Z. Kostrzycka
M. Zaionc

Asymptotic densities in logic and type theory

Abstract. This paper presents a systematic approach for obtaining results from the area of quantitative investigations in logic and type theory. We investigate the proportion between tautologies (inhabited types) of a given length n against the number of all formulas (types) of length n. We investigate an asymptotic behavior of this fraction. Furthermore, we characterize the relation between number of premises of implicational formula (type) and the asymptotic probability of finding such formula among the all ones. We also deal with a distribution of these asymptotic probabilities. Using the same approach we also prove that the probability that randomly chosen fourth order type (or type of the order not greater than 4), which admits decidable lambda definability problem, is zero.

Keywords: propositional logic, asymptotic density of tautologies, probabilistic methods in logic and type theory.

1. Introduction

The research described in this paper is a part of the project of quantitative investigations in logic and type theory. This paper summarizes the research in which we develop methods of finding the asymptotic probability in some propositional logics. Probabilistic methods appear to be very powerful in combinatorics and computer science. From a point of view of these methods we investigate a typical object chosen from some set. For propositional formulas, we investigate the proportion between the number of valid formulas of a given length n against the number of all formulas of length n. Our interest lays in finding the limit of that fraction when $n \to \infty$. If the limit exists, then it is represented by a real number which we may call *the density of truth* of the investigated logic. In general, we are also interested in finding the *'density'* of some other classes of formulas. Good presentation and overview of asymptotic methods for random boolean expressions can be found in the paper [4] of Gardy.

We assume that the set of formulas $\mathcal{F}$ of a given propositional calculus is equipped with a norm $\|.\|$ which is a function $\|.\| : \mathcal{F} \mapsto \mathbb{N}$. Moreover, we assume that for any n the set of formulas $\{\phi \in \mathcal{F} : \|\phi\| = n\}$ is finite.

Presented by **Name of Editor**; *Received* December 1, 2006

Typical norms are presented in Definitions 6 and 19. In Definition 6 the norm $\|\phi\|$ means the total number of appearances of propositional variables in the formula ϕ, while in Definition 19, $\|\phi\|$ is the number of all characters (without parentheses) in formula ϕ.

Definition. 1. *We associate the density $\mu(\mathcal{A})$ with a subset $\mathcal{A}$ of formulas as follows:*

$$\mu(\mathcal{A}) = \lim_{n\to\infty} \frac{\#\{t \in \mathcal{A} : \|t\| = n\}}{\#\{t \in \mathcal{F} : \|t\| = n\}} \tag{1}$$

if the appropriate limit exists.

The number $\mu(\mathcal{A})$ if exists is an asymptotic probability of finding a formula from the set $\mathcal{A}$ among all formulas. It may be also interpreted as the asymptotic density of the set $\mathcal{A}$. It can be immediately seen that the density μ is finitely additive. So, if $\mathcal{A}$ and $\mathcal{B}$ are disjoined classes of formulas such that $\mu(\mathcal{A})$ and $\mu(\mathcal{B})$ exist then $\mu(\mathcal{A} \cup \mathcal{B})$ also exists and

$$\mu(\mathcal{A} \cup \mathcal{B}) = \mu(\mathcal{A}) + \mu(\mathcal{B}). \tag{2}$$

It is straightforward to observe that for any finite set $\mathcal{A}$ the density $\mu(\mathcal{A})$ exists and is 0 and dually for co-finite sets $\mathcal{A}$ the density $\mu(\mathcal{A}) = 1$. Unfortunately, the density μ is not countably additive, so in general, the formula (3) given below

$$\mu\left(\bigcup_{i=0}^{\infty} \mathcal{A}_i\right) = \sum_{i=0}^{\infty} \mu\left(\mathcal{A}_i\right) \tag{3}$$

is not true for all pairwise disjoint classes of sets $\{\mathcal{A}_i\}_{i\in\mathbb{N}}$. The good counterexample for the equation (3) is to take as $\mathcal{A}_i$ the singleton consisting of the $i-th$ formula from our language. On the left hand side of (3) we get 1 but on right hand side $\mu\left(\mathcal{A}_i\right) = 0$ for all $i \in \mathbb{N}$.

Definition. 2. *By a random variable X we understand the function*

$$X : \mathcal{F} \mapsto \mathbb{N}$$

which assigns a number $n \in \mathbb{N}$ to a formula in such a way that for any n the density $\mu\left(\{\phi : X(\phi) = n\}\right)$ exists and moreover

$$\sum_{n=0}^{\infty} \mu\left(\{\phi : X(\phi) = n\}\right) = 1.$$

Definition. 3. *By a distribution of a random variable* X *we mean the function* $\overline{X} : \mathbb{N} \mapsto \mathbb{R}$ *defined by:*

$$\overline{X} \ : \ \mathbb{N} \ni n \mapsto \mu\left(\{\phi : X(\phi) = n\}\right) \in \mathbb{R}$$

Definition. 4. *The expected value, variance and standard deviation are defined in the conventional way by:*

$$\begin{aligned} E(X) &= \sum_{p=0}^{\infty} p \cdot \overline{X}(p), & (4)\\ D^2(X) &= E((X - E(X))^2) = E(X^2) - (E(X))^2, & (5)\\ &= \sum_{p=0}^{\infty} p^2 \cdot \overline{X}(p) - (E(X))^2, \end{aligned}$$

so the standard deviation of X *is* $\sqrt{\mathcal{D}^2(X)}$

2. Generating functions

In the whole paper we present some properties of numbers characterizing the amount of formulas in different classes and languages, and we are concerned with the asymptotic behavior of these numbers. The main tool for dealing with asymptotics of sequences of numbers are known in combinatorics as *generating functions.* A nice exposition of this method can be found in [14], [1] and [2]. See also papers [16], [15], [9] [11] for the presentation of this method in logics.
Suppose that we have a system of non-linear equations $\overrightarrow{y_j} = \Phi_j(z, y_1, ...y_m)$ for $1 \leq j \leq m$, where any $y_j = \sum_{n=0}^{\infty} a_j z^n$. The following result known as Drmota-Lalley-Woods theorem (see [2], Thm. 8.13, p.71) is of great importance in the both cases of solving the system explicitly or implicitly.

Theorem. 5. *Consider a nonlinear polynomial system, defined by a set of equations*

$$\{\overrightarrow{y} = \Phi_j(z, y_1, ..., y_m)\}, \quad 1 \leq j \leq m$$

satisfying the following properties:

1. *a-properness:* Φ *is a contraction, i.e. satisfies the Lipschitz condition*

$$d(\Phi(y_1, ..., y_m), \Phi(y'_1, ..., y'_m)) < K d((y_1, ..., y_m), (y'_1, ..., y'_m)), \ \ K < 1$$

2. a-positivity: all terms of the series $\Phi_j(\overrightarrow{y})$ are ≥ 0

3. a-irreducibility: the dependency graph of the algebraic system is built on m vertices: 1,2,...,m; there is an edge from a vertex k to a vertex j if y_j appears in Φ_k. The algebraic system is a irreducible if its dependency graph is strongly connected.

4. a-aperiodicity: z (not z^2 or z^3...) is the right variable, that means for each y_j there exist three monomials z^a, z^b, and z^c such that $b-a$ and $c-a$ are relatively prime

Then

1. All component solutions y_i have the same radius of convergence $\rho < \infty$.

2. There exist functions h_j analytic at the origin such that

$$y_j = h_j(\sqrt{1 - z/\rho}), \quad (z \to \rho^-). \tag{6}$$

3. All other dominant singularities are of the form $\rho\omega$ with ω being a root of unity.

4. If the system is a-aperiodic then all y_j have ρ as unique dominant singularity. In that case, the coefficients admit a complete asymptotic expansion of the form:

$$[z^n]y_j(z) \sim \rho^{-n} \left(\sum_{k \geq 1} d_k n^{-1-k/2} \right). \tag{7}$$

From (7) there is a simple transition by the so called transfer lemma from [3], to a formula defining the value of the coefficients $[z^n]y_j(z)$. So, the a-aperiodicity of a system of equations is a very desirable property. The application of the above theorem will proceed in the following way. Suppose that we have two functions f_T and f_F enumerating the tautologies of some logic and all formulas. Suppose they have the same dominant singularity ρ and there are the suitable constants α_1, α_2, β_1, β_2 such that:

$$f_T(z) = \alpha_1 - \beta_1\sqrt{1 - z/\rho} + O(1 - z/\rho), \tag{8}$$

$$f_F(z) = \alpha_2 - \beta_2\sqrt{1 - z/\rho} + O(1 - z/\rho). \tag{9}$$

Then the *density of truth* (probability that a random formula is a tautology) is given by:

$$\mu(T) = \lim_{n\to\infty} \frac{[z^n]f_T(z)}{[z^n]f_F(z)} = \frac{\beta_1}{\beta_2}. \tag{10}$$

We apply the approach in two examples in the next section.

3. Densities of implicational reducts of logics

In this section we present some results obtained in [11], [16], and [6] characterizing the density of tautologies in a language with the only connective of implication.

DEFINITION. 6. *The set of formulas $\mathcal{F}_k^{\rightarrow}$ over k propositional variables is a minimal set consisting of these variables and closed under implication. In this definition the norm $\|.\|$ measures the total number of appearances of propositional variables in the formula. The set of formulas of length n is denoted by F_n^k.*

The number of formulas in F_n^k is finite for any $n \in N$ and will be denoted by $\left|F_n^k\right|$. From Definition 6 we see that any formula from $\mathcal{F}_k^{\rightarrow}$ may be interpreted as a binary planar tree with the internal nodes labeled by the operator $\rightarrow$, and the external ones by propositional variables. Then we have immediately:

LEMMA. 7. *The numbers $|F_n^k|$ are given by the following recursion:*

$$|F_0^k| = 0, \; |F_1^k| = k, \tag{11}$$

$$|F_n^k| = \sum_{i=1}^{n-1} |F_i^k||F_{n-i}^k|. \tag{12}$$

PROOF. Obvious. ■

LEMMA. 8. *The generating function f_{F^k} for the numbers F_n^k is the following:*

$$f_{F^k}(z) = \frac{1-\sqrt{1-4kz}}{2}. \tag{13}$$

PROOF. From the recurrence (12) we see that the generating function f_{F^k} has to fulfil the following equation:

$$f_{F^k}(z) = f_{F^k}^2(z) + kz. \tag{14}$$

By solving (14) with the boundary condition $f_{F^k}(0) = 0$ we obtain (13). ■

Now, we study the case when $k = 1$. Let us notice that in this case the numbers F_n^1 are the well-known Catalan numbers (see [2], [11] and [14]). In the paper [11] it is shown that:

LEMMA. 9. *The Lindenbaum algebra $AL(Cl_1^{\rightarrow})$ for the implicational reduct of classical logic of one variable consists of the following two classes:*

$$N^1 = [p]_{\equiv},$$
$$T^1 = [p \rightarrow p]_{\equiv}.$$

The truth-table for the algebra $AL(Cl_1^{\rightarrow})$ is as follows:

$\rightarrow$	N^1	T^1
N^1	T^1	T^1
T^1	N^1	T^1

Table 1

From Table 1 we obtain a suitable recurrence for the numbers T_n^1, N_n^1, which may be transformed into the following system of functional equations:

$$f_{T^1}(z) = f_{F^1}(z) f_{T^1}(z) + f_{N^1}^2(z), \tag{15}$$
$$f_{N^1}(z) = f_{N^1}(z) f_{T^1}(z) + z. \tag{16}$$

From finite additivity (2) we know that $f_{T^1} = f_{F^1} - f_{N^1}$. We are able to solve the system explicitly (compare [11] Lemma 7.4) and:

LEMMA. 10. *The generating function f_{T^1} is the following:*

$$f_{T^1}(z) = \frac{3 - \sqrt{1-4z} - \sqrt{2 + 2\sqrt{1-4z} + 12z}}{4}. \tag{17}$$

To take advantage of Theorem 5 and the formula (10), let us notice that the system of equations (15)-(16) is a-proper, a-positive, a-irreducible and a-aperriodic. Hence the functions: f_{F^1}, f_{T^1} and f_{N^1} have the unique dominant singularity $z_0 = 1/4$. The expansions of f_{F^1} and f_{T^1} around $z_0 = 1/4$ are the following:

$$f_{T^1}(z) = \frac{3-\sqrt{5}}{4} - \left(1 + \frac{\sqrt{5}}{5}\right)\sqrt{1-4z} + O(1-4z), \tag{18}$$
$$f_{F^1}(z) = \frac{1}{2} - 2\sqrt{1-4z} + O(1-4z). \tag{19}$$

THEOREM. 11. *The density of the class $Cl_1^{\rightarrow}$ is the following*[1]*:*

$$\mu(Cl_1^{\rightarrow}) = \frac{-(1+\frac{\sqrt{5}}{5})}{-2} = \frac{1}{2} + \frac{\sqrt{5}}{10} \approx 0.7236\ldots\ . \tag{20}$$

[1] The number $\mu(Cl_1^{\rightarrow})$ is closely related to the golden ratio. Namely: golden ratio= $5\mu(Cl_1^{\rightarrow}) - 2$.

The above theorem describes the asymptotic density of the set of tautologies in the simplest implicational language, which is the language of one propositional variable (see [11]). The natural inspiration for this research comes from the typed lambda calculus, in which the set of simple types can be identified under the Curry-Howard isomorphism with a set of implicational formulas. Under this isomorphism the class of provable formulas can be understood as the class of inhabited types. Notice that the density of provable formulas in this language is surpassingly hight. Notice also that the classical tautologies in this language coincide with the intuitionistic ones.

THEOREM. 12. *(see [11] page 592) For $k = 1$ the asymptotic density of the set of intuitionisticaly provable formulas $\mathcal{T}_1^{\rightarrow}$ exists and is:*

$$\mu(\mathcal{T}_1^{\rightarrow}) = \frac{1}{2} + \frac{\sqrt{5}}{10} \approx 0.7236\ldots\ .$$

THEOREM. 13. *Implicational classical and intuitionistic logics of one variable are identical.*

A proof by counting can be found in [15]. It is based on the fact that generating functions for the classical and intuitionistic logics are identical.

Analogously, we may consider the implicational classical and intuitionistic logics with two propositional variables, which now are different. We will show however, that from the quantitative point of view, the difference is not significant. We will give a detailed approach in the case of classical logic $Cl_2^{\rightarrow}$; see [6].

LEMMA. 14. *The Lindenbaum algebra $AL(Cl_2^{\rightarrow})$ (presented in the form of diagram) for the implicational reduct of classical logic of two variables consists of the following six classes:*

$$\begin{aligned} A &= [p]_{\equiv}, \\ B &= [q]_{\equiv}, \\ C &= [p \rightarrow q]_{\equiv}, \\ D &= [q \rightarrow p]_{\equiv}, \\ E &= [(p \rightarrow q) \rightarrow q]_{\equiv}, \\ T^2 &= [p \rightarrow p]_{\equiv}. \end{aligned}$$

Below the diagram of the algebra $AL(Cl_2^{\rightarrow})$ is presented

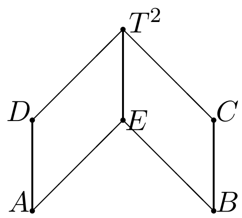

Diagram 1.

The appropriate truth-table is the following:

$\rightarrow$	A	B	C	D	E	T^2
A	T^2	C	C	T^2	T^2	T^2
B	D	T^2	T^2	D	T^2	T^2
C	A	E	T^2	D	E	T^2
D	E	B	C	T^2	E	T^2
E	D	C	C	D	T^2	T^2
T^2	A	B	C	D	E	T^2

Table 2.

From Table 2 we obtain a suitable recurrence (for the numbers T_n^2, A_n, B_n, C_n, D_n, E_n). It may be transformed into the following system of six functional equations:

$$f_A = f_A(f_C + f_{T^2}) + z, \tag{21}$$
$$f_B = f_B(f_D + f_{T^2}) + z, \tag{22}$$
$$f_C = f_C(f_D + f_E + f_{T^2} + f_A) + f_B(f_A + f_E) \tag{23}$$
$$f_D = f_D(f_B + f_C + f_{T^2} + f_E) + f_A(f_B + f_E) \tag{24}$$
$$f_E = f_E(f_C + f_D + f_{T^2}) + f_B f_C \tag{25}$$
$$f_{T^2} = f_{F^2} f_{T^2} + f_E(f_A + f_B + f_E) + f_D(f_A + f_D) + f_C(f_B + f_C) + +f_B^2 + f_A^2. \tag{26}$$

We are able to solve the system explicitly (compare with Lemma 22 of [6]).

LEMMA. 15. *The generating function f_{T^2} is the following:*

$$\begin{aligned} f_{T^2}(z) &= -\frac{\sqrt{2}\sqrt{SU - 4z - f_{F^2} + 1} + S + f_{F^2} - 3}{2}, \\ \textit{where} \quad & S = \sqrt{1 - f_{F^2} + 2z} \\ & U = \sqrt{6z - f_{F^2} + 1}. \end{aligned} \tag{27}$$

Let us notice that the system of equations (21)-(26) is again *a-proper, a-positive, a-irreducible* and *a-aperiodic*. The common dominant singularity is $z_0 = 1/8$. The expansions of f_{F^2} and f_{T^2} around $z_0 = 1/8$ are the following:

$$\begin{aligned} f_{T^2}(z) &= \frac{1}{4}(5 - \sqrt{3} - \sqrt{2}\sqrt[4]{15}) \\ &\quad + 2\left(1 - \frac{1}{\sqrt{3}} - \frac{\sqrt{2} + \sqrt{\frac{3}{10}} + \sqrt{\frac{5}{6}}}{\sqrt[4]{15}}\right)\sqrt{1-8z} + O(1-8z), \\ f_{F^2}(z) &= \frac{1}{2} - 4\sqrt{1-8z} + O(1-8z). \end{aligned}$$

THEOREM. 16. *(see Kostrzycka in [6]) The asymptotic density of the set of classical tautologies $Cl_2^{\rightarrow}$ exists and is the following:*

$$\mu(Cl_2^{\rightarrow}) = -\frac{1}{2}\cdot\left(1 - \frac{1}{\sqrt{3}} - \frac{\sqrt{2} + \sqrt{\frac{3}{10}} + \sqrt{\frac{5}{6}}}{\sqrt[4]{15}}\right) \approx 0.5190\ldots \ . \tag{28}$$

The same approach we may apply to intuitionistic implicational logic of two variables:

THEOREM. 17. *(see Kostrzycka in [6]) The asymptotic density of the set of intuitionistically provable formulas $I_2^{\rightarrow}$ exists and is the following:*

$$\mu(I_2^{\rightarrow}) \approx 0.5043... \ . \tag{29}$$

The exact analytical formula for (29) is extremely complicated and too long to be written here. From (28) and (29) we have another characterization of the implicational fragments of intuitionistic and classical logics:

THEOREM. 18. *[Relative density] The relative density of intuitionistic tautologies among the classical ones in the language $\mathcal{F}_2^{\rightarrow}$ is more than 97%.*

$$\mu((I_2^{\rightarrow})/(Cl_2^{\rightarrow})) \approx 0.5043/0.5190 \approx 0.9715... \ . \tag{30}$$

4. Densities of implicational-negational tautologies

In the next two theorems we change the language. We consider formulas built by means of implication and negation from one variable. We can see that adding the functor of negation has negative impact on the density of tautologies. Moreover, as a result of the paper [9], we are able to find the exact density of the intuitionistic fragment of classical logic in this language. We can also see by Theorem 23 that within the reacher language with negation the density of purely implicational tautologies in the class of all tautologies is 0.

DEFINITION. 19. *The set* $\mathcal{F}_k^{\to,\neg}$ *over k propositional variables is the minimal set consisting of these variables and closed under implication and negation. In this definition the norm* $\|\phi\|$ *means the total number of characters (without parentheses) in the formula* ϕ.

THEOREM. 20. *(see Zaionc in [15]) The asymptotic density of the set of classical tautologies* $Cl_1^{\to,\neg}$ *exists and is:*

$$\begin{aligned} \mu(Cl_1^{\to,\neg}) &= \frac{1}{(4\sqrt{13})} + \frac{1}{(4\sqrt{17})} + \frac{1}{2\sqrt{2\left(\sqrt{221}-9\right)}} + \frac{15}{2\sqrt{442\left(\sqrt{221}-9\right)}} \\ &\approx 0.4232\ldots\ . \end{aligned}$$

THEOREM. 21. *(see Kostrzycka, Zaionc in [9]) The asymptotic density of the set of intuitionisticaly provable formulas* $I_1^{\to,\neg}$ *exists and is:*

$$\mu(I_1^{\to,\neg}) \approx 0.3952\ldots\ . \tag{31}$$

In the paper [9] the reader can find the analytical formula for (31). Putting together Theorems 20 and 21 we obtain:

THEOREM. 22. *[Relative density (see Kostrzycka, Zaionc in [9])] The relative density of intuitionistic tautologies among the classical ones in the language* $\mathcal{F}_1^{\to,\neg}$ *is more than 93 %.*

$$\mu((I_1^{\to,\neg})/(Cl_1^{\to,\neg})) \approx 0.93 \tag{32}$$

In the paper [9] the analytical formula for (32) is presented.

Let us also compare the sets of classical tautologies in different languages: $\mathcal{F}_1^{\to,\neg}$ and $\mathcal{F}_1^{\to}$.

THEOREM. 23. *The probability of finding classical implicational tautology among classical implicational-negational ones is* 0 *(in the sense of the norm in Definition 19).*

5. Densities of modal tautologies

To distinguish modal tautologies from the non-modal ones we need to consider a language with some modal operator. We have chosen the operator of necessity $\Box$. We will consider set $\mathcal{F}_1^{\to,\Box}$ of formulas built up from one propositional variable p by means of necessitation and implication only.

Our research so far has focussed on two modal logic: some normal extensions of Grzegorczyk logic and the Lewis logic **S5**. In the choice we were governed by the simplicity of the appropriate modal algebras.

The Grzegorczyk logic **Grz** is characterized as an extension of **S4** by the axiom

$$\Box(\Box(p \to \Box p) \to p) \to p.$$

In [7] we examined normal extensions of Grzegorczyk's logic obtained by adding to the set of axioms the formulas J_n defined as follows:

Definition. 24.

$$\begin{aligned} J_1 &= \Diamond\Box p_1 \to p_1, \\ J_{n+1} &= \Diamond(\Box p_{n+1} \wedge \sim J_n) \to p_{n+1}. \end{aligned}$$

We will consider the logics $\mathbf{Grz}^{\leq n} = \mathbf{Grz} \oplus J_n$. They contain the logic **Grz** and the following inclusions hold:

$$\mathbf{S4} \subset \mathbf{Grz} \subset ... \subset \mathbf{Grz}^{\leq n} \subset \mathbf{Grz}^{\leq n-1} \subset ... \subset \mathbf{Grz}^{\leq 1}. \tag{33}$$

The Lewis modal logic **S5** is characterized as an extension of **S4** by the axiom

$$(sym)\;\; p \to \Box\Diamond p.$$

and the following inclusion holds:

$$\mathbf{S4} \subset \mathbf{S5}. \tag{34}$$

In [7] and [8] we characterized the implicational-necessitional modal algebras of $\mathbf{Grz}^{\leq n}$ and **S5**. In the case of $\mathbf{Grz}^{\leq 2}$ we obtained four-element modal algebra, whereas in the cases of $\mathbf{Grz}^{\leq 3}$ and **S5** we had two different eight-element modal ones. In the first case we were able to determine the generating function for the class of tautologies explicitly and count the density as follows:

THEOREM. 25. *(see Kostrzycka in [7]) The asymptotic density of the set of Grzegorczyk's tautologies* $\mathbf{Grz}_1^{\leq 2}$ *exists and is:*

$$\mu(\mathbf{Grz}_1^{\leq 2}) \quad \approx \quad 0.6127... \ .$$

In the cases of eight-element modal algebras counting the densities is connected with solving a system of eight non-linear functional equations. We were able to do it only numerically and obtained the following results:

THEOREM. 26. *(see Kostrzycka in [8]) The densities of truth of the logics* $\mathbf{Grz}_1^{\leq 3}$ *and* $\mathbf{S5}_1$ *exist and:*

$$\begin{aligned} \mu(\mathbf{Grz}_1^{\leq 3}) &\approx 0.6088... \ , \\ \mu(\mathbf{S5}_1) &\approx 0.6081... \ . \end{aligned}$$

6. Probability distribution, typical formulas, typical tautologies

In this section we will discuss some questions concerning probability distribution (see Definition 3) of formulas written in the implicational language $\mathcal{F}_k^{\rightarrow}$ (see Definition 6) equipped with the norm $\|.\|$ measuring the total number of appearances of propositional variables in a formula.

DEFINITION. 27. *By* $\mathcal{F}_k^{\rightarrow}(p)$ *we mean the set of formulas having p premises, i.e. formulas which are of the form:* $\tau = \tau_1 \rightarrow (\cdots \rightarrow (\tau_p \rightarrow \alpha))$*, where* α *is a propositional variable.*

DEFINITION. 28. *A* simple tautology *is the formula* $\tau \in \mathcal{F}_k^{\rightarrow}$ *in the form*

$$\tau = \tau_1 \rightarrow (\cdots \rightarrow (\tau_p \rightarrow \alpha)),$$

such that there is at least one component τ_i *identical with* α*. Let* $\mathcal{G}_k$ *be the set of all simple tautologies in* $\mathcal{F}_k^{\rightarrow}$ *and* $\mathcal{G}_k(p)$ *be the set of simple tautologies with p premises .*

Evidently, a simple tautology is a tautology. Our goal is to find how big asymptotically is the fragment of simple tautologies within the set of all formulas and also how big is the fragment of simple tautologies with p premises in the set of all simple tautologies.

DEFINITION. 29. *Let us defined the random variable* $X : \mathcal{F}_k^{\rightarrow} \mapsto \mathbb{N}$ *(see Definition 2) which assigns to implicational formula the number of its premisses.*

In Theorem 31 we will check the correctness of the above definition since for any n the density $\mu(\{\phi : X(\phi) = n\})$ exists and moreover

$$\sum_{n=0}^{\infty} \mu(\{\phi : X(\phi) = n\}) = 1.$$

We wish to answer two questions:
QUESTION 1: *What is the probability that a randomly chosen implicational formula admits p premises?*
QUESTION 2: *What is the probability that a randomly chosen implicational simple tautology admits p premises?*

LEMMA. 30. *(see [11]) The asymptotic density of the set of all formulas with p premisses $\mathcal{F}_k^{\rightarrow}(p)$ exists and is:*

$$\mu(\mathcal{F}_k^{\rightarrow}(p)) = \frac{p}{2^{p+1}} \tag{35}$$

THEOREM. 31. *The random variable X has the following distribution (see Definition 3):*

$$\overline{X}(p) = \frac{p}{2^{p+1}}$$

The expected value is $E(X) = 3$, the variance is $D^2(X) = 4$. The standard deviation of X is 2.

From the whole discussion we can surprisingly see that a typical implicational formula has exactly 3 premisses. For example, the amount of formulas with the number of premises laying between 1 and 5 ie. which are typical $\pm$ standard deviation is 57/64 which is about 89%.

Now, we will answer the second question. We will show the difference between distribution of any formulas with p premisses and the distribution of simple tautologies only.

DEFINITION. 32. *For every $k \geq 1$ separately, let us define the random variable Y_k which assigns to an implicational simple tautology in the language $\mathcal{F}_k^{\rightarrow}$ the number of its premises.*

THEOREM. 33. *(Zaionc [16]) The random variable Y_k has the following distribution:*

$$\overline{Y_k}(p) = \frac{(2k+1)^2}{4k+1}\left(\frac{p}{2^{p+1}} - p\frac{(2k-1)^{p-1}}{4^p k^{p-1}}\right)$$

The natural questions are: what does the distribution of true sentences for very large numbers k look like and does there exist a uniform asymptotic distribution when k, the number of propositional variables in the logic, tends to infinity? The answers are the following:

LEMMA. 34. *(Zaionc [16]) In this lemma the number of premises $p \geq 0$ is fixed.*

$$\lim_{k\to\infty} \frac{(2k+1)^2}{4k+1}\left(\frac{p}{2^{p+1}} - p\frac{(2k-1)^{p-1}}{4^p k^{p-1}}\right) = \frac{p(p-1)}{2^{p+2}}$$

Let us name the limit distribution by $Y_\infty(p) = \frac{p(p-1)}{2^{p+2}}$. Since:

$$\sum_{p=0}^{\infty} \frac{p(p-1)}{2^{p+2}} = 1 \tag{36}$$

then the expected value of Y_∞ is:

$$E(Y_\infty) = \sum_{p=0}^{\infty} p\frac{p(p-1)}{2^{p+2}} = 5$$

The variance of Y_∞ is:

$$D^2(Y_\infty) = \sum_{p=0}^{\infty} p^2\frac{p(p-1)}{2^{p+2}} - 25 = 31 - 25 = 6$$

Comparing this result with the distribution $\overline{X}(p)$, the reader can easily check that starting with $k = 1$ the expected value of the number of premises for simple tautologies is substantially greater than 3 and is growing asymptotically to 5 and

$$\lim_{k\to\infty} E(Y_k) = 5. \tag{37}$$

The asymptotical behavior of $D^2(Y_k)$ is

$$\lim_{k\to\infty} D^2(Y_k) = 6. \tag{38}$$

So, it is clear now that:

$$\forall p \geq 0 \lim_{k\to\infty} \overline{Y_k}(p) = \overline{Y_\infty}(p) \tag{39}$$

$$\lim_{k\to\infty} E(Y_k) = E(Y_\infty) \tag{40}$$

$$\lim_{k\to\infty} D^2(Y_k) = D^2(Y_\infty) \tag{41}$$

The componentwise convergence presented in Lemma 34 and summarized by the formula (39) can be extended to a much stronger uniform convergence. In fact, the distribution $\overline{Y_\infty}$ can be treated as a good model of a distribution for simple tautologies of the language $\mathcal{F}$ when the number k of atomic propositional variables is large.

THEOREM. 35. *(Zaionc in [16]) The sequence of distributions $\overline{Y_k}$ converges uniformly to the distribution $\overline{Y_\infty}$.*

We can also see the following surprising result

THEOREM. 36. *(Zaionc [16]) For fixed $k > 0$ and $p > 0$*

$$\mu[(\mathcal{G}_k(p))/(\mathcal{F}_k^{\to}(p))] = 1 - \left(\frac{2k-1}{2k}\right)^{p-1}. \tag{42}$$

This result is somehow intriguing. It shows how big asymptotically is the fraction of simple tautologies with p premises among all formulas with p premises. We can see that with p growing, this fraction becomes closer and closer to 1. Of course the fraction of all tautologies (not only the simple ones) with p premises is even larger. So the *density of truth* within the classes of formulas of p premises can be as big as we wish. For every $\varepsilon > 0$, we can effectively find p such that almost all formulas with p premises (except a tiny fraction of the size ε) asymptotically are tautologies. This should be contrasted with the results proved in Theorem 6.3 and Corollary 6.10 page 587 in [11]. It shows that *density of truth* for all $p's$ together is always of the size $O(1/k)$. The result for every p treated separately is very different.

7. Counting types in typed λ calculus

In this section we are going to consider again the language $\mathcal{F}_1^{\to}$ of pure implication with one propositional variable O, introduced in the section 3. Under the Curry-Howard isomorphism we are going to look at this language from the point of view of simple types. We shall consider a simple typed lambda calculus with a single ground type O. Functions $rank$ and number of arguments arg for type τ are defined in the standard way by: $arg(O) = 0$, $rank(O) = 1$ and $arg(\tau_1 \to ... \to \tau_n \to O) = n$ and $rank(\tau_1 \to ... \to \tau_n \to O) = \max_{i=1...n} rank(\tau[i]) + 1$. A type τ is called regular if $rank(\tau) \leq 4$ and every component of τ has $arg \leq 1$. This implies that the only allowed components for regular types are O, $O \to O$ and $(O^k \to O) \to O$ for any $k \geq 1$. We will refer to the **lambda definability problem** defined independently by Statman and Plotkin (see Statman-Plotkin conjecture in [13] and [12]). A *full type hierarchy* $\{D_\tau\}_{\tau \in \mathcal{F}}$ is a collection of finite domains, one for each type. The whole hierarchy is determined by D_O. We assume that the set D_O is a given finite set and $D_{\tau \to \mu}$ is a collection of all functions from D_τ to D_μ. Therefore all D_τ are finite. It can be noted that any type τ term is interpreted as an object from D_τ. Hence, closed terms in the *full type hierarchy* have a fixed interpretation. A function $f \in D_\tau$ is called lambda definable if there is a closed type τ term T that is interpreted as f. For the given type τ, the *τ-lambda definability problem* is the decision problem to determine whether or not for the given finite D_O and the given object $f \in D_\tau$, f is lambda definable. In this case the type τ is not a part of the problem. It has been proved by Loader [10] that the lambda definability problem in general is undecidable which means that Statman-Plotkin conjecture fails (see [13] and [12]). There is a simple characterization of the types with decidable lambda definability problem.

THEOREM. 37. *λ definability problem is decidable for all rank 1, 2, 3 types and for regular rank 4 types.*

PROOF. First three cases are trivial. The decidability for rank 4 regular types is proved in [16]. ∎

THEOREM. 38. *λ definability problem is undecidable for any non regular rank 4 type.*

PROOF. The proof is based on the observation that the type $\mathbb{L} = ((O \to O) \to O) \to ((O \to (O \to O) \to O)$ is the simplest non regular type of rank 4. Therefore by a simple lambda definable coding, $\mathbb{L}$ can be embedded into any non regular type of rank 4. But then the λ definability problem is undecidable for $\mathbb{L}$ (see [5]) . ∎

As we can see from Theorems 37 and 38 there are rank 4 types with decidable and with undecidable lambda definability problem. So, the following natural problem arises.

QUESTION: What is the probability that a randomly chosen rank 4 type admits decidable lambda definability problem?

7.1. Counting types

In this section we present some properties of numbers characterizing the amount of types of different ranks.

DEFINITION. 39. *By F_k^n and G_k^n we mean respectively the total number of types of rank k and size n and the total number of types of rank $\leq k$ and size n, so:*

$$F_k^n = \#\{\phi \in \mathcal{F} : \|\phi\| = n \text{ and } rank(\phi) = k\}, \tag{43}$$
$$G_k^n = \#\{\phi \in \mathcal{F} : \|\phi\| = n \text{ and } rank(\phi) \leq k\}. \tag{44}$$

LEMMA. 40. *F_k^n and G_k^n are given by the following mutual recursion:*

$$F_1^n = \text{if } n = 1 \text{ then } 1 \text{ else } 0 \tag{45}$$
$$G_1^n = \text{if } n = 1 \text{ then } 1 \text{ else } 0 \tag{46}$$
$$F_{k+1}^n = \sum_{i=0}^{n} F_k^i G_k^{n-i} + \sum_{i=0}^{n} G_k^i F_{k+1}^{n-i} \tag{47}$$
$$G_{k+1}^n = G_k^n + F_{k+1}^n \tag{48}$$

PROOF. Obvious application of the definition of *rank*. ∎

7.2. Generating functions for counting types

We are going to use the generating functions technique for proving the asymptotic behavior of the appropriate fractions. For this purpose, let us define for every $k \geq 1$ the pair of the generating functions: $f_k(z) = \sum_{i=0}^{\infty} F_k^i z^i$ and $g_k(z) = \sum_{i=0}^{\infty} G_k^i z^i$.

LEMMA. 41. *The functions f_k and g_k satisfy the following recursive definitions:*

$$f_{k+1}(x) = \frac{f_k(x) g_k(x)}{1 - g_k(x)} \tag{49}$$
$$g_{k+1}(x) = g_k(x) + f_{k+1}(x) \tag{50}$$

PROOF. By a simple encoding of the recursive equations (45), (46), (47) and (48). ■

THEOREM. 42. *Let $\mathcal{R} \subset \mathbb{T}$ be the set of all regular types of rank 4. Let R_n be the number of elements of $\mathcal{R}$ of size n. The generating function f_R for the sequence R_n is*

$$f_R(x) = \frac{x^3}{(1-2x)(1-x-x^2)}$$

PROOF. Long but elementary proof may be found in [16]. ■

THEOREM. 43. *The density of rank 4 types with decidable λ definability problem among all rank 4 types is 0.*

PROOF. It is enough to find the closed form for the involved generating functions. Namely, for all rank 4 types the function calculated from equality (49) is

$$f_4(x) = \frac{x^4}{1-5x+7x^2-2x^3}$$

which can be easily turned into the power series $O\left(\left(\frac{3+\sqrt{5}}{2}\right)^n\right)$. The closed form term for the function $f_R(x)$ enumerating regular 4 order types returns rate of Fibonacci numbers; namely: $O\left(\left(\frac{1+\sqrt{5}}{2}\right)^n\right)$. ■

It doesn't help much to add all the third order types, for which we know that the lambda definability problem is decidable. Undecidability is again a dominating factor.

THEOREM. 44. *The density of types of rank ≤ 4 with the decidable λ definability problem among all types of rank ≤ 4 is again 0.*

PROOF. It is enough to find the closed form for the involved generating functions. For all rank ≤ 4 types the function is

$$g_4(x) = \frac{x-2x^2}{1-3x+x^2}$$

The respective power series is again $O\left(\left(\frac{3+\sqrt{5}}{2}\right)^n\right)$. The function g_3 obtained from the equality (50) is $\frac{x^3}{(1-2x)(1-x)}$. The closed form term for the function

$f_R(x) + g_3(x)$ enumerating regular 4 order types plus all third order types returns rate of $O(2^n)$. ■

References

[1] Comtet, L. *Advanced combinatorics. The art of finite and infinite expansions.* Revised and enlarged edition. Reidel,Dordrecht 1974.

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

[3] Flajolet, P. and Odlyzko, A. M. Singularity analysis of generating functions , *SIAM J. on Discrete Math.,* 3(2), 1990, pp 216-240.

[4] Gardy D. *Random Boolean expressions*, Colloquium on Computational Logic and Applications, Chambery (France), June 2005. Proceedings in DMTCS, 2006, pp 1-26.

[5] Joly T., On the λ definability I, the Model Problem and Generalizations of the Matching Problem, *Fundamentae Informaticae* 65 (1-2) 2005 pp 135-151.

[6] Kostrzycka, Z., On the density of implicational parts of intuitionistic and classical logics, *Journal of Applied Non-Classical Logics,* Vol. 13, Number 3, 2003, pp 295-325.

[7] Kostrzycka Z., On the density of truth in Grzegorczyk's modal logic, *Bulletin of the Section of Logic*, Vol. 33, Number 2, June 2004, pp 107-120.

[8] Kostrzycka Z., On the density of truth in modal logics, *Colloquium on Mathematics and Computer Science,* Nancy (France), September 2006. Proceedings in DMTCS, 2006, pp 161-170.

[9] Kostrzycka Z., Zaionc M. Statistics of intuitionistic versus classical logics, *Studia Logica,* **76**, 2004, 307-328.

[10] Loader R., *The Undecidability of λ - Definability,* In Logic, Meaning and Computation: Essays in Memory of Alonzo Church, 331-342, C.A. Anderson and Zeleny editors, Kluwer Academic Publishers, 2001.

[11] Moczurad M., Tyszkiewicz J., Zaionc M. Statistical properties of simple types, *Mathematical Structures in Computer Science,* vol 10, 2000, pp 575-594.

[12] Plotkin G. *λ definability and logical relations*, Memorandum SAI-RM-4, School of Artificial Intelligence, University of Edinburgh, Ocober 1973

[13] Statman R. *Equality of functionals revisited*, in L.A. Harrington et al. (Eds.), Harvey Friedman's Research on the Foundations of Mathematics, North- Holland, Amsterdam, 1985, 331-338.

[14] Wilf, H.S. *generatingfunctionology.* Second edition. Academic Press, Boston 1994.

[15] Zaionc, M. On the asymptotic density of tautologies in logic of implication and negation, *Reports on Mathematical Logic,* vol. 39, 2004.

[16] Zaionc, M. Probabilistic approach to the lambda definability for fourth order types, *Electronic Notes in Theoretical Computer Science,* vol. 140, 2005, pp 41-54.

[17] Zaionc, M. Probability distribution for simple tautologies, *Theoretical Computer Science,* vol. 355, Issue 2, 11 April 2006, pp 243-260.

Zofia Kostrzycka
University of Technology
Luboszycka 3,
45-036 Opole, Poland,
z.kostrzycka@po.opole.pl

Marek Zaionc
Theoretical Computer Science,
Jagiellonian University,
Gronostajowa 3,
30-072 Kraków, Poland
zaionc@ii.uj.edu.pl