REPORTS ON MATHEMATICAL LOGIC
36 (2002), 71–93

Grzegorz SOZA

ASYMPTOTIC DENSITY AS A METHOD OF EXPRESSING QUANTITATIVE RELATIONS IN INTUITIONISTIC LOGIC

A b s t r a c t. Our efforts in this work are mainly directed towards the statistical properties of tautologies and non-tautologies in intuitionistic logic (which is equivalent to research in typed lambda calculus because of the Curry-Howard isomorphism, see [1]). This article is a part of my master's thesis, which I defended at the Computer Science Department of Jagiellonian University in 2000. The inspiration for the thesis were the scientific works of the supervisor of my thesis dr hab. Marek Zaionc. In his [2] and [3] he dealt with typed lambda calculus considered over a finite number of ground types. His aim was to study the properties of types according to their length, defined as the number of occurrences of ground type variables in a type.

The goal here is quite similar, though we start from a different definition of the length of a type. In this work the complexity measure function (the "length" of a type) is defined as the height of its constructing tree. As we show the statistical behaviour of the type depends vitaly on the definition of its length. In Section 2 we prove that the asymptotic probability (defined precisely there) that a random one-variable formula is valid in intuitionistic logic (with implication only) is exactly 1, while by the linear definition of the length of a type (as discussed in [2] and [3]) this probability is equal to $\frac{1}{2} + \frac{\sqrt{5}}{10}$.

In Section 3 we shall be concerned with formulas their corresponding types consist of more than one ground type. We define a subset

Received 3 January 2001

of tautologies (called simple tautologies). Then we show that for each number k of ground type variables, and for each number $n > 1$, the ratio of types corresponding to this class of length n to all types of the same length n expressed as a fraction is always positive and at the same time bounded by $\frac{1}{k}$. We show also that the similarly defined ratio of all types representing tautologies to all types expressed as a fraction is greater than $\frac{1}{k}$, which implies that we have noticeably more (in terms of asymptotic density which is defined in Section 1) tautologies than simple tautologies. However, it does not give us precise information about all tautologies (as was the case in Section 2). Later on in Section 3 we shall be occupied with a subset of non-tautologies whose asymptotic density is positive by the linear definition of the length of a type, and moreover, this density tends to 1 as k tends to infinity. We show that by our double exponential definition of the length of a type this density equals 0.

In the last chapter we state some conjectures, which seem to hold but are not contained in this work.

1. Preliminary remarks

Although the considerations here touch on only types in typed lambda calculus, a similar method can be used to discuss the statistical properties of other finite mathematical objects. The idea is to apply a complexity measure function f arising from a given poset (partially ordered set) $\mathcal{B}$ of finite objects (not necessarily restricted in any other way) to positive natural numbers (meaning that $b \prec b'$ is a "subject"). The natural interpretation of $f(b)$ is a complexity of the object $b \in \mathcal{B}$.

$$f : \mathcal{B} \to N \setminus \{\emptyset\}$$

Usually we demand that such a function satisfy some assumptions, such as

1. $\forall\, x\ |f^{-1}(x)| < \omega$
2. $\mu \prec \tau \Rightarrow f(\mu) < f(\tau)$, $\prec$ - a partial order relation on $\mathcal{B}$
3. $x < y \Rightarrow |f^{-1}(x)| < |f^{-1}(y)|$

which is only formalization of some very intuitive expectations.

Then we define an equivalence relation on $\mathcal{B}$

$$\forall\, b_i, b_j \in \mathcal{B} \quad [b_i] = [b_j] \Leftrightarrow f(b_i) = f(b_j)$$

Such a relation partitions set $\mathcal{B}$ into $\aleph_0$ distinct equivalence classes. We formulate then properties of the elements of this set according to their belonging to particular classes. Using this method we can state the statistical properties of infinite sets, while concerning actually only finite subsets (we assumed all equivalence classes to be finite). Our complexity measure function gives us the opportunity to estimate the asymptotic probability of choosing an element fulfilling concrete assumptions from the whole set. In other words, we can estimate the asymptotic density of a specified group of elements in the set of all elements.

Here we shall apply the described method to determine the asymptotic density of some concrete subsets of types in typed lambda calculus.

We begin by introducing a definition of length of a type.

Definition 1.0.1. The length of a type is defined as follows

$|a| = 1$

$|\tau \to \mu| = \max(|\tau|, |\mu|) + 1$

It can easily be seen that a length so defined is in fact the height of the constructing tree of a given type.

Definition 1.0.2. The number of elements in a set $\mathcal{A}$ that are of length n is denoted by $\mathcal{A}_n$.

Definition 1.0.3. The asymptotic density of a set $\mathcal{A}$ in a set $\mathcal{B}$ is

$$d_{\mathcal{B}}(\mathcal{A}) = \liminf_{n\to\infty} \frac{\mathcal{A}_n}{\mathcal{B}_n}$$

If the sequence $\frac{\mathcal{A}_n}{\mathcal{B}_n}$ has a limit, we say that $\mathcal{A}$ has a natural density in a set $\mathcal{B}$, $d_{\mathcal{B}}(\mathcal{A})$.

We will deal here only with the special case where $\mathcal{B}$ is the set of all formulas in intuitionistic logic with implication (over a fixed number k of ground propositional variables). Thus we will omit it while talking about density, and later on the asymptotic density of a set $\mathcal{A}$ will mean the asymptotic density of this set $\mathcal{A}$ in the set of all formulas (built from implications and the following from a context number k of ground propositional variables).

The following lemma will be of great importance. It tells us that we can investigate the partial sums of the sequences $\mathcal{A}_n$ and $\mathcal{B}_n(\sum_{i=1}^n \mathcal{A}_i$ and $\sum_{i=1}^n \mathcal{B}_i$ respectively) when concerned with the asymptotic density of a given set $\mathcal{A}$.

Lemma 1.0.1. *Let A_n and B_n be given sequences of real numbers. Let us denote*

$$a_n = \sum_{i=1}^n A_i$$

$$b_n = \sum_{i=1}^n B_i$$

Suppose that

1. $\lim_{n\to\infty} \frac{a_n}{b_n}$ *exists or* $\lim_{n\to\infty} \frac{A_n}{B_n}$ *exists*
2. $\lim_{n\to\infty} \frac{a_n}{a_{n-1}} = \infty$
3. $\lim_{n\to\infty} \frac{b_n}{b_{n-1}} = \infty$

Then

$$\lim_{n\to\infty} \frac{A_n}{B_n} = \lim_{n\to\infty} \frac{a_n}{b_n}$$

Proof. Let us assume that $\lim_{n\to\infty} \frac{a_n}{b_n}$ exists. We have

$$\begin{aligned} \lim_{n\to\infty} \frac{a_n}{b_n} &= \lim_{n\to\infty} \frac{\frac{1}{b_n}}{\frac{1}{a_n}} = \lim_{n\to\infty} \frac{\frac{b_n - b_{n-1}}{b_n}(a_n - a_{n-1})}{\frac{a_n - a_{n-1}}{a_n}(b_n - b_{n-1})} = \lim_{n\to\infty} \frac{a_n - a_{n-1}}{b_n - b_{n-1}} \\ &= \lim_{n\to\infty} \frac{A_n}{B_n} \end{aligned}$$

If we assumed only the existence of the limit $\lim_{n\to\infty} \frac{A_n}{B_n}$ we would follow our identities in the reverse order. However, as we see, our calculations show that the existence of one of the limits implies the existence of the other. ■

2. Counting one-variable types

In this chapter we shall prove that there exists a natural asymptotic density of one-variable tautologies (considered in intuitionistic logic with implication only) and that the density is equal to 1. We begin with a definition of the length of a type.

Lemma 2.0.2. *The number of different types of length n is given by*

$$F_1 = 1$$

$$F_2 = 1$$

$$F_n = 2x_{n-2}F_{n-1} + F_{n-1}^2 \ for \ n > 2$$

where

$$x_n = \sum_{i=1}^{n} F_i$$

Proof. This lemma is a special case of Lemma 3.1.1. ∎

Let us look at the values of F_n and x_n for a few first n's

n	F_n	x_n
1	1	1
2	1	2
3	3	5
4	21	26
5	651	677
6	457653	458330
7	210065930571	210066388901

Lemma 2.0.3.

$$\lim_{n\to\infty} \frac{F_{n+1}}{F_n} = \infty$$

Proof. Since

$$\frac{F_{n+1}}{F_n} = 2x_{n-1} + F_n \Rightarrow \frac{F_{n+1}}{F_n} \geq F_n + F_{n-1} \text{ (by Lemma 2.0.2)}$$

it follows

$$\lim_{n\to\infty} \frac{F_{n+1}}{F_n} = \infty$$

∎

Definition 2.0.4. By numbers N_n and T_n we mean the number of non-tautologies (which correspond to non-inhabited types) and tautologies (which correspond to inhabited types) in intuitionistic logic with one propositional variable and implication only, respectively of length n.

It is well-known that intuitionistic tautologies with one propositional variable and implication only are simply classic provable formulas ([2]). Hence the following lemma holds.

Lemma 2.0.4. *$\tau \to \mu$ is a non-tautology $\Leftrightarrow$ τ is a tautology and μ is not a tautology.*

Proof. A proof of this lemma can be found in [2], [3]. ■

We will need Lemma 2.0.4 to state the following recurrence identities

Lemma 2.0.5. *Numbers N_n and T_n are given by the formulas*

$$N_1 = 1 \ ; \ N_n = T_{n-1}\sum_{i=1}^{n-1} N_i + N_{n-1}\sum_{i=1}^{n-1} T_i - N_{n-1}T_{n-1} \ for \ n \geq 2$$

$$T_1 = 0 \ ; \ T_n = F_n - T_{n-1}\sum_{i=1}^{n-1} N_i - N_{n-1}\sum_{i=1}^{n-1} T_i + N_{n-1}T_{n-1} \ for \ n \geq 2$$

Proof. This is a consequence of Lemma 2.0.4. Every type τ of length n has the form $\tau_1 \to \tau_2$, where at least one of the terms τ_i has the length n-1 and neither of the lengths is greater than n-1. We have 3 disjoint cases (see Lemma 3.1.1)

1. $|\tau_1| = n - 1 \ and \ |\tau_2| < n - 1$ - we have $T_{n-1}\sum_{i=1}^{n-2} N_i$ such types
2. $|\tau_1| < n - 1 \ and \ |\tau_2| = n - 1$ - we have $\sum_{i=1}^{n-2} T_i N_{n-1}$ such types
3. $|\tau_1| = n - 1 \ and \ |\tau_2| = n - 1$ - we have $N_{n-1}T_{n-1}$ such types

Hence we have

$$\begin{aligned} N_n &= T_{n-1}\sum_{i=1}^{n-2} N_i + N_{n-1}\sum_{i=1}^{n-2} T_i + N_{n-1}T_{n-1} \\ &= T_{n-1}\sum_{i=1}^{n-1} N_i + N_{n-1}\sum_{i=1}^{n-1} T_i - N_{n-1}T_{n-1} \end{aligned}$$

■

Below we present the list of numbers of tautologies (T_n), non-tautologies (N_n) and all (F_n) formulas with the fractions of $\frac{T_n}{F_n}$ for the first few n's.

n	T_n	N_n	F_n	$\frac{T_n}{F_n}$
1	0	1	1	0.0000
2	1	0	1	1.0000
3	2	1	3	0.6666
4	16	5	21	0.7619
5	524	127	651	0.8049
6	385024	72629	457653	0.8413
7	182010991712	28054938859	210065930571	0.8664
8	.	.	.	0.8843
9	.	.	.	0.8977
10	.	.	.	0.9081
11	.	.	.	0.9166
12	.	.	.	0.9235
13	.	.	.	0.9294
14	.	.	.	0.9344
15	.	.	.	0.9387
16	.	.	.	0.9424
17	.	.	.	0.9458
18	.	.	.	0.9487
19	.	.	.	0.9513
20	.	.	.	0.9537

We will now investigate partial sums of the sequences T_n and N_n, which will allow us to find the natural asymptotic densities of tautologies and non-tautologies (by Lemma 1.0.1). In some of the following lemmas we will make use of a recurrence equation for x_n, which can be found in Chapter 3, Lemma 3.1.2, formulated for a more general case (for x_n^k).

Definition 2.0.5. By y_n we mean the number of tautologies of a length not greater than n. By z_n we mean the number of nontautologies of a length not greater than n.

$$y_n = \sum_{i=1}^{n} T_i$$

$$z_n = \sum_{i=1}^{n} N_i$$

Lemma 2.0.6. *Numbers* x_n, y_n *and* z_n *are given by the recurrence formulas*

$$x_1 = 1,\ x_n = x_{n-1}^2 + 1,\ n \geq 2$$
$$y_1 = 0,\ y_n = x_n - y_{n-1}x_{n-1} + y_{n-1}^2 - 1,\ n \geq 2$$
$$z_1 = 1,\ z_n = y_{n-1}z_{n-1} + 1,\ n \geq 2$$

Proof. The formula for x_n follows from Lemma 3.1.2.
For y_n it holds

$$y_n = x_n - z_n = x_n - y_{n-1}z_{n-1} - 1 = x_n - y_{n-1}x_{n-1} + y_{n-1}^2 - 1,\ n \geq 2$$

The recursion for z_n is a simple consequence of Lemma 2.0.4. ■

Lemma 2.0.7.

$$\lim_{n\to\infty} \frac{y_n}{x_n} = 1$$

Proof. Firstly, we show that this limit exists.

1. The sequence $\frac{y_n}{x_n}$ is bounded by 0 and 1

2. We will show that the sequence is increasing. We have

$$\begin{aligned} \frac{y_n}{x_n} &= \frac{x_n - y_{n-1}x_{n-1} + y_{n-1}^2 - 1}{x_n} \\ &= 1 + \frac{y_{n-1}}{x_{n-1}^2 + 1}(y_{n-1} - x_{n-1}) - \frac{1}{x_n} \\ &> 1 + \frac{y_{n-1}}{x_{n-1}}(\frac{y_{n-1}}{x_{n-1}} - 1) - \frac{1}{x_n} \end{aligned}$$

We need now only prove that

$$1 + \frac{y_{n-1}}{x_{n-1}}(\frac{y_{n-1}}{x_{n-1}} - 1) - \frac{1}{x_n} - \frac{y_{n-1}}{x_{n-1}} \geq 0,\ n \geq 2$$

For a fixed $n \geq 2$ let us define $\frac{y_{n-1}}{x_{n-1}} = a$. We have to consider the difference $1 + a(a-2) - \frac{1}{x_n} = (1-a)^2 - \frac{1}{x_n}$. But we can bound $\frac{1}{x_n}$:

$$\frac{1}{x_n} = \frac{1}{x_{n-1}^2 + 1} < \frac{1}{x_{n-1}^2} \leq \frac{z_{n-1}^2}{x_{n-1}^2} = (\frac{z_{n-1}}{x_{n-1}})^2 = (1-a)^2$$

This means that our inequality holds, so the sequence is increasing.

An increasing and bounded sequence converges, hence the limit exists. Let us note that

$$\lim_{n\to\infty} \frac{1}{x_n} = 0 \text{ since } \lim_{n\to\infty} x_n = \infty$$

$$\lim_{n\to\infty} \frac{y_{n-1}x_{n-1}}{x_n} = \lim_{n\to\infty} \frac{y_{n-1}x_{n-1}}{x_{n-1}^2+1} = \lim_{n\to\infty} \frac{y_{n-1}}{x_{n-1}} = \lim_{n\to\infty} \frac{y_n}{x_n}$$

(exists as already mentioned),

$$\lim_{n\to\infty} \frac{y_{n-1}^2}{x_n} = \lim_{n\to\infty} \frac{y_{n-1}^2}{x_{n-1}^2+1} = \lim_{n\to\infty} \frac{y_{n-1}^2}{x_{n-1}^2} = (\lim_{n\to\infty} \frac{y_n}{x_n})^2 \text{ (exists).}$$

We can conclude that

$$\lim_{n\to\infty} \frac{y_n}{x_n} = \lim_{n\to\infty} (1 - \frac{y_{n-1}x_{n-1}}{x_n} + \frac{y_{n-1}^2}{x_n} - \frac{1}{x_n}) = 1 - \lim_{n\to\infty} \frac{y_n}{x_n} + (\lim_{n\to\infty} \frac{y_n}{x_n})^2$$

Let us define

$$\lim_{n\to\infty} \frac{y_n}{x_n} = g$$

which satisfies the equation $g = 1 - g + g^2$, which only solution is g=1. Hence the lemma holds. ∎

Lemma 2.0.8.

$$\lim_{n\to\infty} \frac{x_n}{x_{n-1}} = \infty \textit{ and } \lim_{n\to\infty} \frac{y_n}{y_{n-1}} = \infty$$

Proof. This follows directly from the recurrence formulas for x_n and y_n. ∎

Theorem 1.

$$\lim_{n\to\infty} \frac{T_n}{F_n} = 1$$

Proof. This theorem is just a simple consequence of Lemma 2.0.7, Lemma 2.0.8 and Lemma 1.0.1. ∎

3. Types built from k type variables

3.1 All formulas

Let us introduce a definition similar to the definition of the number F_n from the first paragraph

Definition 3.1.1. By F_n^k we mean the number of formulas which corresponding types are of length n and built from k ground variables. By T_n^k and N_n^k we mean the number of tautologies and non-tautologies, respectively. This means that $T_n^k + N_n^k = F_n^k$.

Lemma 3.1.1. *The number F_n^k is given by the formula*

$$F_1^k = k$$

$$F_n^k = 2 * (\sum_{i=1}^{n-2} F_i^k) * F_{n-1}^k + (F_{n-1}^k)^2 \; for \; n \geq 2$$

Proof. For $n = 1$ the lemma holds - the only types of length 1 are ground variables and we have k such variables. Let us consider a composite type of the form $\tau \to \mu$. This type can have length n in 3 disjoint cases

1. $|\tau| = n-1$ and $|\mu| < n-1$ - we have $F_{n-1}^k \sum_{i=1}^{n-2} F_i^k$ such types
2. $|\tau| < n-1$ and $|\mu| = n-1$ - we have $\sum_{i=1}^{n-2} F_i^k F_{n-1}^k$ such types
3. $|\tau| = n-1$ and $|\mu| = n-1$ - we have $(F_{n-1}^k)^2$ such types

Hence the lemma holds. ■

Definition 3.1.2. The number x_n^k is the number of all binary trees of a height not exceeding n and whose leaves can be labeled with one of the k variables. In terms of F_n^k this means

$$x_n^k = \sum_{i=1}^{n} F_n^k$$

Lemma 3.1.2. *The number x_n^k is given by the formula*

$$x_{n+1}^k = {x_n^k}^2 + k, \; n \geq 1$$

Proof. x_{n+1}^k is the number of binary trees of a height not greater than n+1. Let us consider one such tree and its left and right subtree. We can choose the left or the right subtree in x_n^k ways, which gives us ${x_n^k}^2$ possibilities of constructing our chosen tree. The only remaining cases are trees consisting only of one node roots (both its subtrees are empty). Altogether, this gives us ${x_n^k}^2 + k$ trees of a height not exceeding n+1 (as one node can be labeled with one of the k variables). ■

3.2 Tautologies

In this section we define a special subset of tautologies called simple tautologies. Next we show that for each k a natural asymptotic density of this subset exists and is positive (greater than $\frac{1}{k+1}$), but at the same time it is upper-bounded by $\frac{1}{k}$.

Definition 3.2.1. By a simple tautology we mean a formula which corresponding type has a form $\tau = \tau_1, \ldots, \tau_n \rightarrow a$ for some ground type a, such that there is at least one component τ_i identical with an a. Evidently simple tautology is a tautology with the proof being a projection $\lambda x_1 \ldots x_n.x_i$. Let G_n^k be the number of types of length n built from k ground type variables corresponding to simple tautologies.

Lemma 3.2.1. *The number G_n^k is given by the formula*

$$G_1^k = 0\ ;\ G_2^k = k$$

$$G_n^k = (\sum_{i=2}^{n-2} G_i^k) * F_{n-1}^k + (\sum_{i=1}^{n-1} F_i^k) * G_{n-1}^k + (F_{n-1}^k - G_{n-1}^k)\ for\ n \geq 3$$

Proof. For n=1 and for n=2 the lemma is obvious. Let us consider a composite type of the form $\tau_1 \rightarrow \tau_2$. The proof is based on the following observation: $\tau_1 \rightarrow \tau_2$ is simple if τ_2 is simple. This simple tautology can have length n in 3 disjoint cases

1. $|\tau_1| = n - 1$ *and* $|\tau_2| < n - 1$ - we have $F_{n-1}^k \sum_{i=2}^{n-2} G_i^k$ such simple tautologies. The sum starts from i=2 because $G_1^k = 0$.

2. $|\tau_1| < n - 1$ *and* $|\tau_2| = n - 1$ - we have $\sum_{i=1}^{n-2} F_i^k G_{n-1}^k$ such simple tautologies

3. $|\tau_1| = n - 1$ *and* $|\tau_2| = n - 1$ - we have $F_{n-1}^k * G_{n-1}^k$ such simple tautologies

The only other possible simple tautologies are if τ_1 is the same ground type that τ_2 points to. Therefore for every type τ_2 of length n-1 which corresponding formula is not a simple tautology (there are exactly $F_{n-1}^k - G_{n-1}^k$ of such types) we have exactly one type representing a simple tautology $a \rightarrow \tau_2$ where a is the ground type which the type τ_2 proves. Hence the lemma holds. ■

Below you can find G_n^k for n=1,2,3,4 contrasted with the numbers F_n^k of all formulas with k variables.

$$\begin{array}{ll} G_1^k = 0 & F_1^k = k \\ G_2^k = k & F_2^k = k^2 \\ G_3^k = k^3 + 2k^2 - k & F_3^k = k^4 + 2k^3 \\ G_4^k = k^7 + 4k^6 + 5k^5 + 4k^4 & F_4^k = k^8 + 4k^7 + 6k^6 + 6k^5 + 4k^4 \\ \quad + 2k^3 - 3k^2 + k & \\ . & . \\ . & . \\ . & . \end{array}$$

Lemma 3.2.2.

$$\forall m \geq 1 \; \exists M \; \forall k \geq M \; \forall n \geq 2 \;\; \frac{m-1}{mk} F_n^k \leq G_n^k$$

$$\forall k \; \forall n \geq 2 \;\; G_n^k \leq \frac{1}{k} F_n^k$$

Proof. m is fixed, induction over n. For n=2 the lemma is obvious (see the table above). Suppose the inequality holds for i=2,3, ... ,n-1. We have

$$\begin{aligned} G_n^k &= (\sum_{i=2}^{n-2} G_i^k) * F_{n-1}^k + (\sum_{i=1}^{n-1} F_i^k) * G_{n-1}^k + (F_{n-1}^k - G_{n-1}^k) \\ &= (\sum_{i=2}^{n-2} G_i^k) * F_{n-1}^k + (\sum_{i=1}^{n-2} F_i^k) * G_{n-1}^k + F_{n-1}^k + (F_{n-1}^k - 1) G_{n-1}^k \\ &\geq \frac{m-1}{mk} (\sum_{i=2}^{n-2} F_i^k) * F_{n-1}^k + (\sum_{i=1}^{n-2} F_i^k) * \frac{m-1}{mk} F_{n-1}^k + F_{n-1}^k \\ &\quad + (F_{n-1}^k - 1) * \frac{m-1}{mk} F_{n-1}^k \\ &= \frac{m-1}{mk} (2 * (\sum_{i=1}^{n-2} F_i^k) * F_{n-1}^k + (F_{n-1}^k)^2) - \frac{m-1}{mk} * k * F_{n-1}^k \\ &\quad + F_{n-1}^k - \frac{m-1}{mk} F_{n-1}^k \\ &= \frac{m-1}{mk} F_n^k + F_{n-1}^k (\frac{1}{m} - \frac{m-1}{mk}) \geq \frac{m-1}{mk} F_n^k \end{aligned}$$

since $\frac{1}{m} - \frac{m-1}{mk} \geq 0$ for $k \geq m-1$. That means that we can assign $M = m-1$.

Upper bound

$$\begin{aligned} G_n^k &= (\sum_{i=2}^{n-2} G_i^k) * F_{n-1}^k + (\sum_{i=1}^{n-1} F_i^k) * G_{n-1}^k + (F_{n-1}^k - G_{n-1}^k) \\ &\leq \frac{1}{k}(\sum_{i=1}^{n-2} F_i^k) * F_{n-1}^k - \frac{1}{k} F_1^k F_{n-1}^k + \frac{1}{k}(\sum_{i=1}^{n-2} F_i^k) * F_{n-1}^k \\ &\quad + \frac{1}{k}(F_{n-1}^k)^2 + (F_{n-1}^k - G_{n-1}^k) \\ &= \frac{1}{k}(2 * (\sum_{i=1}^{n-2} F_i^k) * F_{n-1}^k + (F_{n-1}^k)^2) - G_{n-1}^k = \frac{1}{k} F_n^k - G_{n-1}^k \leq \frac{1}{k} F_n^k \end{aligned}$$

■

Definition 3.2.2. The number g_n^k is the number of all simple tautologies over k ground variables of length not exceeding n.

$$g_n^k = \sum_{i=1}^{n} G_i^k$$

Lemma 3.2.3.

$$g_1^k = 0$$

$$g_n^k = x_{n-1}^k g_{n-1}^k + x_{n-1}^k - g_{n-1}^k \; for\ n \geq 2 \; and\ all\ k$$

Proof. For $n = 1$ the lemma is obvious. Let us consider a composite type of the form $\tau_1 \to \tau_2$. The proof is based on the following observation: $\tau_1 \to \tau_2$ is simple if τ_2 is simple. We have $x_{n-1}^k g_{n-1}^k$ of such simple tautologies.

The only other possible simple tautologies are if τ_1 is the same ground type that τ_2 points to. Therefore for every type τ_2 of a length not exceeding n-1 which does not correspond to a simple tautology (there are exactly $x_{n-1}^k - g_{n-1}^k$ of such types) we have exactly one type representing a simple tautology $a \to \tau_2$ where a is the ground type which the type τ_2 proves. Hence the lemma holds. ■

Lemma 3.2.4.

$$\lim_{n\to\infty} \frac{g_n^k}{x_n^k} \; exists$$

Proof. 1. The sequence $\frac{g_n^k}{x_n^k}$ is bounded by 0 and 1

2. We will show that the sequence is increasing. Using Lemma 3.2.3 we obtain

$$\begin{aligned}\frac{g_n^k}{x_n^k} - \frac{g_{n-1}^k}{x_{n-1}^k} &= \frac{x_{n-1}^k g_{n-1}^k + x_{n-1}^k - g_{n-1}^k}{x_n^k} - \frac{g_{n-1}^k}{x_{n-1}^k} \\ &= \frac{{x_{n-1}^k}^2 g_{n-1}^k + {x_{n-1}^k}^2 - g_{n-1}^k x_{n-1}^k - ({x_{n-1}^k}^2 + k) g_{n-1}^k}{({x_{n-1}^k}^2 + k) x_{n-1}^k} \\ &= \frac{{x_{n-1}^k}^2 - g_{n-1}^k x_{n-1}^k - k g_{n-1}^k}{({x_{n-1}^k}^2 + k) x_{n-1}^k} \text{ for } \ n \geq 2\end{aligned}$$

Since $({x_{n-1}^k}^2 + k) x_{n-1}^k$ is always positive, we have

$$\frac{g_n^k}{x_n^k} - \frac{g_{n-1}^k}{x_{n-1}^k} > 0 \Leftrightarrow (x_{n-1}^k - g_{n-1}^k) x_{n-1}^k > k g_{n-1}^k$$

This last inequality holds since $x_1^k - g_1^k = k$, and not all types correspond to simple tautologies. Thus the sequence is increasing.

An increasing and bounded sequence converges, hence the lemma holds. ■

Lemma 3.2.5.

$$\lim_{n\to\infty} \frac{G_n^k}{F_n^k} = \lim_{n\to\infty} \frac{g_n^k}{x_n^k} \textit{ for all } k$$

Proof. The lemma follows from Lemma 1.0.1 (by Lemma 3.2.4 and because $\lim_{n\to\infty} \frac{g_n^k}{g_{n-1}^k} = \infty$ for all k). ■

Lemma 3.2.6.

$$\frac{1}{k+1} < \lim_{n\to\infty} \frac{G_n^k}{F_n^k} \leq \frac{1}{k}$$

Proof. The first inequality we obtain by Lemma 3.2.4 (as $\frac{g_2^k}{x_2^k} = \frac{k}{k^2+k}$ and the sequence $\frac{g_n^k}{x_n^k}$ is increasing).

The second inequality ($\lim_{n\to\infty} \frac{G_n^k}{F_n^k} \leq \frac{1}{k}$) follows from Lemma 3.2.2. ■

We will now show that not all tautologies are simple ones (in terms of asymptotic density). We will define a subclass of tautologies (called almost-simple tautologies), and will show that the ratio of the types corresponding to these formulas to all types expressed as a fraction is greater than the corresponding fraction of types representing simple tautologies.
All the properties that we stated for simple tautologies have their corresponding formulations for almost-simple tautologies. Moreover, the proofs are almost the same, so we do not include them separately here. We present only the very lemmas.

Definition 3.2.3. By an almost-simple tautology we mean a formula which corresponding type is a simple tautology or of the form

$$\tau = \tau_1, \dots, \tau_i, \mu \to \mu$$

for some (not necessarily ground) type μ. It is obvious that an almost-simple tautology is a tautology. Let $\bar{G}_n^k$ be the number of types representing almost-simple tautologies of length n built from k ground type variables.

Lemma 3.2.7. *The number $\bar{G}_n^k$ is given by the formula*

$$\bar{G}_1^k = 0 \ ; \ \bar{G}_2^k = k$$

$$\bar{G}_n^k = (\sum_{i=2}^{n-2} \bar{G}_i^k) * F_{n-1}^k + (\sum_{i=1}^{n-1} F_i^k) * \bar{G}_{n-1}^k + 2(F_{n-1}^k - \bar{G}_{n-1}^k) \ \textit{for } n \geq 3$$

Definition 3.2.4. The number $\bar{g}_n^k$ is the number of all almost-simple tautologies over k ground variables of a length not exceeding n.

$$\bar{g}_n^k = \sum_{i=1}^{n} \bar{G}_i^k$$

Lemma 3.2.8.

$$\bar{g}_1^k = 0$$

$$\bar{g}_n^k = x_{n-1}^k \bar{g}_{n-1}^k + 2x_{n-1}^k - 2\bar{g}_{n-1}^k - k \ \textit{for } n \geq 2 \textit{ and all } k$$

Lemma 3.2.9.

$$\lim_{n\to\infty} \frac{\bar{g}_n^k}{x_n^k} \ \textit{exists}$$

Lemma 3.2.10.

$$\lim_{n\to\infty}\frac{\bar{G}_n^k}{F_n^k}=\lim_{n\to\infty}\frac{\bar{g}_n^k}{x_n^k}$$

Theorem 2.

$$\frac{1}{k+1}<\lim_{n\to\infty}\frac{G_n^k}{F_n^k}\le\frac{1}{k}<\lim_{n\to\infty}\frac{\bar{G}_n^k}{F_n^k}\quad \textit{for all } k\ge 3$$

Proof. The first two inequalities follow from Lemma 3.2.6.
To prove that $\frac{1}{k}<\lim_{n\to\infty}\frac{\bar{G}_n^k}{F_n^k}$ we will consider the sequence $\frac{\bar{g}_n^k}{x_n^k}$ which has the same limit as the sequence $\frac{\bar{G}_n^k}{F_n^k}$. This sequence is increasing (this was proved in Lemma 3.2.9). Hence we have

$$\lim_{n\to\infty}\frac{\bar{g}_n^k}{x_n^k}\ge\frac{\bar{g}_3^k}{x_3^k}=\frac{k^3+3k^2-k}{k^4+2k^3+k^2+k}=\frac{1}{k}+\frac{k^2-2k-1}{k^4+2k^3+k^2+k}$$

Since $k^2-2k-1>0$ for $k\ge 3$, the inequality holds. ■

3.3 Non-tautologies

In this section we investigate a subgroup of non-tautologies (called simple nontautologies), which was considerably large by the linear definition of the length of a type (see [3]). We show, however, that though a natural asymptotic density of this subset exists, it is equal to 0.
In this section we differentiate types according to the number of premises, which is useful while discussing properties of simple non-tautologies.

Definition 3.3.1. By $F_n^k(\overline{p})$ we mean the number of types of length n built from k ground type variables and having p premises - types which are of the form: $\tau_1\to\ldots\to\tau_p\to a$ where a is a ground type. Since the numbers $F_n^k(\overline{p})$ describe disjoint sets of types for different p's and since there are no types of length n having more than n-1 premises for $n\ge 2$ we have

$$F_n^k=F_n^k(\overline{0})+\ldots+F_n^k(\overline{n-1})$$

Lemma 3.3.1. *The number $F_n^k(\overline{p})$ is given by recurrence (on p)*

$$F_1^k(\overline{0})=k,\ F_n^k(\overline{0})=0\ \textit{for } n\ge 2$$

$$F_1^k(\overline{1})=0,\ F_n^k(\overline{1})=kF_{n-1}^k\ \textit{for } n\ge 2$$

$$F_n^k(\overline{p}) = (\sum_{i=1}^{n-1} F_i^k) * F_{n-1}^k(\overline{p-1}) + (\sum_{i=p}^{n-2} F_i^k(\overline{p-1})) * F_{n-1}^k \; for \; p \geq 2$$

Proof. The only types having 0 premises are simply ground type variables. These variables have obviously length 1.

$F_n^k(\overline{1}) = kF_{n-1}^k$ is the number of types of the form $\tau \to a$ since for each type of length $n-1$ we can choose k ground types a.

The formula for $p \geq 2$ is correct because any type $\tau \to \mu$ has to have p premises if μ has p-1 premises. This type can have length n in 3 disjoint cases

1. $|\tau| = n-1$ and $|\mu| < n-1$ - we have $F_{n-1}^k \sum_{i=p}^{n-2} F_i^k(\overline{p-1})$ such types. The sum starts with i=p because $F_{p-1}^k(\overline{p-1}) = \ldots = F_1^k(\overline{p-1}) = 0$.
2. $|\tau| < n-1$ and $|\mu| = n-1$ - we have $\sum_{i=1}^{n-2} F_i^k F_{n-1}^k(\overline{p-1})$ such types
3. $|\tau| = n-1$ and $|\mu| = n-1$ - we have $F_{n-1}^k F_{n-1}^k(\overline{p-1})$ such types

Hence the lemma holds. ■

Definition 3.3.2. By "a-type" for some ground type a we mean any type of the form $\tau_1, \ldots, \tau_p \to a$. A type is also "$a$-type" for $p = 0$.

By a simple non-tautology we mean a formula which corresponding type is an "a-type" τ such that all its components are not "a-types" for any ground type a.

The number of simple non-tautologies of length n built from k variables we will denote by C_n^k.

The number of simple non-tautologies of length n built from k variables with exactly p premises will be denoted by $C_n^k(\overline{p})$.

It is worth noticing that a simple non-tautology is neither an intuitionistic nor a classic tautology (we can cause the term to have value zero by assigning 1 to all ground types different from a, and 0 to a).

Lemma 3.3.2. *The number $C_n^k(\overline{p})$ is given by*

$$\left(\frac{k-1}{k}\right)^p F_n^k(\overline{p})$$

Proof. There are $\frac{F_n^k(\overline{p})}{k^{p+1}}$ patterns of types with p premises of the form:

$$(\ldots \to \bigcirc), (\ldots \to \bigcirc), \ldots, (\ldots \to \bigcirc) \to \bigcirc$$

in which $\bigcirc$ are places in the type left for some ground types to fill. There are exactly $p+1$ such places. There are $k(k-1)^p$ ways of filling a pattern in such a way as to obtain a simple non-tautology. Altogether there are

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

of simple non-tautologies with p premises. ■

Lemma 3.3.3. *The total number C_n^k of types corresponding to simple non-tautologies of length n built from k ground type variables is exactly*

$$\left(\frac{k-1}{k}\right)^0 F_n^k(\overline{0}) + \left(\frac{k-1}{k}\right)^1 F_n^k(\overline{1}) + \ldots + \left(\frac{k-1}{k}\right)^{n-1} F_n^k(\overline{n-1})$$

Proof. The formula for all simple non-tautologies can be obtained just by adding all simple non-tautologies for every possible p, since all classes of simple non-tautologies described by the numbers $C_n^k(\overline{p})$ are disjoint. ■

Lemma 3.3.4. *The sequence $\frac{F_n^k(\overline{n-1})}{x_n^k}$ is decreasing for $n \geq 2$ and for every k and the seqeunce $\frac{F_n^k(\overline{n-p})}{x_n^k}$ is increasing for every $p \geq 2$, $n \geq p+1$ and every k.*

Proof. Using Lemma 3.3.1 we have

$$\frac{F_n^k(\overline{n-p})}{x_n^k} - \frac{F_{n-1}^k(\overline{n-p-1})}{x_{n-1}^k} = \frac{1}{({x_{n-1}^k}^2 + k)x_{n-1}^k} \times$$

$$\Big((x_{n-1}^k F_{n-1}^k(\overline{n-p-1}) + (\sum_{i=n-p}^{n-2} F_i^k(\overline{n-p-1}))F_{n-1}^k)x_{n-1}^k$$

$$-F_{n-1}^k(\overline{n-p-1})({x_{n-1}^k}^2 + k)\Big)$$

Since $({x_{n-1}^k}^2 + k)x_{n-1}^k$ is positive, we have

$$\frac{F_n^k(\overline{n-p})}{x_n^k} - \frac{F_{n-1}^k(\overline{n-p-1})}{x_{n-1}^k} > 0 \Leftrightarrow$$

$$x_{n-1}^{k}{}^{2} F_{n-1}^{k}(\overline{n-p-1}) + (\sum_{i=n-p}^{n-2} F_i^k(\overline{n-p-1}))F_{n-1}^k x_{n-1}^k$$

$$-x_{n-1}^{k}{}^{2} F_{n-1}^{k}(\overline{n-p-1}) - kF_{n-1}^k(\overline{n-p-1}) > 0 \Leftrightarrow$$

$$x_{n-1}^k F_{n-1}^k(\sum_{i=n-p}^{n-2} F_i^k(\overline{n-p-1})) > kF_{n-1}^k(\overline{n-p-1})$$

For p=1 the sum occurring in the inequality vanishes. For every $p \geq 2$ and for every $n \geq p+1$ and every k the sum is a positive natural number. Since $x_{n-1}^k > k$ for all positive integers k and every natural $n \geq 2$, and $F_{n-1}^k > F_{n-1}^k(\overline{n-p-1})$, it implies that the sequence is increasing. ■

Lemma 3.3.5.

$$\lim_{n\to\infty} \frac{F_n^k(\overline{n-p})}{x_n^k} \textit{ exists for all } k,\ p = 1, \dots, n-1$$

Proof. The sequence whose limit is considered is bounded. On the other hand, it is monotonous, as we proved in Lemma 3.3.4. More precisely for $p = 1$ and $n \geq 2$ this sequence is decreasing, for other p's less or equal $n-1$ it is increasing, and for $p = n$ $(n \geq 2)$ it is equal 0. Hence the limit exists. ■

Now we shall provide a proof which is somewhat stronger than needed here, from which we shall deduce Lemma 3.3.7.

Lemma 3.3.6. $\forall k \geq 2\ \forall p \geq 2\ \forall n \geq \max(5, p+1)$

$$\frac{k-1}{kx_{n-3}^k} F_n^k(\overline{p}) > F_n^k(\overline{p-1}) + F_{n-1}^k$$

Proof. Let us assume that n is fixed. Induction over p.

Let us consider the case where $p = 2$. We have

$$\begin{aligned}
\frac{k-1}{kx_{n-3}^k}F_n^k(\overline{2}) &= \frac{k-1}{kx_{n-3}^k}((\sum_{i=1}^{n-1}F_i^k)F_{n-1}^k(\overline{1}) + (\sum_{i=2}^{n-2}F_i^k(\overline{1}))F_{n-1}^k) \\
&= \frac{k-1}{kx_{n-3}^k}(x_{n-1}^k kF_{n-2}^k + (\sum_{i=2}^{n-2}kF_{i-1}^k)F_{n-1}^k) \\
&= \frac{k-1}{x_{n-3}^k}(x_{n-1}^k F_{n-2}^k + x_{n-3}^k F_{n-1}^k) \\
&\geq \frac{1}{x_{n-3}^k}(x_{n-1}^k F_{n-2}^k + x_{n-3}^k F_{n-1}^k) \\
&> \frac{1}{x_{n-3}^k}(F_{n-1}^k kx_{n-3}^k + x_{n-3}^k F_{n-1}^k) = F_n^k(1) + F_{n-1}^k
\end{aligned}$$

The last inequality holds since $x_{n-1}^k > F_{n-1}^k$ and it can easily be proved that $F_{n-2}^k > kx_{n-3}^k$ for $n \geq 5,\ k \geq 1$.

Let us suppose the lemma holds for every $2 \leq i \leq p$. We have

$$\begin{aligned}
\frac{k-1}{kx_{n-3}^k}F_n^k(\overline{p+1}) &= \frac{k-1}{kx_{n-3}^k}((\sum_{i=1}^{n-1}F_i^k)F_{n-1}^k(\overline{p}) + (\sum_{i=p+1}^{n-2}F_i^k(\overline{p}))F_{n-1}^k) \\
&= (\sum_{i=1}^{n-1}F_i^k)\frac{k-1}{kx_{n-3}^k}F_{n-1}^k(\overline{p}) + (\sum_{i=p+1}^{n-2}\frac{k-1}{kx_{n-3}^k}F_i^k(\overline{p}))F_{n-1}^k \\
&> (\sum_{i=1}^{n-1}F_i^k)(F_{n-1}^k(\overline{p-1}) + F_{n-2}^k) \\
&\quad +(\sum_{i=p+1}^{n-2}(F_i^k(\overline{p-1}) + F_{i-1}^k))F_{n-1}^k \\
&= F_n^k(\overline{p}) - F_p^k(\overline{p-1})F_{n-1}^k + (\sum_{i=1}^{n-1}F_i^k)F_{n-2}^k \\
&\quad +(\sum_{i=p+1}^{n-2}F_{i-1}^k)F_{n-1}^k \\
&\geq F_n^k(\overline{p}) - F_p^k(\overline{p-1})F_{n-1}^k + F_{n-1}^k(F_{n-2}^k - 1 + 1) \\
&= F_n^k(\overline{p}) - F_p^k(\overline{p-1})F_{n-1}^k + F_{n-1}^k + (F_{n-2}^k - 1)F_{n-1}^k \\
&\geq F_n^k(\overline{p}) + F_{n-1}^k
\end{aligned}$$

The last inequality holds since $F^k_{n-2} - 1 > F^k_p(\overline{p-1})$ for $p \leq n-2$ and $n \geq 5$. Hence the lemma holds. ■

As an immediate derivation of this lemma we obtain

Lemma 3.3.7. $\forall k \geq 2\ \forall p \geq 2\ \forall n \geq \max(5, p+1)$

$$\frac{k-1}{2k}\, F^k_n(\overline{p}) > F^k_n(\overline{p-1})$$

Now we will show the following lemma

Lemma 3.3.8.

$$\forall\ k \lim_{n\to\infty} \frac{C^k_n(\overline{n-p})}{x^k_n} = 0$$

Proof. We know by Lemma 3.3.5 that $\lim_{n\to\infty} \frac{F^k_n(\overline{n-p})}{x^k_n}$ exists ($p = 1, \ldots, n$). This limit is finite since $\frac{F^k_n(\overline{n-p})}{x^k_n}$ is bounded. Since

$$\lim_{n\to\infty} (\frac{k-1}{k})^{n-p} = 0,$$

this gives

$$\lim n \to \infty \frac{(\frac{k-1}{k})^{n-p} F^k_n(\overline{n-p})}{x^k_n} = \lim_{n\to\infty} \frac{C^k_n(\overline{n-p})}{x^k_n} = 0$$

■

Lemma 3.3.9.

$$\forall\ k \lim_{n\to\infty} \frac{C^k_n}{x^k_n} = 0$$

Proof. By Lemma 3.3.8 (taking $p = 1$) we have

$$\forall\, \epsilon\ \exists\, N \geq 4\ \forall\, n > N\ \frac{(\frac{k-1}{k})^{n-1} F^k_n(\overline{n-1})}{x^k_n} \leq \epsilon$$

But by Lemma 3.3.7 we have $\forall\, n > N,\ p = 1, \ldots, n-2$

$$\begin{aligned} \frac{(\frac{k-1}{k})^{n-p-1} F^k_n(\overline{n-p-1})}{x^k_n} &< \frac{(\frac{k-1}{k})^{n-p-1} \frac{k-1}{2k} F^k_n(\overline{n-p})}{x^k_n} \\ &= \frac{1}{2} \frac{(\frac{k-1}{k})^{n-p} F^k_n(\overline{n-p})}{x^k_n} \leq \frac{1}{2}\epsilon \end{aligned}$$

Thus

$$\forall\, n > N\ \frac{(\frac{k-1}{k})^{n-p} F_n^k(\overline{n-p})}{x_n^k} \leq \frac{\epsilon}{2^{p-1}} \text{ for } p = 1, \dots, n-1$$

Hence

$$\forall\, n > N\ \frac{(\frac{k-1}{k})^0 F_n^k(\overline{0}) + (\frac{k-1}{k})^1 F_n^k(\overline{1}) + \ldots + (\frac{k-1}{k})^{n-1} F_n^k(\overline{n-1})}{x_n^k}$$

$$\leq 0 + \frac{\epsilon}{2^{n-2}} + \frac{\epsilon}{2^{n-3}} + \ldots + \epsilon < 2\epsilon$$

which means that

$$\forall\, k\ \lim_{n\to\infty} \frac{(\frac{k-1}{k})^0 F_n^k(\overline{0}) + (\frac{k-1}{k})^1 F_n^k(\overline{1}) + \ldots + (\frac{k-1}{k})^{n-1} F_n^k(\overline{n-1})}{x_n^k}$$

$$= \lim_{n\to\infty} \frac{C_n^k}{x_n^k} = 0$$

■

Theorem 3.

$$\forall\, k\ \lim_{n\to\infty} \frac{C_n^k}{F_n^k} = 0$$

Proof. This is a simple consequence of Lemma 2.0.8 and Lemma 3.3.9. ■

This result should be contrasted with the analogous one in [2] and [3], where (by the linear definition of the length of a type) the correspondent fraction for every k tends to a positive number as k tends to infinity. Moreover, it tends exactly to 1, which intuitively means that the types corresponding to simple non-tautologies are statistically almost the only ones occurring in typed lambda calculus.

Here we can only state that this special subclass of non-tautologies is in terms of asymptotic density of no importance. However, we do not have enough information to conclude the statistical behaviour of all non-tautologies by our complexity measure of a type.

4. Conjectures

In Chapter 2 we showed that a natural asymptotic density of one-variable tautologies exists and is equal to 1. In Chapter 3 we investigated types built from more ground type variables, in particular we dealt with simple tautologies and simple non-tautologies. Our results prove that the asymptotic density of tautologies (as they contain simple tautologies) is positive for every k. However, it does not tell us what value the asymptotic density exactly has. We conjecture that the asymptotic density decreases as k increases. It is not known what we would get if we evaluated the limit of these densities over k tending to infinity. We suppose that, as in [3], we would obtain 0.

These conjectures expressed in terms of our notation are

1.

$$\liminf_{n\to\infty} \frac{T_n^k}{F_n^k} > \liminf_{n\to\infty} \frac{T_n^{k+1}}{F_n^{k+1}}, \; for\ k \geq 1$$

2.

$$\lim_{k\to\infty} \liminf_{n\to\infty} \frac{T_n^k}{F_n^k} = 0$$

References

[1] J. R. Hindley, Basic Simple Type Theory, Cambridge University Press, 1997

[2] M. Zaionc, J. Tyszkiewicz and M. Moczurad, *Statistical Properties of Simple Types*, Mathematical Structures in Computer Science 10 (5), 2000, pp. 575 – 594

[3] M. Zaionc, *How Many Types Are Inhabited*, 1998, unpublished

Institute of Computer Science
Jagiellonian University
Cracow, Poland

soza@informatik.uni-erlangen.de
grzegorz_soza@yahoo.com