Quantitative comparison of Intuitionistic and Classical logics - full propositional system *

Antoine Genitrini[1] and Jakub Kozik[2]

[1] PRiSM, CNRS UMR 8144, Université de Versailles Saint-Quentin
45 av. des États-Unis, 78035 Versailles cedex, France.
antoine.genitrini@prism.uvsq.fr
[2] Theoretical Computer Science, Jagiellonian University,
ul. Łojasiewicza 6, 30-348 Kraków, Poland.
jakub.kozik@uj.edu.pl

Abstract. We address the problem of quantitative comparison of classical and intuitionistic logics within the language of the full propositional system. We apply two different approaches, to estimate the asymptotic fraction of intuitionistic tautologies among classical tautologies, obtaining the same results for both. Our results justify informal statements such as "about 5/8 of classical tautologies are intuitionistic".

1 Introduction

It is a standard approach to use the notion of density [6,1] to analyse quantitative relations between countable sets. The general idea is to consider subsets of elements of bounded size, and to 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.

One of the first papers to address the quantitative aspects of intuitionistic logic was [6], which (according to the authors) 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*". Most results of that paper were formulated in terms of inhabitation of types in simple λ-calculus. However, under Curry-Howard isomorphism (see e.g. [8]), they translate directly to the framework of intuitionistic logic.

The authors of [6] considered calculus with a finite number of ground types, and only functional types. In terms of logical formulae it means that the number of different variables in a formula was bounded by some constant, and the only allowed connective was $\Rightarrow$. The authors proved that at least 1/3 of classical

* Research described in this paper was partially supported by the A.N.R. project *SADA*, French government research grant for young scientists (program number 0185) and by POLONIUM grant *Quantitative research in logic and functional languages*, cooperation between Jagiellonian University of Krakow, L' École Normale Supérieure de Lyon and L'Université de Versailles Saint-Quentin, contract number 7087/R07/R08

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 have also conjectured that, among the formulae with the 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 was known to be true for the formulae using only one variable for the trivial reason that both sets of tautologies are equal in this case.

Although the conjecture of [6] is false, a slight reformulation turned out to be true. The authors of [3] proved 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. This counter-intuitive result raised a question about the appropriateness of the approach. In fact, the assumption about the bounded number of variables seemed to have a strong influence on the result. In the paper [4] the authors suggested another approach, in which formulae was considered up to a renaming of variables (i.e. two formulae which differ only in the naming of variables were assumed to be equal). In that case the authors could 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 densities tend 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 $\bot$. 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.

2 Prerequisites and results

For any set of finite elements A and $n \in \mathbb{N}$ we denote by $A(n)$ the number of elements of set A with size n (the element is finite if it has finite size).

Formulae and terms. Let $\mathrm{Var} = \{x_1, x_2, x_3, \ldots\}$ be a countable set of variables, $\bot$ 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 labelled by the elements of $\mathrm{Var} \cup \{\bot\}$ (precisely the tree is rooted and planar i.e. the order of descendants matters). For every $k \in \mathbb{N}$ let $\mathcal{F}_k$ denote the set of terms whose 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 tree 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 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. We denote Term/α by $\mathcal{F}_\infty$. We use the name formula both for terms and for elements from $\mathcal{F}_\infty$.

Intuitionistic logic. For a general reference about intuitionistic logic we suggest e.g. [8]. These are well known facts that every intuitionistic tautology is classical, and that the converse is not true (even for implicational fragment). In our proofs we use also the fact that the Heyting algebra of open subsets of $\mathbb{R}$ (with respect to euclidean topology) is a complete model for the propositional intuitionistic logic.

2.1 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 $(d_k(n))_{n\in\mathbb{N}}$ be defined as: $d_k(n) = Int_k(n)/Cl_k(n)$. Each fraction $d_k(n)$ equals the probability that a formula, chosen uniformly at random among the set of elements of Cl_k of size n, is an intuitionistic tautology. 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 existence of D_k. We use the following bounds instead:

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

The first of our main results says:

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

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: $d_\infty(n) = Int_\infty(n)/Cl_\infty(n)$.

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 [4] that the analogous approach for the implicational fragments gives the density 1.

We derive both results in an unified way from some structural properties of tautologies.

2.2 Structure and labelling

For every formula 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 a ∞-labelling of size n.

As usual the size of a formula is the number of its leaves. We use the same convention for the size of a structure. We denote by $\mathcal{T}(n)$ the number of trees from $\mathcal{T}$ of size n.

Note, that in all the 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. That fact is reflected in simple expressions for the numbers of formulae of size n. We have

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

where $B(n+1)$ is the number of equivalence relations on the set $\{0, 1, \ldots, n\}$, known as Bell number (see e.g. [5]).

2.3 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 ones.

We use a notation which always exposes the formal parameters of a generating function. E.g. we write $g(z)$ instead of 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. 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. We denote it by $t(z)$. By a standard constructions we get an algebraic equation for $t(z)$, where z marks the size:

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

Solving this equation (and choosing the proper solution) we get

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

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

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*

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

with both d_f, d_g being nonzero. Then: $\lim_{n \to \infty} \frac{[z^n]f(z)}{[z^n]g(z)} = \lim_{z \to_{\mathbb{R}} \varrho^-} \frac{f'(z)}{g'(z)}$.

Using Theorem VII.8 from [2] we obtain this equality because both sides are equal to d_f/d_g.

3 Structural properties of tautologies

Within this section we consider structural properties of tautologies, which are independent of the approach we use (bounded or unbounded). In order to obtain results independent from the kind of labelling, we use $\mathcal{F}$ to denote the set of formulae under consideration, and the function $Lab : \mathbb{N} \to \mathbb{N}$ which for every $n \in \mathbb{N}$ returns the number of all different labellings of the structure of size n. In particular we get results for the unbounded approach by setting $\mathcal{F}$ equal to $\mathcal{F}_\infty$ and $Lab(n) = B(n+1)$. 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

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

Pointed structures. An m-pointed structure is a pair (t, s) of a structure t and a sequence 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 elements of A by the 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 give a formula from $\mathcal{F}_A$ from the structure a. Therefore $\mathcal{F}_A(n) \leqslant A(n)Lab(n-1)$.

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 $\Rightarrow$. We say it 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 $\Rightarrow$ 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 $\Rightarrow$-node, and u marking the remaining leaves (t denotes a formal parameter, not the generating function for all trees which we denote by $t(z)$). The generating function is given implicitly by an initial condition and by the equation

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

which reflects the fact that every such a tree is either an implication with the left subtree being a 1-layer leaf and the right subtree belonging to P, or a conjunction or a disjunction with both subtrees belonging to P, or a leaf (which is 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 $p_{\leqslant n}(t,u)$ 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 $\mathcal{T}^{(m)}_{\leqslant s}$ 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*

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

Proof. 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 holomorphic in the set $D_\varepsilon = \{(t,u) \in \mathbb{C}^2 : |t| \leqslant \frac{1}{6} + \varepsilon,\ |u| \leqslant \frac{1}{12} + \varepsilon\}$ for some small positive $\varepsilon \in \mathbb{R}$ (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 holomorphic in D_ϵ (for some positive $\epsilon \in \mathbb{R}$) and so are all its partial derivatives, in particular $\frac{\partial^m p_{\leqslant s}(t,u)}{(\partial u)^m}$. 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 the leaves (after pointing we are no longer interested in s-shallow leaves). As a result we obtain a function

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

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 holomorphic 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 singularity in ρ. Therefore we know that $p_{m,s}(z)$ has unique dominating singularity in ρ. In fact we know also that the limit $\lim_{z \to_{\mathbb{R}} \rho^-} p_{m,s}(z) < \infty$, 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 $t(v^2)$ is analytic at ρ we get that $p_{m,s}(v^2)$ 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.

In fact we need the Proposition 1 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.

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 is the set of k-shallow leaves we say that φ has r k-shallow repetitions. Note, that the occurrence of the constant is treated as a repetition e.g. the formula $x \Rightarrow \bot$ has one repetition 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 from above 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*

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

Proof. 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 $\bot$;
B . φ contains one 3-shallow leaf labelled with $\bot$ 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 from $\varphi \in \mathcal{F}^{[\geqslant 2]}_{\leqslant 3}$ 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 the inequality is usually strict.

Every formula from $\mathcal{F}^A$ contains at least two 3-shallow leaves labelled with $\perp$. Therefore it can be constructed from a 3-shallow 2-pointed structure by some labelling 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:

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

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

Using these equations and Proposition 1 we obtain

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

The above proposition will be used to show that we can neglect all formulae with at least two 3-shallow repetitions, since the number of them will be shown to be essentially smaller than the number of tautologies.

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

Proposition 3 *If a formula φ does not contain at least one 1-shallow repetition, it is not a classical tautology.*

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

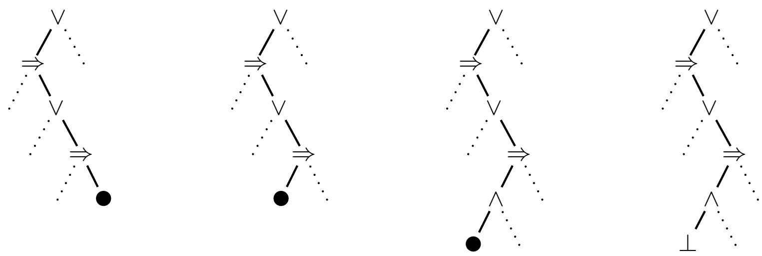

Fig. 1. From left to right: a positive path, not a positive path, a negative path, a tree with a negative leaf labelled with $\bot$.

Definition 2 *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 left from* h *and then follows only* $\wedge$*-nodes (if any). A node is called negative if there exists a negative path to it (see Fig. 1).*

For every formula it is enough to valuate one of its positive nodes to $True$ or one of its negative nodes to $False$, to ensure that the valuation of the whole formula is $True$.

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

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

Observation 2 *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*.*

We call the formulae from the set $S_R \cup S_\bot$, *simple tautologies*. We focus on the formulae with exactly one 2-shallow repetition and no more than one 3-shallow repetition. 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* $\varphi \in \mathcal{H} \setminus S_\bot$ *contains a 3-shallow leaf* l *labelled with* $\bot$*, then it is not a tautology.*

Proof. 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 a tautology (Proposition 3). If l is 0-shallow then we can use the valuation v_φ^1 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_φ^1 to $False$.

In the remaining case l is a 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 a 1-layer node, because l is not negative.

Suppose that s is labelled by $\vee$. One of its subtrees does not contain shallow occurrences of l. In that subtree all the 0-shallow leaves are valuated by $v_\varphi^{1,3}$ to

$True$ (because they are all 1-layer in φ) therefore the whole subtree with root s is valuated to $True$ by $v_\varphi^{1,3}$.

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 $\bot$, all the labels of 2-layer and 3-layer leaves are not repeated among the 3-shallow leaves. Therefore the valuation $v_\varphi^{1,3}$ assigns $False$ to all the 2-layer leaves, and $True$ to all the 3-layer leaves. Consequently, every 2-layer node is valuated to $False$. Is 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 $v_\varphi^{1,3}$ 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 variable repetition, then it is not a tautology.*

Proof. 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 repeated variable are on the same level, then the valuation v_φ^1 valuates all the 0-shallow leaves to $False$ and all 1-layer leaves to $True$, and the formula is valuated by v_φ^1 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_φ^1 are below s. But s is valuated to $False$, because it is a $\wedge$-node and one of its subtrees is valuated by v_φ^1 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 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 $\bot$, to prove that φ is not a tautology.

We have

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

The lower bound comes from the fact that every formula which belongs to $S_\bot \cap S_R$ 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 $\mathcal{F}_{\leqslant 3}^{[\geqslant 2]}$.

Simple Intuitionistic Tautologies. It is easy to show that all the formulae from $S_\bot$ are intuitionistic tautologies. This is not true for S_R, and a simple counterexample is $x \vee (x \Rightarrow y)$. However, we can prove the following proposition (e.g. by using the Heyting algebra of open subsets of $\mathbb{R}$).

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 (i.e. the last common node is a $\Rightarrow$-node). The set of those formulae is denoted by S_{RI} (see Fig. 2).*

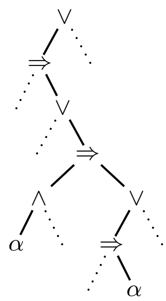

Fig. 2. A tree with a negative path and a positive path with the same prefix.

Analogously to the inequality (3) we get

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

4 Counting simple families

Within this section we denote by $\mathcal{T}^{(2,3,4)}_{\leq 3}(n)$ the value $\mathcal{T}^{(2)}_{\leqslant 3}(n) + 2\mathcal{T}^{(3)}_{\leqslant 3}(n) + \mathcal{T}^{(4)}_{\leqslant 3}(n)$. For any $i \in \mathbb{N}$, a i-positive-pointed structure is a 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 one 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

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

Proof. 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:

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

The inequality is usually strict, since some formulae can be generated with more than one structure-labelling pair of considered type. Those are exactly the formulae, that have at least two negative leaves labelled with $\perp$ (hence they have at least two 3-shallow repetitions). But 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 Proposition 2),

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

An analogous result holds for S_R.

Proposition 8

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

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

We omit the technical proof.

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

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

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

4.1 Structural limits

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

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

Proposition 9 $\quad D_N = \lim_{n\to\infty} \frac{\mathcal{T}_N(n)}{\mathcal{T}(n)} = \frac{5}{8}.$

Proof. Let $T(z)$ be the generating function for $\mathcal{T}$ and $g_N(y,z)$ be the generating function for $\mathcal{T}$, with z marking the size and y marking the leaves that can be obtained from root by paths containing only $\wedge$-nodes. It satisfies:

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

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

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

The first term is obtained when the root is labelled by $\vee$. The second one, by $\Rightarrow$, and the third term corresponds to $\wedge$.

By a 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, and then to substitute y by 1 (we no longer need bivariate function). Algebraic calculations and the application of the Lemma 1 give:

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

In the similar way we prove the following two propositions.

Proposition 10 $\quad D_{PN} = \lim_{n\to\infty} \frac{\mathcal{T}_{PN}(n)}{\mathcal{T}(n)} = \frac{11}{8}.$

Proposition 11 $\quad D_{\widehat{PN}} = \lim_{n\to\infty} \frac{\mathcal{T}_{\widehat{PN}}(n)}{\mathcal{T}(n)} = \frac{5}{8}.$

Using the bounds from the Proposition 7 and the limits we have computed, we get:

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

and

$$\begin{aligned}\frac{S_\perp(n)}{\mathcal{F}(n)} &\geqslant \frac{\mathcal{T}_N(n)}{\mathcal{T}(n)}\frac{Lab(n-1)}{Lab(n)} - \frac{\mathcal{T}^{(2,3,4)}_{\leq 3}(n)}{\mathcal{T}(n)}\frac{Lab(n-2)}{Lab(n)} \\ &\sim \frac{5}{8}\frac{Lab(n-1)}{Lab(n)} - C\frac{Lab(n-2)}{Lab(n)},\end{aligned}$$

for some $C \in \mathbb{R}$. Analogous estimates (using values D_{PN} and $D_{\widehat{PN}}$) hold for S_R and S_{RI}.

4.2 Main results

Unbounded case. The asymptotic behaviour of the Bell numbers is known due to the result of Moser and Wyman [7]. We are going to use the following property: $B(n-2)/B(n) = o(B(n-1)/B(n))$. The estimates from the previous subsection specialize to the unbounded case; inequalities 3 and 4 gives:

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

$$\sim \frac{B(n)}{B(n+1)}\left(\frac{10}{8}+o(1)\right),$$

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

$$\sim \frac{B(n)}{B(n+1)}\left(\frac{16}{8}+o(1)\right)$$

therefore

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

Bounded case. We specialize now to the case with the number of variables bounded by k. We get

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

and

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

Analogous reasoning for families S_R and S_{RI} gives

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

and

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

Therefore we get

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

$$= \frac{\frac{10}{8k}-o(\frac{1}{k})}{\frac{2}{k}+o(\frac{1}{k})} \sim_k \frac{5}{8}$$

and

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

$$= \frac{\frac{10}{8k}-o(\frac{1}{k})}{\frac{2}{k}+o(\frac{1}{k})} \sim_k \frac{5}{8}.$$

Hence

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

which is the second of our main results.

5 Final remarks

The reasoning we used for the full propositional system is also appropriate for other sets of connectives. If we allow only implication the method we presented reconstructs the results from [3] and [4] (i.e. in this case the density of intuitionistic logic in classical is 1). Adding conjunction and $\perp$ to the system does not change the situation. However, it suffices to consider the language which uses only $\Rightarrow$ and $\vee$ to observe a difference. For this language the asymptotic density of intuitionistic logic in the classical one equals $3/13$.

Finally, we want once again to emphasize 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 & Computing **13** (2004).
2. P. Flajolet and R. Sedgewick, *Analytic combinatorics*, in preparation, preprint available at http://algo.inria.fr/flajolet/Publications/book.pdf, 2008.
3. H. Fournier, D. Gardy, A. Genitrini, and M. Zaionc, *Classical and intuitionistic logic are asymptotically identical*, CSL, 2007, pp. 177–193.
4. A. Genitrini, J. Kozik, and M. Zaionc, *Intuitionistic vs classical tautologies, quantitative comparison*, Lecture Notes in Computer Science (2008), no. 4941, 100–109.
5. 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.
6. M. Moczurad, J. Tyszkiewicz, and M. Zaionc, *Statistical properties of simple types*, Mathematical Structures in Computer Science **10** (2000), no. 5, 575–594.
7. L. Moser and M. Wyman, *An asymptotic formula for the bell numbers*, Transactions of the Royal Society of Canada **XLIX** (1955).
8. M. Sorensen and P. Urzyczyn, *Lectures on the curry-howard isomorphism*, 1998.