

On the density and the structure of the Peirce-like formulae[†]

Antoine Genitrini¹ and Jakub Kozik² and Grzegorz Matecki²

¹ PRISM, CNRS UMR 8144, Université de Versailles,
Saint-Quentin en Yvelines, 45 av. des États-Unis, 78035 Versailles cedex, France.
antoine.genitrini@prism.uvsq.fr

² Theoretical Computer Science, Jagiellonian University,
Gronostajowa 3, Kraków, Poland.
[Jakub.Kozik,Grzegorz.Matecki]@tcs.uj.edu.pl

Within the language of propositional formulae built on implication and a finite number of variables k , we analyze the set of formulae which are classical tautologies but not intuitionistic (we call such formulae - Peirce's formulae). We construct the large family of so called simple Peirce's formulae, whose sequence of densities for different k is asymptotically equivalent to the sequence $\frac{1}{2k^2}$. We prove that the densities of the sets of remaining Peirce's formulae are asymptotically bounded from above by $\frac{c}{k^3}$ for some constant $c \in \mathbb{R}$. The result justifies the statement that in the considered language almost all Peirce's formulae are simple. The result gives a partial answer to the question stated in the recent paper by H. Fournier, D. Gardy, A. Genitrini and M. Zaionc - although we have not proved the existence of the densities for Peirce's formulae, our result gives lower and upper bound for it (if it exists) and both bounds are asymptotically equivalent to $\frac{1}{2k^2}$.

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

1 Introduction

Intuitionistic logic was developed in the beginning of the XX-th century, in search for a basis for constructive mathematics. Apart from philosophical origins, intuitionistic logic emerged independently in many different fields of mathematics. One of the most interesting examples is the Curry-Howard isomorphism, which relates intuitionistic proofs to programs in lambda calculus. The intuitionistic logic is known to be a proper subset of a classical one. An interesting formula which witnesses this fact is the Peirce's law $((p \rightarrow q) \rightarrow p) \rightarrow p$ which cannot be proved constructively (it needs some form of the law of excluded middle in the proof). Since implication turned out to be the most interesting connector in the intuitionistic logic we focus on the language of formulae which does not allow other connectors. One of the first results on the quantitative comparison between implicational fragments of both logics was obtained in the paper

[†]Research described in this paper was partially supported 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

of M. Moczurad, J. Tyszkiewicz and M. Zaionc [MTZ00]. They defined a family of *simple tautologies* which are intuitionistic tautologies with some specific structure. The conjecture stated in [MTZ00], that *almost all classical tautologies are simple*, was recently proved by H. Fournier, D. Gardy, A. Genitrini and M. Zaionc in [FGGZ07]. This surprising result can be reformulated as *almost all classical tautologies are intuitionistic*. In the present paper we extend one of the results from [FGGZ07] by estimating the density of formulae which are classical tautologies but not intuitionistic (we call them Peirce's formulae). This work is also a continuation of some studies on Boolean formulas. Most of them are dealing with *and/or* formulas [LS97, CFGG04, Koz08]. All are dealing with the probability and complexity of a Boolean function. But the last one, moreover describes in details the inside structure of almost all formulas computing a given function. In our work we give the inside details of a formula too, but the way of investigation is different.

1.1 Main results

Let $\mathcal{V} = \{x_1, x_2, x_3, \dots\}$ be a countable set of variables. Let \mathcal{T}_k be the set of implicational formulae t such that all the variables used in t belongs to the set $\mathcal{V}_k = \{x_1, \dots, x_k\}$. Let $Cl_k \subset \mathcal{T}_k$ be the set of all *Classical tautologies* and $Int_k \subset \mathcal{T}_k$ be the set of all *Intuitionistic tautologies*. For any set $\mathcal{A} \subset \mathcal{T}_k$ by $A(n)$ we denote the number of elements of the set \mathcal{A} of size n . We prove that:

$$\liminf_{n \rightarrow \infty} \frac{(Cl_k \setminus Int_k)(n)}{\mathcal{T}_k(n)} \sim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{(Cl_k \setminus Int_k)(n)}{\mathcal{T}_k(n)} \sim_{k \rightarrow \infty} \frac{1}{2k^2}.$$

Our proof is based on the construction of families of formulae whose union has density $1 - O(1/k^3)$. These families are easily defined by *family schemes* (see e.g. Fig. 2). Similar approach was taken in [FGGZ07]. However, to estimate the density of Peirce's formulae we need to consider another, more detailed, partition of the set of all formulae. Also, the density of each presented family must be calculated more precisely, since we are interested in the order of $1/k^2$. (The order $1/k$ is completely consumed by the simple tautologies.)

2 Basic facts

In a straightforward way we identify implicational formulae from \mathcal{T}_k with rooted binary planar trees with leaves labelled by the variables from \mathcal{V}_k and the inner nodes by \rightarrow . For a formula (tree) $\varphi \in \mathcal{T}_k$, the goal of φ (denoted by $r(\varphi)$) is the label of its rightmost leaf. For a set or sequence of trees S we denote by $r(S)$ the set of all goals of trees from S . The set of premises of a formula of the kind $\varphi \rightarrow \psi$ is the set of premises of ψ enlarged by the element φ , formula which is a leaf has no premises. For all considered types of trees the size of a tree is the number of its leaves.

2.1 Generating functions

Generating functions and results of singularity analysis for algebraic functions are important tools for our development. The exhaustive treatment of this subject can be found in [FS08]. The generating function for some set of trees \mathcal{A} is denoted by $g_{\mathcal{A}}(z)$ (formally it is the generating function for the sequence $(\mathcal{A}(n))_{n \in \mathbb{N}}$).

Easy construction shows that $f_k(z) = (1 - \sqrt{1 - 4kz})/2$ is the generating function for \mathcal{T}_k . For any distinguished subset of variables D with cardinality $d < k$ the generating function for all the formulae whose goal is not labelled with a variable from D is $b_k^d(z) = \frac{k-d}{k} f_k(z)$.

2.2 Tree families

Definition 1 A family scheme is a finite planar rooted unary-binary tree whose:

- leaves are labelled by the elements of some countable set of scheme variables (we use Greek letters),
- edges are labelled by either R (we call these regular edges) or S (scheme edges)

We demand additionally that all the edges which go to the left from some node are labelled with R and all the edges going down from unary nodes are scheme edges. We do not distinguish family schemes which can be transformed one into another (and back) by a bijective renaming of the scheme variables. In the pictures regular edges are represented by solid lines, scheme edges by dashed lines. We also usually assign names to the scheme edges (see e.g. Fig. 3).

For a family scheme T let $s(T), d(T)$ denote respectively the number of scheme edges in T and the number of different labels assigned to the leaves of T . The number of repetitions $rep(T)$ in a family scheme T is the total number of its leaves diminished by $d(T)$ (e.g. $((p \rightarrow q) \rightarrow p) \rightarrow p$ has 2 repetitions). We write s, d, rep instead of $s(T), d(T), rep(T)$ if the family scheme is clear from the context. As usual, the size of a scheme T is the number of its leaves (we denote it by $|T|$).

For a family scheme T an admissible substitution is any element of the set $((\mathcal{T}_k)^*)^{s(T)} \times \mathcal{V}^{d(T)}$. Let T be a family scheme, $\sigma = ((s_1, \dots, s_a), (v_1, \dots, v_b))$ be an admissible substitution for T , let (e_1, \dots, e_a) and (l_1, \dots, l_b) be the lists of scheme edges of T and scheme variables occurring in T both listed according to some fixed tree traversal order (lets say DFS). The application of the substitution σ to the family scheme T proceeds as follows:

- each leaf labelled with l_i is relabelled to v_i
- each scheme edge e_i is locally expanded by the sequence $s_i = (t_1, \dots, t_{k_i})$ as depicted below:

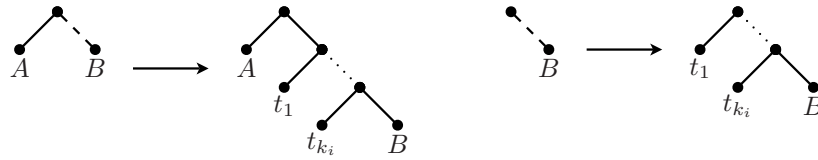


Fig. 1: scheme edge substitution.

The treatment of the substitution with sequences is straightforward. If the substituting sequence is empty, the scheme edge whose parent is a binary node become a regular edge. In case when the parent of the edge is unary, and the substituting sequence is empty the parent node of the scheme edge is replaced by the child node. Obviously, the result of an application of a substitution is a formula.

Definition 2 The substitution $((s_1, \dots, s_a), (v_1, \dots, v_b))$ is not proper if there exists a tree in some sequence s_i whose goal $r(t)$ equals some v_j or there exist different $i, j \leq b$ such that $v_i = v_j$. Otherwise, the substitution is proper.

The family of trees defined by the family scheme T (denoted by \mathcal{F}_k^T) consists of all the trees which can be constructed from T by the application of some **proper** substitution. By $\widehat{\mathcal{F}_k^T}$ we denote the set of formulae which can be constructed from T by any substitution (not necessarily proper). If there is a family scheme T such that $\mathcal{F}_k^T \subset \mathcal{H}_k \subset \widehat{\mathcal{F}_k^T}$ we say that the family \mathcal{H}_k corresponds to the scheme T .

2.3 Densities

For any set of trees $\mathcal{A} \subset \mathcal{T}_k$ we say that \mathcal{A} has density $\mu_k(\mathcal{A}) \in \mathbb{R}$ if

$$\mu_k(\mathcal{A}) = \lim_{n \rightarrow \infty} \frac{\mathcal{A}(n)}{\mathcal{T}_k(n)}.$$

It is easy to observe that there are sets whose density does not exist. However, the following limits always exist

$$\mu_k^-(\mathcal{A}) = \liminf_{n \rightarrow \infty} \frac{\mathcal{A}(n)}{\mathcal{T}_k(n)} \quad \text{and} \quad \mu_k^+(\mathcal{A}) = \limsup_{n \rightarrow \infty} \frac{\mathcal{A}(n)}{\mathcal{T}_k(n)}.$$

We cite below a technical Lemma which is a consequence of the Theorem VII.8 from [FS08]. Generating functions of all families of trees we use, have the assumed property.

Lemma 3 *Let $f(z), g(z)$ be generating functions having both a unique dominating singularity of square root type in $\rho \in \mathbb{R}_+$. Then, the limit $\lim_{n \rightarrow \infty} \frac{[z^n]f(z)}{[z^n]g(z)}$ exists and equals $\lim_{z \rightarrow \rho^-} \frac{f'(z)}{g'(z)}$.*

We use this lemma to estimate densities of families of trees defined by some family schemes.

Lemma 4 *For every family scheme T and $k \in \mathbb{N}$, the density $\mu_k(\mathcal{F}_k^T)$ exists and*

$$\mu_k(\mathcal{F}_k^T) = \frac{s(T)}{2^{2\text{rep}(T)+2d(T)-s(T)-1} \cdot k^{\text{rep}(T)}} + O\left(\frac{1}{k^{\text{rep}(T)+1}}\right).$$

Proof: Let us fix $k \in \mathbb{N}$. We have one to one correspondence between the proper substitutions admissible for T and the elements of \mathcal{F}_k^T . Let us define the size of a substitution $((s_1, \dots, s_{s(T)}), (v_1, \dots, v_{d(T)}))$ as a sum of the sizes of all trees from the sequences $s_1, \dots, s_{s(T)}$. It is easy to see that the size of the formula corresponding to the substitution equals the size of the substitution increased by the size of the scheme T . The generating function for the formulae whose goal does not contain any of the variable occurring in (v_1, \dots, v_b) is easily seen to be $b_k^{d(T)}(z) = \frac{k-d(T)}{k} f_k(z)$. Consequently, the generating function for the sequences of such trees is $(1 - b_k^{d(T)}(z))^{-1}$. Since we need $s(T)$ of such sequences and the substitution is proper the generating function for \mathcal{F}_k^T is

$$\left(1 - \frac{k-d(T)}{k} f_k(z)\right)^{-s(T)} \cdot k^{\frac{d(T)}{2}} \cdot z^{|T|},$$

where $k^{\frac{d(T)}{2}} = k(k-1) \dots (k-d(T)+1)$. For $s(T) > 0$ and $d(T) < k$ (the remaining cases are trivial) the function is easily seen to have unique dominating singularity of the square root type in $\frac{1}{4k}$ (in the same point that $f_k(z)$). The application of the Lemma 3 yields

$$\mu_k(\mathcal{F}_k^T) = \frac{k^{\frac{d(T)}{2}} \cdot s(T) \cdot \left(1 - \frac{k-d(T)}{2k}\right)^{-s(T)-1} \cdot \frac{k-d(T)}{k}}{(4k)^{|T|}} = \frac{k^{\frac{d(T)}{2}} \cdot s(T) \cdot 2^{s(T)+1}}{(4k)^{|T|}} + O\left(\frac{k^{\frac{d(T)}{2}-1}}{k^{|T|}}\right).$$

This estimation together with the fact that $|T| = d(T) + \text{rep}(T)$ gives the estimation claimed in the Lemma. \square

By a similar reasoning we get the following Lemma.

Lemma 5 For every family scheme T , we have:

$$\mu_k^+(\widehat{\mathcal{F}_k^T}) = \mu_k(\mathcal{F}_k^T) + O\left(\frac{1}{k^{\text{rep}(T)+1}}\right). \quad (1)$$

Proof: The number of elements of $\widehat{\mathcal{F}_k^T}$ of size n is not greater than the number of all substitutions admissible for T of size $n - |T|$. The generating function for these substitutions is $(1 - f_k(z))^{-s(T)} \cdot k^{d(T)}$, hence the generating function for the family $\widehat{\mathcal{F}_k^T}$ is coordinatewise not greater than the function

$$h_k(z) = (1 - f_k(z))^{-s(T)} \cdot k^{d(T)} \cdot z^{|T|}.$$

(i.e. for every $n \in \mathbb{N}$ we have $[z^n]h_k(z) \geq \widehat{\mathcal{F}_k^T}(n)$). Application of the Lemma 3 gives

$$\mu_k^+(\widehat{\mathcal{F}_k^T}) \leq \lim_{n \rightarrow \infty} \frac{[z^n]h_k(z)}{[z^n]f_k(z)} = \frac{s(T)}{2^{2\text{rep}(T)+2d(T)-s(T)-1} \cdot k^{\text{rep}(T)}} + O\left(\frac{1}{k^{\text{rep}(T)+1}}\right).$$

The last equation together with Lemma 4 gives (1). \square

All considered families corresponding to some family scheme have densities. In most cases we omit the proofs of the existence, which are typical but needs a lot of calculations.

2.4 Intuitionistic logic

Within this section we present a simple characterization of the propositional intuitionistic logic. The proof of its equivalence with other definitions and far more general view of the subject can be found in [SU98].

Let τ be the set of open subset of \mathbb{R} with respect to Euclidean topology. The functions $v : \mathcal{V}_k \rightarrow \tau$ are called valuations in τ . We can extend every valuation $v : \mathcal{V}_k \rightarrow \tau$ to the set of all the formulae by the following rule:

$$v[\varphi \rightarrow \psi] := \text{interior}((\mathbb{R} \setminus v(\varphi)) \cup v(\psi)).$$

The following theorem belongs to folklore (we treat this theorem as a definition for the intuitionistic tautologies):

Theorem 6 A formula $\varphi \in \mathcal{T}_k$ is an intuitionistic tautology if and only if for every valuation $v : \mathcal{V}_k \rightarrow \tau$ we have $v[\varphi] = \mathbb{R}$.

It is easy to derive from the statement above that all the intuitionistic tautologies are classical. The converse is not true. To see it, let us analyze the formula $((p \rightarrow q) \rightarrow p) \rightarrow p$ known as Peirce's law. Simple check shows that it is a classical tautology. To show that it is not intuitionistic one, let us consider a valuation $v : \mathcal{V}_k \rightarrow \tau$ such that $v[p] = \mathbb{R} \setminus \{0\}$ and $v[q] = \emptyset$. Then we get

$$v[((p \rightarrow q) \rightarrow p) \rightarrow p] = \mathbb{R} \setminus \{0\}.$$

Such formulae are the main subject of our interest. We denote by $Peirce_k$ the set of formulae from \mathcal{T}_k which are classical but not intuitionistic tautologies, we call them Peirce's formulae.

3 Densities of Peirce's formulae

To estimate the density of Peirce's formulae we need to consider more detailed partitions of the sets of all formulae that the one considered in [FGGZ07]. We need also more precise estimations for the densities of considered families, since the density of Peirce's formulae is at most of the order k^{-2} . (According to the results of [FGGZ07] the order k^{-1} is completely consumed by the simple tautologies). A simple tautology is a formula in which at least one premise is a leaf labelled by the same variable that the goal of the whole formula. Simple tautologies are easily seen to be intuitionistic tautologies (see [FGGZ07]). Note also that if all the goals of the premises of some formula are labelled by different variables than the goal of the formula, then the formula cannot be a classical tautology.

We present several families with estimations of their densities. In the end we use a quantitative argument to show that the set of formulae not belonging to the considered families has low density (i.e. of the order k^{-3}). Therefore we need to consider also families which does not contain any Peirce's formulae.

We start with the family of classical tautologies which are not simple tautologies and which have at least two premises with the same goal as the whole tree. This family was also considered in [FGGZ07]. In the Lemma 7 we give an alternative simple proof that the density of this family is $O(1/k^3)$. Then we analyze the family of non simple tautologies with only one premise A such that A has the same goal as the whole tree and A has at least two premises. We prove in the Lemma 8 that the density of this set is $O(1/k^3)$. In the last step we consider trees as above but with A having exactly one premise. We show how to split this kind of trees into six disjoint sets with high densities. Finally, we use a quantitative argument to prove in Theorem 10 that the density of Peirce's formulae is $1/(2k^2) + O(1/k^3)$. One of the considered set is a family containing only Peirce's formulae (we call them simple Peirce's formulae) and the set of formulae which have not been considered has density of the order k^{-3} .

Lemma 7 *Let \mathcal{G}_k be the set of tautologies, but non simple tautologies $t \in \mathcal{T}_k$ for which at least two premises have goals equal to $r(t)$. We have $\mu_k^+(\mathcal{G}_k) = O(1/k^3)$.*

Proof: Let \mathcal{H}_k be the set of formulae $t \in \mathcal{T}_k$ which are not simple tautologies and which have at least two premises with goal equal to $r(t)$. Let T be the family scheme depicted in Fig. 2. It is easy to see that $\mathcal{F}_k^T \subset \mathcal{H}_k \subset \widehat{\mathcal{F}_k^T}$ and that no element of \mathcal{F}_k^T is a tautology (put $\alpha = \beta = \gamma = 0$ and all the other variables to 1). From the Lemma 4 we know that $\mu_k(\mathcal{F}_k^T) = \frac{7}{4k^2} + O(1/k^3)$. We know that $\mathcal{G}_k \subset \mathcal{H}_k \setminus \mathcal{F}_k^T$. From the Lemma 5 we get $\mu_k^+(\mathcal{H}_k \setminus \mathcal{F}_k^T) = O(1/k^3)$, which proves the Lemma. \square

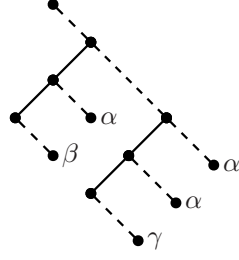
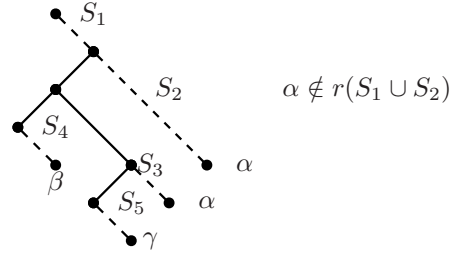
Let \mathcal{S}_k be the family of trees from \mathcal{T}_k such that each tree $t \in \mathcal{S}_k$ satisfies the following conditions:

- t is not a simple tautology, i.e. no premise of t is a leaf labelled with $r(t)$,
- t has exactly one premise, say A , such that $r(A) = r(t)$ and A has at least two premises.

Lemma 8 *Let \mathcal{G}_k^2 be the set of tautologies belonging to \mathcal{S}_k . We have $\mu_k^+(\mathcal{G}_k^2) = O(1/k^3)$.*

Proof: Family \mathcal{S}_k can be constructed from the scheme in Fig. 3 by substitutions in which no tree from the sequences substituted for S_1, S_2 has the goal labelled with the variable assigned to α .

We show that densities of tautologies from \mathcal{S}_k are $O(1/k^3)$, by constructing a large subfamily of non tautologies. First, we estimate the density of \mathcal{S}_k . The generating function for \mathcal{S}_k can be easily found and


 Fig. 2: A tree with at least two premises with a goal α .

 Fig. 3: The scheme corresponding to the family \mathcal{S}_k .

it equals $g_{\mathcal{S}_k}(z) = k^3 z^4 (1 - b_k^1(z))^{-2} (1 - f_k(z))^{-3}$. Therefore, we get

$$\mu_k(\mathcal{S}_k) = \frac{5}{4k} - \frac{7}{2k^2} + O(1/k^3). \quad (2)$$

We use the scheme from Fig. 3 to define four disjoint subfamilies of non tautologies by imposing restrictions on the allowed substitutions. For a substitution $((S_1, S_2, S_3, S_4, S_5), (\alpha, \beta, \gamma))$ we consider the following cases (we abuse the notation using the same names for the scheme variables (resp. scheme edges) and variables (resp. sequences) assigned to them by the substitution):

- (a) $\beta = \alpha, \gamma \neq \alpha, \alpha, \gamma \notin r(S_1) \cup \dots \cup r(S_5)$,
- (b) $\beta \neq \alpha, \beta \notin r(S_1) \cup r(S_2) \cup r(S_4), \alpha \notin r(S_1) \cup r(S_2)$ (no restrictions for γ),
- (c) $\beta \neq \alpha, \beta$ occurs exactly once among the goals of trees from $S_1, S_2, S_4, \beta \notin r(S_3) \cup r(S_5), \alpha, \gamma \notin r(S_1) \cup \dots \cup r(S_5)$,
- (d) $\beta \neq \alpha, \alpha$ occurs exactly once among the goals of trees from $S_4, \alpha \notin r(S_1) \cup r(S_2) \cup r(S_3) \cup r(S_5), \gamma \notin r(S_1) \cup \dots \cup r(S_5)$.

We denote by $\mathcal{S}_k^a, \mathcal{S}_k^b, \mathcal{S}_k^c, \mathcal{S}_k^d$ the families of trees from \mathcal{T}_k constructed from the scheme from Fig. 3 and substitutions fulfilling corresponding condition (a), (b), (c), (d). Each of the above sets contains only non tautologies. To falsify the elements of the families $\mathcal{S}_k^a, \mathcal{S}_k^c, \mathcal{S}_k^d$, valuate α and γ to 0 and all the other variables to 1. For the family \mathcal{S}_k^b it suffices to put $\beta = \alpha = 0$ and all the other variables to 1.

The family \mathcal{S}_k^a is easily seen to be defined by some family scheme T_a with parameters $s = 5, d = 2$ and $rep = 2$ (substitute β with α in the scheme from Fig. 3). By the Lemma 4 the density of \mathcal{S}_k^a is

$$\frac{5}{4k^2} + O(1/k^3). \quad (3)$$

For the family \mathcal{S}_k^b we need more accurate estimation. The generating function for \mathcal{S}_k^b is

$$g_{\mathcal{S}_k^b}(z) = k^2(k-1)z^4 \frac{1}{(1 - b_k^2(z))^3} \frac{1}{(1 - f_k(z))^2}.$$

Thus, the density of \mathcal{S}_k^b is

$$\frac{5}{4k} - \frac{47}{4k^2} + O(1/k^3). \quad (4)$$

The family \mathcal{S}_k^c is a disjoint union of three families corresponding to the schemes depicted in Fig. 4. Formally, let T_{c1}, T_{c2}, T_{c3} be the family schemes depicted in Fig. 4, then it is easy to see that

$$\left(\mathcal{F}_k^{T_{c1}} \cup \mathcal{F}_k^{T_{c2}} \cup \mathcal{F}_k^{T_{c3}} \right) \subset \mathcal{S}_k^c \subset \left(\widehat{\mathcal{F}_k^{T_{c1}}} \cup \widehat{\mathcal{F}_k^{T_{c2}}} \cup \widehat{\mathcal{F}_k^{T_{c3}}} \right).$$

Each of this family schemes has parameters $s = 7, d = 3, rep = 2$. By the Lemma 4 and the Lemma 5 we get that

$$\mu_k(\mathcal{S}_k^c) = 3 \cdot \frac{7}{4k^2} + O(1/k^3). \quad (5)$$

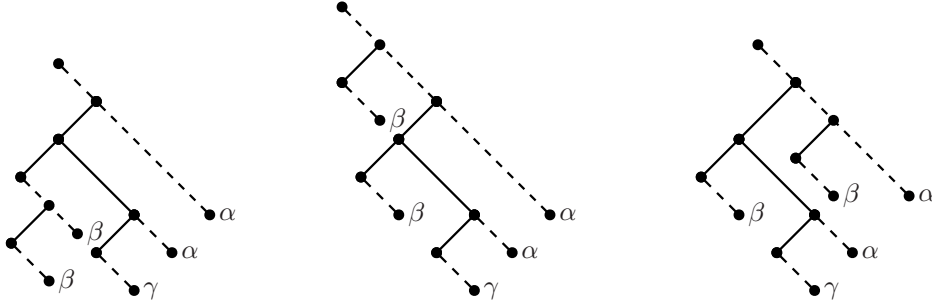


Fig. 4: Scheme for trees in case (c).

The last case – (d) corresponds to the scheme T_d from Fig. 5. We have $\mathcal{F}_k^{T_d} \subset \mathcal{S}_k^d \subset \widehat{\mathcal{F}_k^{T_d}}$. The family scheme has parameters $s = 7, d = 3, rep = 2$, hence, by the Lemma 4 and the Lemma 5, we get

$$\mu_k(\mathcal{S}_k^d) = \frac{7}{4k^2} + O(1/k^3). \quad (6)$$

If we compare the density of \mathcal{S}_k with the density of $\mathcal{S}_k^a \cup \mathcal{S}_k^b \cup \mathcal{S}_k^c \cup \mathcal{S}_k^d$ we get

$$\frac{5}{4k} - \frac{7}{2k^2} - \frac{5}{4k^2} - \frac{5}{4k} + \frac{47}{4k^2} - 3 \cdot \frac{7}{4k^2} - \frac{7}{4k^2} + O(1/k^3) = O(1/k^3).$$

Since no tautology from \mathcal{S}_k belongs to $\mathcal{S}_k^a \cup \mathcal{S}_k^b \cup \mathcal{S}_k^c \cup \mathcal{S}_k^d$, we have $\mu_k^+(\mathcal{G}_k^2) = O(1/k^3)$. \square

From now on we are going to consider trees t with exactly one premise A with $r(t) = r(A)$ and for which the premise A has only one premise. Such a family of trees, denoted by \mathcal{T}_k^{11} , corresponds to the scheme T_{11} presented in Fig. 6.

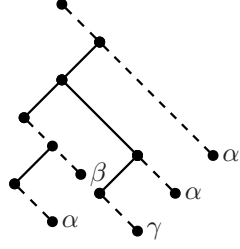
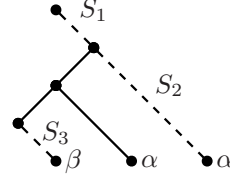


Fig. 5: Scheme for trees in case (d).


 Fig. 6: Scheme corresponding to the family \mathcal{T}_k^{11} .

The generating function for \mathcal{T}_k^{11} is $g_{\mathcal{T}_k^{11}}(z) = k^2 z^3 (1 - b_k^1(z))^{-2} (1 - f_k(z))^{-1}$, thus its density equals

$$\mu_k(\mathcal{T}_k^{11}) = \frac{3}{4k} - \frac{5}{2k^2} + O(1/k^3). \quad (7)$$

We divide the family \mathcal{T}_k^{11} into six disjoint subfamilies with large densities. One of these subfamilies, denoted by \mathcal{P}_k , will contain only Peirce's formulae, we call the elements of \mathcal{P}_k simple Peirce's formulae.

The subfamilies, mentioned above, arise by putting some restrictions for the substitutions allowed for the scheme T_{11} . For a substitution $((S_1, S_2, S_3), (\alpha, \beta))$ (again we abuse the notation) we consider the following four cases:

- (a) $\beta = \alpha$, $\alpha \notin r(S_1) \cup r(S_2) \cup r(S_3)$,
- (b) $\beta \neq \alpha$, $\alpha, \beta \notin r(S_1) \cup r(S_2) \cup r(S_3)$,
- (c) $\beta \neq \alpha$, $\beta \notin r(S_1) \cup r(S_2) \cup r(S_3)$, α occurs exactly once among the goals of trees from S_3 , $\alpha \notin r(S_1) \cup r(S_2)$,
- (d) $\beta \neq \alpha$, $\alpha \notin r(S_1) \cup r(S_2) \cup r(S_3)$ and β occurs exactly once among the goals of the trees from S_1, S_2, S_3 .

We denote by \mathcal{NT}_k^a the set of formulae from \mathcal{T}_k which can be constructed from T_{11} by substitutions fulfilling the condition (a) (the set \mathcal{NT}_k^b is defined analogously).

Let T_a be the scheme from Fig. 6 with scheme variable β replaced by α . We have $\mathcal{NT}_k^a = \mathcal{F}_k^{T_a}$, hence, by the Lemma 4, the density of \mathcal{NT}_k^a is

$$\mu_k(\mathcal{NT}_k^a) = \frac{3}{4k^2} + O(1/k^3). \quad (8)$$

The formulae from \mathcal{NT}_k^a are easily seen not to be tautologies (valuate α to 0 and all the other variables to 1).

The formulae from \mathcal{NT}_k^b are also non tautologies (to falsify them put $\alpha = \beta = 0$ and all the other variables to 1). We need more accurate estimation for the densities $\mu_k(\mathcal{NT}_k^b)$ then the one provided by the Lemma 4. The generating function for \mathcal{NT}_k^b is $g_{\mathcal{NT}_k^b}(z) = k(k-1)z^3(1 - b_k^2(z))^{-3}$. With simple computations we get

$$\mu_k(\mathcal{NT}_k^b) = \frac{3}{4k} - \frac{33}{4k^2} + O(1/k^3). \quad (9)$$

The family of formulae which can be constructed from the scheme T_{11} by the substitutions fulfilling the condition (c) is divided into two sets. In every such formula t there is only one tree A in S_3 such that $r(A) = \alpha$. The first family, \mathcal{P}_k , (see its corresponding scheme in Fig. 7) is defined by substitutions for which the subformula A is just a variable α . The second one, \mathcal{NT}_k^c , (see Fig. 7) is defined by the remaining substitutions fulfilling the condition (c).

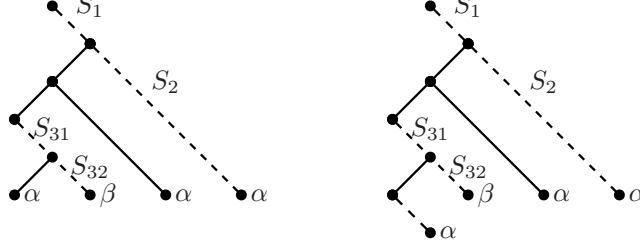


Fig. 7: Scheme of the family \mathcal{P}_k (left one) and the family \mathcal{NT}_k^c (right one).

The set \mathcal{P}_k turns out to consist of Peirce's formulae. Indeed, it suffices to consider only valuations for the variable α to see that every formula from \mathcal{P}_k is a classical tautology. To see that each $t \in \mathcal{P}_k$ is not an intuitionistic tautology check the following valuation in τ , let $v[\alpha] = \mathbb{R} \setminus \{0\}$, $v[\beta] = \emptyset$ and $v[z] = \mathbb{R}$ for all other variables. It gives $v[t] = \mathbb{R} \setminus \{0\}$. Trees from the second set \mathcal{NT}_k^c are non tautologies (put $\alpha = \beta = 0$ and all the other variables to 1). By the Lemma 4 we have

$$\mu_k(\mathcal{P}_k) = \frac{1}{2k^2} + O(1/k^3). \quad (10)$$

Let T_c be the scheme on the right in Fig. 7. We have $\mathcal{F}_k^{T_c} \subset \mathcal{NT}_k^c \subset \widehat{\mathcal{F}_k^{T_c}}$, therefore by Lemma 4 and 5 we get

$$\mu_k(\mathcal{NT}_k^c) = \frac{3}{4k^2} + O(1/k^3). \quad (11)$$

For the case (d) we also consider two sets of trees. The first one, \mathcal{IT}_k , is defined by the substitutions fulfilling the condition (d) for which the unique tree from S_1, S_2, S_3 with a goal labelled by β is a leaf (see its schemes in Fig. 8). The second one, \mathcal{NT}_k^d , is defined by the remaining substitutions fulfilling the condition (d) (see its corresponding schemes in Fig. 9). Trees in \mathcal{NT}_k^d are not tautologies (put $\alpha = \beta = \gamma = 0$ and all the other variables to 1). On the other hand, each tree in \mathcal{IT}_k is an intuitionistic tautology. We prove this fact in the Lemma 9.

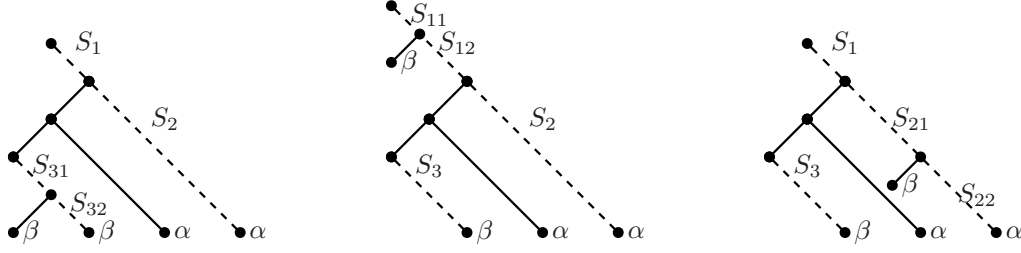
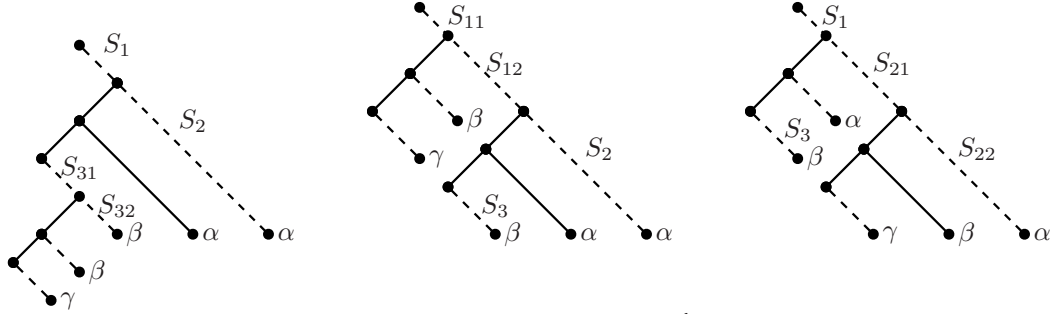
The same reasoning as in the previous cases (involving application of Lemma 4 and Lemma 5) yields:

$$\mu_k(\mathcal{IT}_k) = 3 \cdot \frac{1}{2k^2} + O(1/k^3), \quad (12)$$

$$\mu_k(\mathcal{NT}_k^d) = 3 \cdot \frac{3}{4k^2} + O(1/k^3). \quad (13)$$

Lemma 9 Each tree from the set \mathcal{IT}_k is an intuitionistic tautology.

Proof: The set \mathcal{IT}_k is a disjoint union of three families defined by the family schemes depicted in Fig. 8. Let $\mathcal{IT}_k^1, \mathcal{IT}_k^2, \mathcal{IT}_k^3$ denote the families corresponding to the consecutive schemes from the picture.


 Fig. 8: Scheme of the family \mathcal{IT}_k .

 Fig. 9: Scheme of the family \mathcal{NT}_k^d .

Let $t \in \mathcal{IT}_k^1$. There is a premise A in t whose goal is labelled by the same variable that $r(T)$ and whose only premise is a simple tautology. It means that each valuation in τ evaluates this premise to \mathbb{R} . But then, for every valuation in τ , the premise A is evaluated to $v[r(T)]$. Let $A \rightarrow B$ be the subformula of t corresponding to the parent of A . Note that the valuation of implication is increasing with respect to the second argument (i.e. $v[\phi \rightarrow \psi] \supseteq v[\psi]$). For every valuation v in τ we get $v[B] \supseteq v[r(T)]$ and since $v[A] = v[r(t)]$ we get $v[A \rightarrow B] = \mathbb{R}$. This value is then propagated to the root of the tree (by the increasing property), which means that $v[t] = \mathbb{R}$.

It is easy to check that for every valuation v in τ we have $v[\alpha \rightarrow (\beta \rightarrow \gamma)] = v[\beta \rightarrow (\alpha \rightarrow \gamma)]$. Therefore it is enough to show that the elements of \mathcal{IT}_k^3 are not intuitionistic tautologies. Suppose that $t \in \mathcal{IT}_k^3$ and let $A \rightarrow B$ be the subformula of t such that A is the premise of t with $r(A) = r(t)$. Then, for every valuation v we have $v[B] \supseteq \text{interior}(\mathbb{R} \setminus v(\beta) \cup v(\alpha))$ and $v[A] \subseteq \text{interior}(\mathbb{R} \setminus v(\beta) \cup v(\alpha))$. It gives $v[A] \subseteq v[B]$, but then $v[A \rightarrow B] = \mathbb{R}$ and (again by the monotonicity of the valuation of implication) we get $v[t] = \mathbb{R}$. \square

Theorem 10 For every $k \in \mathbb{N}$ let Peirce_k denote the set of formulae from \mathcal{T}_k which are classical but not intuitionistic tautologies. Then we have

$$\mu_k^+(\text{Peirce}_k) \sim_k \mu_k^-(\text{Peirce}_k) = \frac{1}{2k^2} + O(1/k^3).$$

Proof: Each formula $t \in \text{Peirce}_k$ is a tautology. Therefore it must have at least one premise with goal equal to the goal of t . Moreover, since t is not an intuitionistic tautology, it cannot be a simple tautology, i.e. it has no premise equal to its goal.

We have already found a large set of Peirce's formulae – the family \mathcal{P}_k (see Fig. 7). We know (see the equation (10)) that $1/(2k^2) + O(1/k^3) = \mu(\mathcal{P}_k) \leq \mu_k^-(Peirce_k)$. To find the upper bound for the density $\mu_k^+(Peirce_k)$ we show that the density of the set $Peirce_k \setminus \mathcal{P}_k$ is relatively small.

Let $Peirce_k^1$ be the set of all the formulae $t \in Peirce_k$ which have exactly one premise A with $r(A) = r(t)$ and let $Peirce_k^2$ denote the set of all $t \in Peirce_k$ which have at least two premises A and B such that $r(A) = r(B) = r(t)$. Obviously, $Peirce_k = Peirce_k^1 \cup Peirce_k^2$ and by the Lemma 7 the density of $Peirce_k^2$ is small – we have $\mu_k^+(Peirce_k^2) = O(1/k^3)$.

The set $Peirce_k^1$ can be split further into two subsets. Let $Peirce_k^{11}$ denote the set of all $t \in Peirce_k^1$ with only one premise A such that $r(A) = r(t)$ and A has exactly one premise. Let $Peirce_k^{12} = Peirce_k^1 \setminus Peirce_k^{11}$ (in the elements of $Peirce_k^{12}$ the subformula A has at least two premises.) By the Lemma 8 the density $\mu_k^+(Peirce_k^{12})$ is $O(1/k^3)$.

To estimate the density of $Peirce_k^{11}$ we need to consider all the trees $t \in \mathcal{T}_k$ with only one premise A such that $r(A) = r(t)$, $A \neq r(t)$ (it cannot be a simple tautology) and A has exactly one premise. Such a family is denoted by \mathcal{T}_k^{11} (see Fig. 6). As we know, \mathcal{T}_k^{11} can be split into several subsets (see Fig. 7, 8 and 9) and $\mathcal{T}_k^{11} = \mathcal{N}\mathcal{T}_k^a \cup \mathcal{N}\mathcal{T}_k^b \cup \mathcal{N}\mathcal{T}_k^c \cup \mathcal{N}\mathcal{T}_k^d \cup \mathcal{P}_k \cup \mathcal{I}\mathcal{T}_k \cup Rest_k$. No set $\mathcal{N}\mathcal{T}_k^i$ (for $i = a, b, c, d$) contains a tautology, \mathcal{P}_k consists of Peirce's formulae, each formula in $\mathcal{I}\mathcal{T}_k$ is an intuitionistic tautology (see the scheme in Fig. 8 and the Lemma 9) and $Rest_k$ denotes the set of remaining trees. Therefore we can write that

$$Peirce_k^{11} \subseteq \mathcal{P}_k \cup Rest_k.$$

Since $\mu_k(Rest_k) = \mu_k(\mathcal{T}_k^{11}) - \mu_k(\mathcal{N}\mathcal{T}_k^a) - \mu_k(\mathcal{N}\mathcal{T}_k^b) - \mu_k(\mathcal{N}\mathcal{T}_k^c) - \mu_k(\mathcal{N}\mathcal{T}_k^d) - \mu(\mathcal{P}_k) - \mu(\mathcal{I}\mathcal{T}_k)$, using the equations (7), (8), (9), (10), (11), (12) and (13), we find that $\mu(Rest) = O(1/k^3)$.

Finally, we can estimate the density of all Peirce's formulae. We know that

$$Peirce_k \subseteq \mathcal{P}_k \cup Rest_k \cup Peirce_k^{12} \cup Peirce_k^2$$

and the densities of $Rest_k$, $Peirce_k^{12}$ and $Peirce_k^2$ are small, i.e. each density is $O(1/k^3)$. It gives $\mu_k^+(Peirce_k) = 1/(2k^2) + O(1/k^3)$. \square

4 Final remarks

Although we did not address directly the problem of the existence of the densities of Peirce's formulae, the presented technique can be used to obtain better upper and lower bounds, by the systematic analysis of more detailed partitions.

The intuitionistic logic can be also defined as the set of formulae which are true in every finite Kripke structure. Therefore, a formula is not an intuitionistic tautology, if it can be falsified in some finite Kripke structure. Interestingly, the family \mathcal{P}_k we considered, consists of classical tautologies which can be falsified in the Kripke structure of size 2. It is a minimal size for which the difference between classical and intuitionistic logics can be observed. We know also that the family of formulae which needs Kripke structure of the size 3 to be falsified, has density of the order at most k^{-3} . It is not hard to prove (using Drmota-Lalley-Woods theorem, see [FS08]) that for every $k, m \in \mathbb{N}$ the set of formulae from \mathcal{T}_k which can be falsified in some Kripke structure of the size m has a density. It is interesting to estimate the density of Peirce's formulae which needs a structure of size m to be falsified. It seems also that this approach can be used to prove the existence of density of all Peirce's formulae.

References

- [CFGG04] B. Chauvin, P. Flajolet, D. Gardy, and B. Gittenberger. And/Or trees revisited. *Combinatorics, Probability and Computing*, 13(4-5):475–497, July-September 2004.
- [FGGZ07] H. Fournier, D. Gardy, A. Genitrini, and M. Zaionc. Classical and intuitionistic logic are asymptotically identical. In *CSL*, pages 177–193, 2007.
- [FS08] P. Flajolet and R. Sedgewick. Analytic combinatorics. in preparation, preprint available at <http://algo.inria.fr/flajolet/Publications/book.pdf>, 2008.
- [Koz08] J. Kozik. Subcritical pattern languages for and/or trees. 2008. Fifth Colloquium on Mathematics and Computer Science.
- [LS97] H. Lefmann and P. Savický. Some typical properties of large And/Or Boolean formulas. *Random Structures and Algorithms*, 10:337–351, 1997.
- [MTZ00] M. Moczurad, J. Tyszkiewicz, and M. Zaionc. Statistical properties of simple types. *Mathematical Structures in Computer Science*, 10(5):575–594, 2000.
- [SU98] M. Sorensen and P. Urzyczyn. Lectures on the curry-howard isomorphism, 1998.