S_\perp(n) + S_{RI}(n) + \mathcal{F}_{\leqslant 3}^{[\geqslant 2]}(n))}{\mathcal{F}(n)^{-1}(S_\perp(n) + S_R(n) - \mathcal{F}_{\leqslant 3}^{[\geqslant 2]}(n))}.$$

Since both sequences (in numerator and in denominator) are positive, for large enough n and k, and bounded, we can apply following upper bound:

$$D_k^+ \leqslant \frac{\limsup_{n\to\infty} \mathcal{F}(n)^{-1}(S_\perp(n) + S_{RI}(n) + \mathcal{F}_{\leqslant 3}^{[\geqslant 2]}(n))}{\liminf_{n\to\infty} \mathcal{F}(n)^{-1}(S_\perp(n) + S_R(n) - \mathcal{F}_{\leqslant 3}^{[\geqslant 2]}(n))} \leqslant \frac{\frac{10}{8k} + o(\frac{1}{k})}{\frac{2}{k} - o(\frac{1}{k})} \sim_k \frac{5}{8}$$

The last inequality is a consequence of inequalities 12, 13, 14, 15 and Proposition 2.

In the analogous way we obtain

$$D_k^- \geqslant \frac{\liminf_{n\to\infty} \mathcal{F}_k(n)^{-1}(S_\perp(n) + S_{RI}(n) - \mathcal{F}_{\leqslant 3}^{[\geqslant 2]}(n))}{\limsup_{n\to\infty} \mathcal{F}_k(n)^{-1}(S_\perp(n) + S_R(n) + \mathcal{F}_{\leqslant 3}^{[\geqslant 2]}(n))} = \frac{\frac{10}{8k} - o(\frac{1}{k})}{\frac{2}{k} + o(\frac{1}{k})} \sim_k \frac{5}{8}$$

Hence we get the first of our main results:

$$\lim_{k\to\infty} D_k^- = \lim_{k\to\infty} D_k^+ = \frac{5}{8}.$$

4.3 Main result - unbounded case

In this section we specialize to the case when the number of variables is unbounded, but formulae are considered "up to the names of variables". We put $Lab(n) = B(n+1)$, where $B(n)$ is a Bell number (see [6]). The asymptotic behaviour of the Bell numbers is known due to the result of Moser and Wyman [9]. For our needs it is sufficient to note that Bell numbers satisfy the following property: $\frac{B(n-2)}{B(n)} = o(\frac{B(n-1)}{B(n)})$. Then, from inequalities 4, Proposition 2 and structural limits, we get

$$\frac{Int_\infty(n)}{\mathcal{F}_\infty(n)} = \frac{S_\perp(n) + S_{RI}(n)}{\mathcal{F}_\infty(n)} + o\left(\frac{B(n)}{B(n+1)}\right) \sim \frac{10}{8}\frac{B(n)}{B(n+1)} + o\left(\frac{B(n)}{B(n+1)}\right).$$

Analogously for classical tautologies we have

$$\frac{Cl_\infty(n)}{\mathcal{F}_\infty(n)} = \frac{S_\perp(n) + S_R(n)}{\mathcal{F}_\infty(n)} + o\left(\frac{B(n)}{B(n+1)}\right) \sim \frac{16}{8}\frac{B(n)}{B(n+1)} + o\left(\frac{B(n)}{B(n+1)}\right).$$

Both asymptotic equivalents are precise enough to derive the second of our main results:

$$\frac{Int_\infty(n)}{Cl_\infty(n)} \sim \frac{5}{8}.$$

5 Summary and smaller logics

Going from the simplest language (a single connective: $\Rightarrow$) up to the full propositional logic $(\Rightarrow, \wedge, \vee, \perp)$, we note that there exists a difference between the intuitionistic logic and the classical logic (when the number of variables tends to infinity) as soon as the $\vee$ connective is used. We note that we obtain classical tautologies that are not intuitionistic in S_R, if and only if the positive prefix of the path leading to the negative leaf with the repeated variable is not a prefix of the path leading to the positive leaf with the repeated variable. See Figure 3. So we conclude that there exists a difference between the asymptotic density of intuitionistic

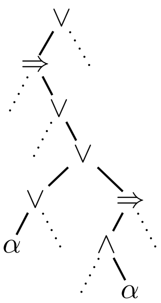

Figure 3: A tree with a negative path and a positive path without the same prefix.

tautologies into classical ones (in both models) if and only if the connective $\vee$ belongs to the set of connectives. Here we summarise some computations of such fractions. For both models the results are always identical. If we restrict the connectives to $\{\Rightarrow, \vee\}$, without $\perp$, the fraction is equal to 3/13. With these connectives and the constant $\perp$, we get 2/7. Now, if the connectives are $\{\Rightarrow, \vee, \wedge\}$, without the constant, the fraction between both logics is 5/11 and it becomes 5/8 is $\perp$ is permitted. Finally, we want once again to emphasise that the coherence of the results in the bounded and unbounded approaches is quite an interesting fact in itself. We believe that Proposition 1 sheds some light on this phenomenon.

References

[1] B. Chauvin, P. Flajolet, D. Gardy, and B. Gittenberger, And/Or *trees revisited*, Combinatorics, Probability and Computing **13** (2004), no. 4-5, 475–497.

[2] P. Flajolet and R. Sedgewick, *Analytic combinatorics*, Cambridge University Press, 2009.

[3] H. Fournier, D. Gardy, A. Genitrini, and M. Zaionc, *Classical and intuitionistic logic are asymptotically identical*, CSL, 2007, pp. 177–193.

[4] D. Gardy, *Random boolean expressions*, Computational Logic and Applications, CLA '05 (René David, Danièle Gardy, Pierre Lescanne, and Marek Zaionc, eds.), DMTCS Proceedings, vol. AF, Discrete Mathematics and Theoretical Computer Science, 2005, pp. 1–36.

[5] A. Genitrini, J. Kozik, and M. Zaionc, *Intuitionistic vs classical tautologies, quantitative comparison*, Lecture Notes in Computer Science (2008), no. 4941, 100–109.

[6] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete mathematics: a foundation for computer science*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1989.

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

[8] M. Moczurad, J. Tyszkiewicz, and M. Zaionc, *Statistical properties of simple types*, Mathematical Structures in Computer Science **10** (2000), no. 5, 575–594.

[9] L. Moser and M. Wyman, *An asymptotic formula for the bell numbers*, Transactions of the Royal Society of Canada **XLIX** (1955).

[10] M. Sorensen and P. Urzyczyn, *Lectures on the curry-howard isomorphism*, 1998.