Statistical Properties of Simple Types

Małgorzata Moczurad*
Jerzy Tyszkiewicz**
Marek Zaionc*

* Computer Science Department,
Jagiellonian University,
Nawojki 11, 30-072 Kraków, Poland
madry@softlab.ii.uj.edu.pl
zaionc@ii.uj.edu.pl

** School CSE, UNSW, 2052 Sydney, Australia
on leave from
Institute of Informatics,
University of Warsaw,
Banacha 2, 02-097 Warszawa, Poland
jty@mimuw.edu.pl

Supported by the ARC grant A 49800112 (1998–2000)

School of Computer Science and Engineering
The University of New South Wales
Sydney 2052, Australia

Abstract

We consider types and typed lambda calculus over a finite number of ground types. We are going to investigate the size of the fraction of inhabited types of the given length n against the number of all types of length n. The plan of this paper is to find the limit of that fraction when $n \to \infty$. The answer to this question is equivalent to finding the "density" of inhabited types in the set of all types, or the so-called asymptotic probability of finding an inhabited type in the set of all types. Under the Curry-Howard isomorphism this means finding the density or asymptotic probability of provable intuitionistic propositional formulas in the set of all formulas. For types with one ground type (formulas with one propositional variable) we prove that the limit exists and is equal to $1/2 + \sqrt{5}/10$, which is approximately 72%. This means that the random type (formula) of the large size is as likely as about 72% to be inhabited (tautology). We also prove that for every finite number k of ground-type variables, the density of inhabited types is always positive and lies between $(4k+1)/(2k+1)^2$ and $(3k+1)/(k+1)^2$. Therefore we can easily see that the density is decreasing to 0 with k going to infinity. From the lower and upper bounds presented we can deduce that at least 1/3 of classical tautologies are intuitionistic.

1 Introduction

In this paper we examine the density of inhabited types among all first order types of a given length. A type is inhabited, if there is at least one closed term of λ-calculus of this type. The inhabitation problem for a simple typed λ-calculus, which became the real motivation for this paper, was deeply studied by Hindley and is fully analyzed in [3]. By the Curry-Howard isomorphism, inhabitation is equivalent to the statement that the type is an intuitionistic tautology (in the language of pure implication). Thus what we indeed attempt to measure is the density of intuitionistic tautologies among all purely implicative formulas. In this statement the λ-calculus is not even mentioned. What we however borrow from it is our interest in formulas with very few propositional variables, which corresponds to the normal situation with just few basic types. Indeed, λ-calculus is a (or perhaps even *the*) mathematical model for functional programming, and in programming languages the number of atomic types is usually pretty small. In the logical setting such restriction would be perhaps questionable, but given our interest in λ-calculus, it is quite natural and legitimate.

Our first main result is a precise estimate of the number of intuitionistic tautologies among formulas of length n with one propositional variable. This result was partially motivated by a short Statman's notice at the end of his paper [4] in which he estimated the fraction of inhabited types as lying between 0.625 and 0.86, and ended his notice with the nice sentence: *"It is a good bet but not a sure thing, that ρ (type) contains a closed term."* As we prove, this is a really *good bet* because the fraction of tautologies approaches the limit $1/2 + \sqrt{5}/10$ as n tends to infinity.

For types with $k > 1$ atomic types we do not know the exact fraction of tautologies, but we prove two bounds which almost match. After isolating special formulas, which we call simple tautologies and simple nontautologies, we prove that their fractions among all formulas of length n approach $(4k+1)/(2k+1)^2$ and $k(k-1)/(k+1)^2$, respectively. So, for large k, the upper bound on the fraction of tautologies is about 3 times the lower bound. We conjecture that indeed the fraction of tautologies is for large k very close to the lower bound. In particular, we conjecture that most of the tautologies are indeed simple (in our terminology).

Since our simple tautologies are even classical tautologies, and simple nontautologies are even classically invalid, our conjecture implies that for large k almost all classical tautologies are intuitionistic, and in fact simple. At the moment we know only that about a third part of classical tautologies are intuitionistic tautologies.

Finally, it is worth observing that, in the light of our results, the number of tautologies with k variables eventually decreases with k: if k grows more than by a factor of 3, the (asymptotic) fraction of tautologies must drop below the old value. We prove separately that asymptotically there are indeed less tautologies with 2 atomic types than with 1 atomic type.

2 Prerequisites

2.1 Simple typed λ-calculus

We shall consider a simple typed λ-calculus with finite ground types. Types are defined as follows: every ground type is a type and if τ and μ are types then $\tau \to \mu$ is a type. We will use the following notation: if τ_1, τ_2, ..., τ_n, μ are types then by $\tau_1, \tau_2, ..., \tau_n \to \mu$ we mean the type $\tau_1 \to (\tau_2 \to (... \to (\tau_n \to \mu)...))$. Therefore, every type τ has the form $\tau_1, \tau_2, ..., \tau_n \to \alpha$ where α is a ground type. If τ has the form $\tau_1, ..., \tau_n \to \alpha$ then τ_i for $i \leq n$ are called components or premises of the type τ and are denoted by $\tau[i]$. This notation can be iterated so by a type $\tau[i_1, ..., i_{k-1}, i_k]$ we mean the type $(\tau[i_1, ..., i_{k-1}])[i_k]$. By the length of a type τ we mean the number of occurrences of ground types in τ.

For any type τ there is given a denumerable set of variables $V(\tau)$. Any type τ variable is a type τ term. If T is a term of type $\tau \to \mu$ and S is a type τ term, then TS is a term which has type μ. If T is a type μ term and x is a type τ variable, then $\lambda x.T$ is a term of type $\tau \to \mu$. The axioms of equality between terms have the form of $\beta\eta$ conversions and the convertible terms will be written as $T =_{\beta\eta} S$. Term T is in a long normal form if $T = \lambda x_1...x_n.yT_1...T_k$, where y is an x_i for some $i \leq n$ or y is a free variable, T_j for $j \leq k$ are in a long normal form and $yT_1...T_k$ is a ground type term. Long normal forms exist and are unique for $\beta\eta$ conversions (see [4]). A closed term is a term without free variables. A type is called inhabited if there is a closed term of this type (for very detailed analysis of inhabited types see for example [3]). Let us introduce a complexity measure π for closed terms. If T is a closed term written in a long normal form and $T = \lambda x_1...x_n.x_i$ then $\pi(T) = 1$. If $T = \lambda x_1...x_n.x_iT_1...T_k$, then $\pi(T) = \max_{j=1...k}(\pi(\lambda x_1...x_n.T_j)) + 1$. Complexity π is just the depth of the Böhm tree of a given term. For more detailed treatment of typed λ-calculus see [3], [6] or more general [1]. Since we are going to use Curry-Howard isomorphism, we will not distinguish between types and formulas.

To make this paper self-contained we prove by the lambda calculus technique very well known fact about the inclusion of the intuitionistic formulas (inhabited types) in the set of formulas which are classically provable.

Lemma 2.1. *If there is a closed term T of type τ (τ is an intuitionistic tautology) then τ (as a formula) is a classical tautology.*

Proof. First we observe that any type (formula) $\tau[1], ..., \tau[n] \to \tau[i]$ is a classical tautology whether or not $\tau[i]$ is a ground type. The proof of the lemma is by induction on the complexity $\pi(T)$. Let the type τ have a form $\tau[1], ..., \tau[n] \to \alpha$, where α is a ground type (since π measures the complexity of terms this is in fact induction on the complexity of proofs). For $\pi(T) = 1$ it is trivial, since the type τ must have the form $\tau[1], ..., \alpha, ..., \tau[n] \to \alpha$ and this is of course a classically valid formula. Suppose the term T is in a long normal form $\lambda x_1...x_n.x_i K_1...K_p$. All p terms $T_1 = (\lambda x_1...x_n.K_1)$,..., $T_p = (\lambda x_1...x_n.K_p)$ are simpler according to the complexity π. Therefore, because of induction argument their types $(\tau[1], ..., \tau[n] \to \tau[i,1])$,..., $(\tau[1], ..., \tau[n] \to \tau[i,p])$ are classically valid formulas. Notice that $\tau[1], ..., \tau[n] \to \tau[i]$, which is a type of the term $\lambda x_1...x_n.x_i$, is also classically valid. Notice that $\tau[i] = \tau[i,1], ..., \tau[i,p] \to \alpha$. Let us consider 0-1 valuation of ground types under which all formulas $\tau[1], ..., \tau[n]$ are true. From the inductive assumption formulas $\tau[i,1], ..., \tau[i,p]$ are true under this valuation. Therefore α is true. □

2.2 Generating functions

The main tool we use for dealing with asymptotics of sequences of fractions are *generating functions.* A nice exposition of the method can be found in [7, 2].

Our main task in this paper is to determine limits of various sequences of real numbers. For this purpose combinatorics has developed an extremely powerful tool, in the form of generating series and generating functions.

Let $A = A_0, A_1, A_2, \ldots$ be a sequence of real numbers (if it is necessary, we tacitly extend A to the left, by assuming $A_n = 0$ for negative n). The *ordinary generating series* for A is the formal power series $\sum_{n=0}^{\infty} A_n z^n$. And, of course, formal power series are in one-to-one correspondence to sequences. However, considering z as a complex variable, this series, as known from the theory of analytic functions, converges uniformly to a function $f_A(z)$ in some open disc $\{z \in \mathbb{C} : |z| < R\}$ of maximal diameter, and $R \geq 0$ is called its radius of convergence. So with the sequence A we can associate a complex function $f_A(z)$, called the *ordinary generating function* for A, defined in a neighbourhood of 0. This correspondence is one-to-one again (unless $R = 0$), since, as it is well known from the theory of analytic functions, the expansion of a complex function $f(z)$, analytic in a neighbourhood of z_0, into a power series $\sum_{n=0}^{\infty} A_n (z - z_0)^n$ is unique, and, moreover, this series is the Taylor series, given by

$$A_n = \frac{1}{n!} \frac{d^n f}{dz^n}(z_0). \tag{1}$$

Many questions concerning the asymptotic behaviour of A can be efficiently resolved by analysing the behaviour of f_A at the complex circle $|z| = R$.

This is the approach we take to determine the asymptotic fraction of intuitionistic tautologies among all types of a given length.

The key tool will be the following result, in which the symbol $[z^n]\{F\}$ stands for the coefficient of z^n in the exponential series expansion of F. It is due to Szegö [5] (our version is simplified, the full version works for any finite number of singularities on the circle of convergence).

Theorem 2.2 (Szegö Lemma). *Let $v(z)$ be analytic in $|z| < 1$ with a single singularity $e^{i\varphi}$ at the circle $|z| = 1$. Suppose that in the neighbourhood of $e^{i\varphi}$, $v(z)$ has the expansion*

$$v(z) = \sum_{p \geq 0} v_p (1 - z e^{i\varphi})^{a+pb}, \tag{2}$$

where $a \in \mathbb{C}$ and $b > 0$ is real, and the branch chosen above for the expansion equals $v(0)$ for $z = 0$. Then

$$[z^n]\{v(z)\} = \sum_{p=0}^{\xi(q)} v_p \binom{a+pb}{n} (-e^{i\varphi})^n + O(n^{-q}), \tag{3}$$

where $\xi(q) = \lceil b^{-1}(q - \Re(a) - 1) \rceil$.

In (3) $\binom{a}{n}$ stands for $a(a-1)\ldots(a-(n-1))/n!$.

As the reader may now expect, while working with propositional intuitionistic logic and simply typed λ-calculus, most of the time we will be concerned with complex analysis, analytic functions and their singularities.

3 Counting types

In this section we present some properties of the Catalan numbers characterising the number of possible bracketings of a type and consequently, the number of types of a given length.

Definition 3.1. *The total number of bracketings of a type of length n is equal to the Catalan number C_n given by the recursion:*

$$C_0 = 0 \qquad C_1 = 1 \qquad C_n = \sum_{i=1}^{n-1} C_i C_{n-i}. \tag{4}$$

Let us show the generating function for the C_n numbers. Denote $f_C(t) = \sum_{n=0}^{\infty} C_n t^n$.

The recursion (4) becomes, after a closer examination, the equality

$$f_C(t) \cdot f_C(t) = f_C(t) - t$$

since the recursion exactly corresponds to the multiplication of power series, with the exception that the linear term t disappears. By solving the equation we get two possible solutions:

$$f_C(t) = \frac{1}{2} - \frac{1}{2}\sqrt{1-4t}$$

or

$$f_C(t) = \frac{1}{2} + \frac{1}{2}\sqrt{1-4t}.$$

We have to choose the first solution, since it corresponds to the assumption $C_0 = 0$ (for the calculation see the proof of Theorem 4.6).

¿From the generating function f_C we can extract, by (1), the well-known nonrecursive formula for C_n (see for example [7]).

$$C_n = \frac{1}{n}\binom{2n-2}{n-1} \tag{5}$$

Below we quote properties which are a consequence of this formula.

For every $n \geq 1$ and for every $k \geq 1$

$$\frac{C_n}{C_{n+1}} = \frac{1}{4} + \frac{3}{8n-4}, \tag{6}$$

$$\frac{C_n}{C_{n+k}} > \frac{1}{4^k}, \tag{7}$$

$$\lim_{n\to\infty} \frac{C_n}{C_{n+k}} = \frac{1}{4^k}. \tag{8}$$

Definition 3.2. *By $C_n(p)$ we mean the number of bracketings of a type of length n having p premises, i.e. formulas which are of the form: $\tau = \tau_1 \to (... \to (\tau_p \to \alpha))$, where α is a ground type. Since numbers $C_n(p)$ are the cardinalities of disjoint sets of formulas for different p's and since there are no types of length n having more then $n-1$ premises, for $n \geq 2$ we have:*

$$C_n = C_n(1) + ... + C_n(n-1).$$

Lemma 3.3. *Number $C_n(p)$ is given by the recursion (on p):*

$$\begin{aligned} C_n(1) &= C_{n-1} \\ C_n(p) &= \sum_{i=1}^{n-p} C_i C_{n-i}(p-1) \end{aligned}$$

Proof. It is obvious that $C_n(1) = C_{n-1}$, since $C_n(1)$ is the number of types of the form $\tau \to \alpha$. Consider

$$\tau = \tau_1 \to \underbrace{(\tau_2 \to (...(\tau_p \to \alpha)...))}_{\mu},$$

where τ_1 is of length i. The number of possible bracketings of τ is the number of bracketings of τ_1 (i.e. C_i) and μ (i.e. $C_{n-i}(p-1)$), summed over all possible divisions at position i. The summation stops at $i = n - p$, since beginning with $i = n - p + 1$ the terms become zero. □

Lemma 3.4. *For fixed p the generating function for $C_n(p)$ is*

$$f_p(x) = x(f_C(x))^p = x\left(\frac{1-\sqrt{1-4x}}{2}\right)^p$$

Proof. The recursion $C_n(p) = \sum_{i=1}^{n-p} C_i C_{n-i}(p-1)$ becomes, after a closer examination, the equality $f_p(x) \cdot f_C(x) = f_{p+1}(x)$ and since $C_n(1) = C_{n-1}$ obviously $f_1(x) = x(f_C(x))$. □

Lemma 3.5. *For all $p > 0$ holds $\lim_{n\to\infty} C_n(p)/C_n = p/2^{p+1}$.*

Proof. See Section 7. □

The following corollary is an easy consequence of Lemma 3.5.

Corollary 3.6. *The random variable X which assigns to a type the number of its premises has the distribution*

$$X(p) = \begin{cases} 0 & \textit{if } p = 0 \\ \frac{p}{2^{p+1}} & \textit{if } p > 0 \end{cases}$$

and has the expected value

$$E(X) = \sum_{p=1}^{\infty} p\frac{p}{2^{p+1}} = 3$$

and has the variance

$$D^2(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2 = \sum_{p=1}^{\infty} p^2 \frac{p}{2^{p+1}} - 9 = 4.$$

so the standard deviation of X is 2.

The already presented properties concern the number of bracketings of a type. It is clear that the number of types of length n over a k-letter alphabet can be obtained by multiplying the number of bracketings by the number of possible letter patterns. In other words, the number of types of length n over an alphabet with k atomic types is

$$F_n^k = k^n C_n$$

and the number of types with p premises of length $n \geq 2$ over an alphabet with k atomic types is

$$F_n^k(p) = k^n C_n(p).$$

Since there are no tautologies of length $n = 1$ we consider only formulas at least two letters long. It can be easily seen that $F_n^k = F_n^k(1) + ... + F_n^k(n-1)$, since there are no formulas of length n with more than $n-1$ premises. From Lemma 3.3 we can see that

$$F_n^k(1) = kF_{n-1}^k \tag{9}$$

$$F_n^k(p) = \sum_{i=1}^{n-p} F_i^k F_{n-i}^k(p-1). \tag{10}$$

Example 3.7.

$F_n^k(p)$	$\lim_{n\to\infty} \frac{F_n^k(p)}{F_n^k} = \frac{p}{2^{p+1}}$
$F_n^k(1) = kF_{n-1}^k$	25.00%
$F_n^k(2) = kF_{n-1}^k$	25.00%
$F_n^k(3) = kF_{n-1}^k - k^2F_{n-2}^k$	18.75%
$F_n^k(4) = kF_{n-1}^k - 2k^2F_{n-2}^k$	12.50%
$F_n^k(5) = kF_{n-1}^k - 3k^2F_{n-2}^k + k^3F_{n-3}^k$	7.81%
...	...

From the table above we can see that statistically 89.06% of types have 5 or less premises and 50% of types have 1 or 2 premises.

4 Types with one ground type

We start the discussion about the density of inhabited types from the types built from just one atomic type α. As noticed by Statman in [4] an implicational formula over one variable is an intuitionistic tautology iff it is a classical tautology. The simple proof of that fact which allows us to count inhabited types and then find an asymptotic probability is given here in Lemma 4.3. In [4], page 529, Statman examines the probability of inhabitation, proving that it must lie between 0.625 and 0.86. We are going to prove that indeed the asymptotic probability exists and is exactly $1/2 + \sqrt{5}/10$.

Lemma 4.1. *τ is a tautology iff $\tau \to \alpha$ is not a tautology.*

Proof. ($\to$) Since τ is a tautology, τ is a classical tautology (see Lemma 2.1). So $\tau \to \alpha$ is not a classical tautology because of valuations which assign 0 to α. Again from Lemma 2.1, it is not a tautology.

($\leftarrow$) By induction on the length of τ. Suppose τ is not a tautology. Let $\tau = \tau[1], ..., \tau[n] \to \alpha$. First we show that all $\tau[i]$ for $1 \le i \le n$ are tautologies. If some $\tau[i]$ is not a tautology then by induction $\tau[i] \to \alpha$ is a tautology so $\tau = \tau[1], ..., \tau[n] \to \alpha$ is also a tautology; contradiction. So let $T_1 : \tau[1] \dots T_n : \tau[n]$ be closed terms. Therefore $\lambda x.xT_1...T_n$ is a closed term of the type $(\tau[1], ..., \tau[n] \to \alpha) \to \alpha$ which ends the proof. □

Lemma 4.2. *$\tau[1], ..., \tau[n] \to \alpha$ is not a tautology iff all $\tau[i]$ are tautologies*

Proof. ($\to$) By the same classical logic argument as in Lemma 4.1. ($\leftarrow$) Assume some $\tau[i]$ is not a tautology. So $\tau[i] \to \alpha$ is a tautology and therefore $\tau[1], ..., \tau[n] \to \alpha$ must be also a tautology. □

Lemma 4.3. *$\tau \to \mu$ is not a tautology iff τ is a tautology and μ is not a tautology.*

Proof. ($\to$) Let $\mu = \mu[1] \dots \mu[n] \to \alpha$. So the type $\tau \to \mu$ is of the form $\tau, \mu[1] \dots \mu[n] \to \alpha$. By Theorem 4.2 all $\mu[i]$ and τ are tautologies. Therefore $\mu[1] \dots \mu[n] \to \alpha$ is not a tautology.
($\leftarrow$) Simply follows from Theorem 4.2. □

Definition 4.4. *By N_n^1 and T_n^1 we mean the number of noninhabited types (nontautologies) and inhabited types (tautologies), respectively, of length n built with one letter.*

Lemma 4.5. *Numbers N_n^1 of noninhabited types (nontautologies) of length n and T_n^1 of inhabited types (tautologies) of length n are given by formulas:*

$$\begin{array}{rclcrcl} N_1^1 & = & 1 & \quad & N_n^1 & = & \sum_{i=1}^{n-1} T_i^1 N_{n-i}^1 \\ T_1^1 & = & 0 & \quad & T_n^1 & = & F_n^1 - \sum_{i=1}^{n-1} T_i^1 N_{n-i}^1 \end{array}$$

Proof. This is a simple consequence of Lemma 4.3. Every type τ of length n has the form $\tau_1 \to \tau_2$, where length of τ_1 is i and length of τ_2 is $n-i$. Since τ is a nontautology if and only if τ_1 is a tautology and τ_2 is a nontautology. Therefore we have the total number of nontautologies of length n as $\sum_{i=1}^{n-1} T_i^1 N_{n-i}^1$. □

The main theorem for this section is that the sequence T_n^1/F_n^1 is convergent. The meaning of this is the following: the asymptotic probability of the fact that a given type is inhabited (or that a given implicational formula is a tautology) exists and is equal to $1/2 + \sqrt{5}/10$.

Theorem 4.6. *The limit of the sequence T_n^1/F_n^1 exists and is equal to $1/2 + \sqrt{5}/10$*

Proof. See Section 7. □

In the sequel we are going to consider the class of simple tautologies which are an important fragment of the set of tautologies. Namely, by a simple tautology (over an arbitrary alphabet) we mean a type $\tau = \tau_1, ..., \tau_n \to \alpha$ such that there is at least one component τ_i identical with α. Evidently, a simple tautology is a tautology with a proof being a projection $\lambda x_1 ... x_n . x_i$. Let G_n^1 be a number of simple tautologies of length n built with one ground-type variable. In the next theorem we will find how big asymptotically is the fragment of simple one-letter tautologies within the set of all one-letter tautologies.

Theorem 4.7. *Asymptotically* $\frac{25-5\sqrt{5}}{18} \approx 76.77\%$ *of all one letter tautologies are simple and* $\frac{5\sqrt{5}-7}{18} \approx 23.22\%$ *are not simple.*

Proof. The sequence G_n^1/T_n^1 is convergent and the limit is $\frac{25-5\sqrt{5}}{18}$ simply because

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

See limits in Theorems 4.6 and 6.3 for $k = 1$. □

5 Counting two-letter types

In this section we are going to find lower and upper bounds for asymptotic probabilities for the fraction of inhabited types of the types built with two letters. We will see that the upper bound for the fraction of tautologies (inhabited types), i.e. the fraction T_n^2/F_n^2, is smaller then the limit of T_n^1/F_n^1. This proves that asymptotically there are fewer tautologies (inhabited types) for two ground types than for one ground type. At the same time the fraction of two-letter tautologies is asymptotically positive.

By T_n^2 and N_n^2 we mean the numbers of tautologies and nontautologies, respectively, of length n over a two-letter alphabet. Of course $T_n^2 + N_n^2 = F_n^2$.

Definition 5.1. *By the projection of a two-letter type* τ *we mean the one-letter type* $\bar{\tau}$ *obtained from* τ *by substituting all occurrences of ground-type letters in* τ *by one letter. We will employ the same notation for terms; so by* $\bar{T}$ *we mean the same term* T *with all typing information in* T *changed from two-letter types* τ *into* $\bar{\tau}$*.*

Lemma 5.2. *Projection of a tautology is a tautology.*

Proof. If there is a closed typed lambda term T of type τ then the term $\bar{T}$ is of type $\bar{\tau}$. □

Definition 5.3. *By* ${}^{\otimes}N_n^2$ *we mean the number of such two-letter nontautologies whose projections are one-letter tautologies, and contrary* ${}^{\downarrow}N_n^2$ *means the number of such two-letters nontautologies whose projections are again one-letter nontautologies. Because classes of types described by numbers* ${}^{\downarrow}N_n^2$ *and* ${}^{\otimes}N_n^2$ *are disjoint, then simply* ${}^{\downarrow}N_n^2 + {}^{\otimes}N_n^2 = N_n^2$*.*

In the next theorem we will find the class of formulas which are nontautologies but their projections are tautologies. Moreover, the asymptotic density of the class is big enough to show that the total amount of nontautologies for two-letter types is bigger then for one-letter ones.

Lemma 5.4. *For every two-letter tautology* τ *each formula in one of the forms*

$$\begin{array}{cc} \tau, \alpha \to \beta & \tau, \alpha \to \alpha \\ \alpha, \tau \to \beta & \beta, \tau \to \alpha \end{array}$$

is a two-letter nontautology while its projection is a tautology. Therefore ${}^{\otimes}N_n^2 \geq 4T_{n-2}^2$ *for* $n \geq 4$.

Proof. It is easy to notice that all those formulas are nontautologies since even classically they are not valid (see Lemma 2.1). Each projection must be a tautology since letters α and β are getting unified. The inequality holds since all four categories are disjoint for $n \geq 4$. □

Lemma 5.5. *For every* $\varepsilon > 0$ *there is a* n_0 *such that for every* $n > n_0$ ${}^{\downarrow}N_n^2 \geq (\frac{1}{2} - \frac{\sqrt{5}}{10} - \varepsilon)F_n^2$

Proof. It follows from Lemma 5.2 since every formula whose projection is a nontautology must be a nontautology. Therefore ${}^{\downarrow}N_n^2 = 2^n N_n^1$. Since N_n^1/F_n^1 is converging to the limit $\frac{1}{2} - \frac{\sqrt{5}}{10}$ which is complementary to $\frac{1}{2} + \frac{\sqrt{5}}{10}$ (see theorem 4.6), then for positive ε there is n_0 such that for every $n > n_0$ $N_n^1 \geq (\frac{1}{2} - \frac{\sqrt{5}}{10} - \varepsilon)F_n^1$. So ${}^{\downarrow}N_n^2 = 2^n N_n^1 \geq (\frac{1}{2} - \frac{\sqrt{5}}{10} - \varepsilon)2^n C_n = (\frac{1}{2} - \frac{\sqrt{5}}{10} - \varepsilon)F_n^2$. □

Lemma 5.6. ${}^{\otimes}N_n^2 \geq \frac{1}{48}F_n^2$ *for all sufficiently large* n.

Proof. The inequality follows from the definition of the class described by the number ${}^{\otimes}N_n^2$ and Lemma 5.4 ${}^{\otimes}N_n^2 \geq 4T_{n-2}^2$. Therefore

$$\frac{{}^{\otimes}N_n^2}{F_n^2} \geq \frac{4T_{n-2}^2}{F_n^2} = \frac{F_{n-2}^2}{F_n^2}\frac{4T_{n-2}^2}{F_{n-2}^2} = \frac{2^{n-2}C_{n-2}}{2^n C_n}\frac{4T_{n-2}^2}{F_{n-2}^2} = \frac{1}{4}\frac{C_{n-2}}{C_n}\frac{4T_{n-2}^2}{F_{n-2}^2} \geq \frac{1}{48}$$

since $\frac{C_{n-2}}{C_n} > \frac{1}{16}$ for all n (see (7)) and also $\frac{T_{n-2}^2}{F_{n-2}^2} \geq \frac{G_{n-1}^2}{F_{n-1}^2} \to \frac{9}{25} > \frac{1}{3}$ (see Theorem 6.3 with $k = 2$). □

Theorem 5.7. *There are fewer two-letters tautologies than single letter tautologies but the fraction of two-letter tautologies is still positive, i.e.*

$$\frac{1}{3} \leq \liminf_{n\to\infty} \frac{T_n^2}{F_n^2} \leq \limsup_{n\to\infty} \frac{T_n^2}{F_n^2} < \lim_{n\to\infty} \frac{T_n^1}{F_n^1} = \frac{1}{2} + \frac{\sqrt{5}}{10}$$

Proof. $N_n^2 = {}^{\downarrow}N_n^2 + {}^{\otimes}N_n^2$. Let us take $\varepsilon = \frac{1}{96}$. From Lemma 5.5 there exists n_0 such that for $n > n_0$ ${}^{\downarrow}N_n^2 \geq (\frac{1}{2} - \frac{\sqrt{5}}{10} - \varepsilon)F_n^2$. In the same time from Lemma 5.6 starting from $n = 4$, ${}^{\otimes}N_n^2 \geq \frac{1}{48}F_n^2 = 2\varepsilon F_n^2$. Therefore for $n > \max(n_0, 4)$ we get:

$$N_n^2 = {}^{\downarrow}N_n^2 + {}^{\otimes}N_n^2 \geq (\frac{1}{2} - \frac{\sqrt{5}}{10} - \varepsilon)F_n^2 + 2\varepsilon F_n^2 = (\frac{1}{2} - \frac{\sqrt{5}}{10} + \varepsilon)F_n^2$$

The lower bound of $\frac{T_n^2}{F_n^2}$ is greater then 0 and can be obtained from Theorem 6.3, so altogether

$$\frac{1}{3} \leq \frac{T_n^2}{F_n^2} \leq \frac{1}{2} + \frac{\sqrt{5}}{10} - \frac{1}{96} < \frac{1}{2} + \frac{\sqrt{5}}{10}$$

which completes the proof. □

6 Types built with k atomic types

In this section we would like to investigate the asymptotic probability of finding inhabited types among types built with k variables. We will see that for any fixed k this probability is positive but the limit of those probabilities as $k \to \infty$ is zero.

By T_n^k and N_n^k we mean the numbers of intuitionistic tautologies and nontautologies, respectively, of length n over a k-letter alphabet. Of course $T_n^k + N_n^k = F_n^k$ and

$$F_n^k = \sum_{i=1}^{n-1} F_i^k \, F_{n-i}^k. \qquad (11)$$

Definition 6.1. *By a simple tautology we mean a type $\tau = \tau_1, ..., \tau_n \to \alpha$ for some ground type α, such that there is at least one component τ_i identical with α. Obviously, simple tautology is a tautology with a proof being a projection $\lambda x_1 ... x_n . x_i$. Let G_n^k be a number of simple tautologies of length n built with k ground type variables.*

Lemma 6.2. *The number G_n^k of simple tautologies is given by a recursive (on n) formula:*

$$\begin{aligned} G_1^k &= 0 \\ G_2^k &= k \\ G_n^k &= \sum_{i=2}^{n-1} F_{n-i}^k G_i^k + (F_{n-1}^k - G_{n-1}^k) \end{aligned}$$

Proof. The proof is based on two observations: First, $\tau_1 \to \tau_2$ is simple if τ_2 is simple. So for every type τ_1 of length $n-i$ and every simple tautology τ_2 of length i we can have one simple tautology $\tau_1 \to \tau_2$ of length n. The sum starts from $i = 2$ because there are no simple tautologies of length 1. This part is responsible for the component $\sum_{i=2}^{n-1} F_{n-i}^k G_i^k$. Additional and only other possible simple tautologies are such that τ_1 is a ground type the same as the ground type that τ_2 points to. Therefore for every formula μ of length $n-1$ which is not a simple tautology (there are exactly $F_{n-1}^k - G_{n-1}^k$

such formulas) we have exactly one simple tautology $a \to \mu$ where α is a ground type the type μ proves. □

Now we present a theorem about the asymptotic probability of the fact that a given type is a simple tautology, i.e. it has a proof which is a projection.

Theorem 6.3. *The limit of the sequence G_n^k/F_n^k exists and is equal to $(4k+1)/(2k+1)^2$.*

Proof. See Section 7 □

Corollary 6.4. *For every $k \geq 1$ the asymptotic probability of finding an inhabited type in the set of all types built with k variables is positive, i.e. for any fixed $k \geq 1$ $\liminf_{n\to\infty} \frac{T_n^k}{F_n^k} > 0$.*

Proof. For $k = 1$ the result was obtained in Theorem 4.6:

$$\liminf_{n\to\infty} \frac{T_n^1}{F_n^1} = \frac{1}{2} + \frac{\sqrt{5}}{10}.$$

For $k > 1$ the set of simple tautologies happens to be big enough to show the result. From Theorem 6.3 it follows that for all k:

$$\liminf_{n\to\infty} \frac{T_n^k}{F_n^k} \geq \lim_{n\to\infty} \frac{G_n^k}{F_n^k} = \frac{4k+1}{(2k+1)^2} > 0$$

□

Definition 6.5. *By "α-type" for some ground type α, we mean any type of the form $\tau_1, ..., \tau_n \to \alpha$. The type α is also an "α-type" with $n = 0$.*

By a simple nontautology we mean the "α-type" τ such that all components of τ are not "α-types", for any ground type α.

It is quite obvious that a simple nontautology is neither an intuitionistic nor classical tautology. Suppose the type $\tau_1, ..., \tau_n \to \alpha$ is a simple nontautology. Thus all τ_i are "α_i-types" where $\alpha_i \neq \alpha$ for all $i \leq n$. Just put the 0-1 valuation to τ which evaluates all α_i to 1 and α to 0. Therefore τ is not classically valid and by Lemma 2.1 τ is not inhabited or τ is not an intuitionistic tautology.

Lemma 6.6. *The number of simple nontautologies of length n built with k ground-type variables with p premises is*

$$SN_n^k(p) = \left(\frac{k-1}{k}\right)^p F_n^k(p)\,. \tag{12}$$

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

$$(... \to \bigcirc), (... \to \bigcirc), ..., (... \to \bigcirc) \to \bigcirc$$

in which ◯ are places in the type, left for some ground types to be filled in. There are exactly $p+1$ such places. There are $k\,(k-1)^p$ ways of filling the pattern in the way to obtain a simple nontatutology. So altogether there are

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

simple nontautologies with p premises. □

As in the case of tautologies, the number of simple nontautologies of the length n built with k ground type variables is

$$SN_n^k = \sum_{p=1}^{n-1} SN_n^k(p).$$

To calculate $\lim_{n\to\infty}(SN_n^k/F_n^k)$ let us first prove the auxiliary lemma:

Lemma 6.7.

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

Proof. Consider the series

$$\sum_{p=1}^{\infty} F_p(n), \tag{13}$$

where $F_p(n) = \left(\frac{k-1}{k}\right)^p\frac{C_n(p)}{C_n}$. For all n the quotients $\frac{C_n(p)}{C_n}$ are bounded above by 1. Thus for any term of (13) we have the inequality

$$\left(\frac{k-1}{k}\right)^p\frac{C_n(p)}{C_n} \le \left(\frac{k-1}{k}\right)^p.$$

The series $\sum_{p=1}^{\infty}\left(\frac{k-1}{k}\right)^p$ converges, thus (13) converges uniformly in $\mathbb{N}$. Since for a uniformly convergent series, the limit of its sum exists and is equal to the sum of the series of the limits of the elements, we can write:

$$\begin{aligned}\lim_{n\to\infty}\sum_{p=1}^{n-1}\left(\frac{k-1}{k}\right)^p\frac{C_n(p)}{C_n} &= \sum_{p=1}^{\infty}\lim_{n\to\infty}\left(\frac{k-1}{k}\right)^p\frac{C_n(p)}{C_n} && \text{(see (*))}\\ &= \sum_{p=1}^{\infty}\left(\frac{k-1}{k}\right)^p\frac{p}{2^{p+1}} && \text{(by Lemma 3.5)}\\ &= \frac{k(k-1)}{(k+1)^2}\end{aligned}$$

(*) $C_n(p) = 0$ for $p \geq n$, since a formula of length n can contain at most $n-1$ premises. □

In Theorem 6.3 we have shown the limit (as n tends to infinity) of the fraction of simple tautologies against all formulas of length n. Now we present the analogical theorem for simple nontautologies.

Theorem 6.8. *The limit of the sequence SN_n^k/F_n^k exists and is equal to $k(k-1)/(k+1)^2$.*

Proof. It is obvious that:

$$\begin{aligned} SN_n^k(p) &= \left(\frac{k-1}{k}\right)^p F_n^k(p) & \text{(by Lemma 6.6)} \\ &= \left(\frac{k-1}{k}\right)^p k^n C_n(p) \\ SN_n^k &= \sum_{p=1}^{n-1} \left(\frac{k-1}{k}\right)^p k^n C_n(p) \end{aligned}$$

Thus

$$\begin{aligned} \lim_{n\to\infty} \frac{SN_n^k}{F_n^k} &= \lim_{n\to\infty} \frac{\sum_{p=1}^{n-1} \left(\frac{k-1}{k}\right)^p k^n C_n(p)}{k^n C_n} \\ &= \lim_{n\to\infty} \sum_{p=1}^{n-1} \left(\frac{k-1}{k}\right)^p \frac{C_n(p)}{C_n} \\ &= \sum_{p=1}^{\infty} \left(\frac{k-1}{k}\right)^p \frac{p}{2^{p+1}} & \text{(by Lemma 6.7)} \\ &= \frac{k(k-1)}{(k+1)^2} \end{aligned}$$

□

Corollary 6.9.

$$\lim_{k\to\infty} \lim_{n\to\infty} \frac{SN_n^k}{F_n^k} = 1$$

It is quite interesting to note that the number of simple nontautologies is growing as k tends to infinity and asymptotically, all formulas are simple nontautologies.

Corollary 6.10. *For all $k \geq 1$,* $\limsup_{n\to\infty} \frac{T_n^k}{F_n^k} \leq \frac{3k+1}{(k+1)^2}$

Proof. Since $T_n^k = F_n^k - N_n^k$ and $N_n^k > SN_n^k$, thus we can calculate the limit of $\frac{T_n^k}{F_n^k}$ in the following way:

$$\begin{aligned}
\limsup_{n\to\infty} \frac{T_n^k}{F_n^k} &= \limsup_{n\to\infty} \frac{F_n^k - N_n^k}{F_n^k} \\
&\leq 1 - \lim_{n\to\infty} \frac{SN_n^k}{F_n^k} \\
&= 1 - \frac{k(k-1)}{(k+1)^2} \qquad \text{(by Theorem 6.8)} \\
&= \frac{3k+1}{(k+1)^2}
\end{aligned}$$

□

7 Types and generating functions

In this section we have gathered three proofs making substantial use of the generating functions. They are ordered here not by their order of appearance in the paper, but by their difficulty and their internal relationships. We suggest that the reader should read them in the order they are presented here.

The first proof is a standard one, done by hand. However, in the other two proofs we have decided to include the (fully commented in *this font*) outputs of two *Maple*™ V R3 sessions, in which we have computed all what we needed to establish the existence and determine the values of the limits. There are several reasons for it: first of all, the manipulations *Maple* has done for us, although tedious, are quite elementary and we have indeed verified all of them by hand, as well. Further, we have made the *Maple* worksheets with the calculations available electronically, so the reader can download them, and repeat the computations setting the tracing on so that all the details of the calculations show up in the output. Finally, we feel that the idea of letting computers do the hard computational work is worth being popularised. And, after all, it is perhaps better to admit that we have used a computer than to say "it is easy to verify that ..." every few lines.

Let us recall what we are aiming at.

Lemma 3.5.

$$\lim_{n\to\infty} C_n(p)/C_n = p/2^{p+1}.$$

Theorem 6.3.

$$\lim_{n\to\infty} G_n^k/F_n^k = \frac{4k+1}{(2k+1)^2}.$$

Theorem 4.6.

$$\lim_{n\to\infty} T_n^1/C_n = \frac{1}{2} + \frac{\sqrt{5}}{10}.$$

We will frequently need the following small lemma, whose proof can be obtained, e.g., by comparing (14) derived below with (5) and using Stirling's Formula.

Lemma 7.1.

$$\binom{1/2}{n} \sim c \cdot n^{-3/2}.$$

Proof. (Proof of Lemma 3.5.)

First of all, let us note that for a fixed p the generating function for $C_n(p)$ is

$$f_p(x) = x(f_C(x))^p = x\left(\frac{1-\sqrt{1-4x}}{2}\right)^p$$

Indeed, the recursion $C_n(p) = \sum_{i=1}^{n-p} C_i C_{n-i}(p-1)$ becomes, after a closer examination, the equality $f_p(x) \cdot f_C(x) = f_{p+1}(x)$ and since $C_n(1) = C_{n-1}$ obviously $f_1(x) = x(f_C(x))$.

We substitute $z = 4x$ in f_p, which amounts to dividing the n-th coefficient by 4^n. We get $f(z) = (z/4)(1/2 - \sqrt{1-z}/2)^p$.

We want to find an approximation of the coefficients of f with an error term $O(n^{-2})$. To do so, we apply the Szegö Lemma to the function $4f/z = (1/2 - \sqrt{1-z}/2)^p$, in which the coefficients are multiplied by 4 and shifted by one to the left, when compared to f. Of course, the only singularity of $4f/z$ in the whole complex plane is $z = 1$.

The expansion (2) becomes by the Newton Binomial Formula

$$4f/z = \sum_{j=0}^{p} \binom{p}{j} \frac{(-1)^j}{2^p}(1-z)^{j/2} = \sum_{j=0}^{p} \binom{p}{j} \frac{(-1)^j}{2^p}(1-z)^{j/2},$$

and our $\xi(q)$ is 2. Hence by (3) we get

$$[z^n]\{4f/z\} = \binom{p}{0}\frac{(-1)^n}{2^p}\binom{0}{n} - \binom{p}{1}\frac{p(-1)^n}{2^p}\binom{1/2}{n} + \binom{p}{2}\frac{(-1)^n}{2^p}\binom{1}{n} + O(n^{-2}),$$

where $\binom{0}{n} = \binom{1}{n} = 0$ for $n > 1$. Therefore $[z^n]\{4f/z\} = -\frac{p(-1)^n}{2^p}\binom{1/2}{n} + O(n^{-2})$. Undoing our substitutions and modifications we find $[x^n]\{f_p\} = 4^{n-1}[z^{n-1}]\{4f/z\}$, so

$$
\begin{aligned}
[x^n]\{f_p\} &= -\left(\frac{p(-1)^{n-1}}{2^p}\binom{1/2}{n-1} + O(n^{-2})\right)4^{n-1} \\
&= -\left(\frac{p(-1)^{n-1}}{2^p}\binom{1/2}{n}\frac{n}{1/2-(n-1)} + O(n^{-2})\right)4^{n-1} \\
&= -\left(\frac{p(-1)^{n-1}}{2^p}\binom{1/2}{n}(-1+O(n^{-1})) + O(n^{-2})\right)4^{n-1} \\
&= \left(\frac{p(-1)^{n-1}}{2^p}\binom{1/2}{n} + O(n^{-2})\right)4^{n-1}.
\end{aligned}
$$

On the other hand, we remember that $C_n(1) = C_{n-1}$, so $C_n = C_{n+1}(1)$, and we can use this formula for calculating the quotient $C_n(p)/C_n$.

$$
\begin{aligned}
\frac{C_n(p)}{C_n} &= \frac{\left(\frac{p(-1)^{n-1}}{2^p}\binom{1/2}{n} + O(n^{-2})\right)4^{n-1}}{\left(\frac{(-1)^n}{2}\binom{1/2}{n+1} + O(n^{-2})\right)4^n} \\
&= \frac{\left(\frac{p(-1)^{n-1}}{2^p}\binom{1/2}{n} + O(n^{-2})\right)4^{n-1}}{\left(\frac{(-1)^{n-1}}{2}\binom{1/2}{n} + O(n^{-2})\right)4^n} \\
&= \frac{\frac{p}{2^p}\binom{1/2}{n} + O(n^{-2})}{2\binom{1/2}{n} + O(n^{-2})} \\
&= \frac{p}{2^{p+1}}
\end{aligned}
$$

because $\binom{1/2}{n} \sim cn^{-3/2}$. This finishes the proof, but before we close it let us note the following useful equality, whose proof appears in the denominator of the fractions in the last calculation.

$$
C_n = (2(-1)^{n-1}\binom{1/2}{n} + O(n^{-2}))4^{n-1}. \tag{14}
$$

□

Proof. (Proof of Theorem 6.3.)

Denote $f_F(z) = \sum_{n=1}^{\infty} F_n^k z^n$.

The recursion

$$
F_n^k = \sum_{i=1}^{n-1} F_i^k F_{n-i}^k
$$

becomes, after a closer examination, the equality

$$
f_F \cdot f_F = f_F - k \cdot z, \tag{15}
$$

since the recursion exactly corresponds to the multiplication of power series, with the exception that the linear term kz disappears.

Similarly, assuming $f_G(z) = \sum_{n=1}^{\infty} G_n^k z^n$, the recursion

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

becomes

$$f_G = f_G \cdot f_F + z \cdot f_F - z \cdot f_G. \tag{16}$$

Now we can turn *Maple* on. The worksheet can be downloaded from `ftp://ftp.cse.unsw.edu.au/pub/doc/papers/UNSW/9808.Thm63.ms`.

First we solve (15).

```
> solve(fF=fF*fF+k*x,fF);
```

$$\frac{1}{2} - \frac{1}{2}\sqrt{1-4\,k\,x}, \frac{1}{2} + \frac{1}{2}\sqrt{1-4\,k\,x}$$

The right solution is the first one, since the value of f_F at 0 should be 0.

```
> fF:=op(1,["]);
```

$$fF := \frac{1}{2} - \frac{1}{2}\sqrt{1-4\,k\,x}$$

Now we determine the generating function for G_n^k, solving (16).

```
> solve(fG=fG*fF+fF*x-fG*x,fG);
```

$$-\frac{-\frac{1}{2}x + \frac{1}{2}\sqrt{1-4\,k\,x}\,x}{\frac{1}{2} + \frac{1}{2}\sqrt{1-4\,k\,x} + x}$$

```
> fG:=";
```

$$fG := -\frac{-\frac{1}{2}x + \frac{1}{2}\sqrt{1-4\,k\,x}\,x}{\frac{1}{2} + \frac{1}{2}\sqrt{1-4\,k\,x} + x}$$

Verification.

```
> series(fG,x=0);
```

$$k\,x^2 + (\,k^2 - k\,(\,-k+1\,)\,)\,x^3 + (\,3\,k^3 - (\,2\,k^2 - k\,)\,(\,-k+1\,)\,)\,x^4 + \\ (\,7\,k^4 + (\,2\,k^2 - k\,)\,k^2 - (\,5\,k^3 - 3\,k^2 + k\,)\,(\,-k+1\,)\,)\,x^5 + \mathrm{O}(\,x^6\,)$$

Everything is OK. The initial coefficients of the series agree with the initial terms of the sequence we are considering, and since the function satisfies the equation, it must be the generating function we are looking for. But indeed we didn't have to verify. The equation to be solved is of degree 1, so it has exactly one solution, and if we have got a solution, it is the only solution, and at the same time the generating function we are looking for.

Substitution which makes everything happen in the vicinity of $z = 1$, *corresponds to the division of the coefficients by* $(4k)^n$ *both in the numerator and in the denominator of the fraction whose limit we are going to compute.*

```
> fG:=subs(z=t/(4*k),fG);
```

$$fG := -\frac{\frac{1}{8}\frac{t\sqrt{-t+1}}{k} - \frac{1}{8}\frac{t}{k}}{\frac{1}{2} + \frac{1}{2}\sqrt{-t+1} + \frac{1}{4}\frac{t}{k}}$$

```
> simplify(");
```

$$-\frac{1}{2}\frac{t\left(\sqrt{-t+1}-1\right)}{2k + 2\sqrt{-t+1}\,k + t}$$

We consider f_G *as a function of* $z = \sqrt{1-t}$. *In particular,* $t = 1 - z^2$.

```
> fH:=subs(sqrt(-t+1)=z,t=1-z^2,fG);
```

$$fH := -\frac{\frac{1}{8}\frac{(1-z^2)\,z}{k} - \frac{1}{8}\frac{1-z^2}{k}}{\frac{1}{2} + \frac{1}{2}z + \frac{1}{4}\frac{1-z^2}{k}}$$

```
> fH:=simplify(fH);
```

$$fH := \frac{1}{2}\frac{z^2 - 2z + 1}{-z + 2k + 1}$$

Let us note that the only singularity of f_H *is located at* $z = 2k + 1$. *Hence we see that the only singularity of* f_G *located at the circle* $|t| = 1$ *is indeed* $t = 1$. *The other singularity, located where the denominator becomes* 0, *is achieved when* $\sqrt{1-t} = 2k + 1$, *i.e., for* t *of a modulus substantially greater than* 1.

We convert f_H *into a sum of simple fractions.*

```
> fH:=convert(fH,parfrac,z);
```

$$fH := -k - \frac{1}{2}z + \frac{1}{2} + 2\frac{k^2}{-z + 2k + 1}$$

Again we can ignore constant terms in the simple fractions representation. The coefficients $[x^n]\{-\frac{1}{2}z\}$ *(remember about the substitutions made) are the same as those of* f_F, *so the give us an additive term 1 always. So what* *remains is to compute* $a = \lim_{n\to\infty}[x^n]\{2k^2/(-z+2k+1)\}/[x^n]\{-z/2\}$, *and then the answer we are looking for is* $1 + a$.

```
> fK:=op(4,fH);
```

$$fK := 2\,\frac{k^2}{-z + 2\,k + 1}$$

```
> S:=series(fK,z=0);
```

$$S := 2\,\frac{k^2}{2\,k+1} + 2\,\frac{k^2}{(\,2\,k+1\,)^2}\,z + 2\,\frac{k^2}{(\,2\,k+1\,)^3}\,z^2 + 2\,\frac{k^2}{(\,2\,k+1\,)^4}\,z^3 + 2\,\frac{k^2}{(\,2\,k+1\,)^5}\,z^4 + 2\,\frac{k^2}{(\,2\,k+1\,)^6}\,z^5 + \mathrm{O}(\,z^6\,)$$

By the comment made a few steps above, f_G *has only one singularity* $t = 1$ *on* $|t| = 1$, *as required by (our version of) Szegö Lemma.*

Hence our expansion (2) *of* f_G *becomes*

```
> subs(z=sqrt(1-t),S);
```

$$2\,\frac{k^2}{2\,k+1} + 2\,\frac{k^2\,\sqrt{-t+1}}{(\,2\,k+1\,)^2} + 2\,\frac{k^2\,(\,-t+1\,)}{(\,2\,k+1\,)^3} + 2\,\frac{k^2\,(\,-t+1\,)^{3/2}}{(\,2\,k+1\,)^4} + 2\,\frac{k^2\,(\,-t+1\,)^2}{(\,2\,k+1\,)^5} + 2\,\frac{k^2\,(\,-t+1\,)^{5/2}}{(\,2\,k+1\,)^6} + \mathrm{O}\left((\,-t+1\,)^3\right)$$

Therefore, with $\xi(q) = 2$ *as before and with all substitutions undone, the expansion* (3) *becomes*

$$\left(\frac{2k^2}{(2k+1)^2}(-1)^n\binom{1/2}{n} + O(n^{-2})\right)(4k)^n.$$

Calculating the quotient K_n^k/F_n^k *and taking into account that* $F_n^k = \left(2(-1)^{n-1}\binom{1/2}{n} + O(n^{-2})\right)4^{n-1}k^n$ *by* (14) *we get*

$$\begin{aligned}\frac{K_n^k}{F_n^k} &= \frac{(\,2k^2/(2k+1)^2(-1)^n\binom{1/2}{n} + O(n^{-2})\,)(4k)^n}{(\,2(-1)^{n-1}\binom{1/2}{n} + O(n^{-2})\,)4^{n-1}k^n}\\ &= -\frac{2k^2/(2k+1)^2}{1/2}(1+o(1))\\ &= -\frac{4k^2}{(2k+1)^2}(1+o(1)).\end{aligned}$$

Consequently $a = -4k^2/(2k+1)^2$ *and the limit we are looking for is* $1 + a = \ldots$

```
> 1-(4*k^2/(2*k+1)^2);
```

$$1 - 4\,\frac{k^2}{(\,2\,k+1\,)^2}$$

```
> simplify(");
```

$$\frac{4\,k+1}{(\,2\,k+1\,)^2}$$

□

Proof. (Proof of Theorem 4.6).

The recursion for N_n^1 and T_n^1 gives the following two identities for their generating functions $f_T = \sum_{n=1}^{\infty} T_n^1 z^n$ and $f_N = \sum_{n=1}^{\infty} N_n^1 z^n$, where F_C is f_F with $k = 1$, of course.

$$\begin{aligned} f_N &= f_N(f_C - f_N) + z \\ f_T &= f_C - f_N. \end{aligned} \tag{17}$$

Now we can turn *Maple* on again. The worksheet can be downloaded from `ftp://ftp.cse.unsw.edu.au/pub/doc/papers/UNSW/9808.Thm46.ms`.

We assume f_F to be already known to Maple.

```
> fC:=subs(k=1,fF);
```

$$fC := \frac{1}{2} - \frac{1}{2}\sqrt{1-4x}$$

Now we solve the equation defining f_N.

```
> solve(fN=fN*(fC-fN)+x,fN);
```

$$-\frac{1}{4} - \frac{1}{4}\sqrt{1-4x} + \frac{1}{4}\sqrt{2+2\sqrt{1-4x}+12x},$$
$$-\frac{1}{4} - \frac{1}{4}\sqrt{1-4x} - \frac{1}{4}\sqrt{2+2\sqrt{1-4x}+12x}$$

Since the solution should be 0 for $x = 0$, we have to choose the first solution. Since the equation is of degree 2 and the solver has found two solutions, there is no risk it has overlooked any. So we do not need to verify the chosen solution.

```
> fN:=op(1,["]);
```

$$fN := -\frac{1}{4} - \frac{1}{4}\sqrt{1-4x} + \frac{1}{4}\sqrt{2+2\sqrt{1-4x}+12x}$$

```
> fT:=fC-fN;
```

$$fT := \frac{3}{4} - \frac{1}{4}\sqrt{1-4x} - \frac{1}{4}\sqrt{2+2\sqrt{1-4x}+12x}$$

We remove the constant term from the functions f_T, since this change corresponds to the change of the first coefficient of the power series, and thus do not affect the asymptotical behaviour of the coefficients.

```
> fT:=fT-3/4;
```

$$fT := -\frac{1}{4}\sqrt{1-4x} - \frac{1}{4}\sqrt{2+2\sqrt{1-4x}+12x}$$

Substitution which makes everything happen in the vicinity of 1, corresponds to the division of the coefficients by 4^n both in the numerator and in the denominator of the fraction whose limit we are going to compute.

```
> fR:=subs(x=z/4,fT);
```

$$fQ := -\frac{1}{4}\sqrt{1-z} - \frac{1}{4}\sqrt{2+2\sqrt{1-z}+3\,z}$$

We can further simplify f_T, removing the term $-\frac{1}{4}\sqrt{1-z}$ from it, since in the quotient $[x^n]\{f_T\}/[x^n]\{f_C\}$ we are going to compute it represents a constant additive term equal $1/2$.

```
> fS:=op(2,fQ);
```

$$fS := -\frac{1}{4}\sqrt{2+2\sqrt{1-z}+3\,z}$$

Now we compute the limit $a = \lim_{n\to\infty}[x^n]\{f_S\}/[x^n]\{f_C\}$ and the limit $\lim_{n\to\infty}[x^n]\{f_T\}/[x^n]\{f_C\}$ we are looking for is $\frac{1}{2}+a$.

But to use Szegö Lemma we must know the singularities of F_S on $|z|=1$. Apart from $z=1$ there can be only a singularity when the expression under the root sign becomes 0. So let us check where it happens.

```
> eq:=op(1,op(2,fS))=0;
```

$$eq := 2+2\sqrt{1-z}+3\,z = 0$$

```
> solve(eq,z);
```

$$\frac{-16}{9}$$

So the only singularity of f_S on $|z|=1$ is indeed $z=1$.

And a comment for those who are surprised: eq is an equation of degree 2 in $t=\sqrt{1-z}$, which has two solutions, but one of them is a real negative one, and is never assumed by the root. The other is real positive and corresponds to $z=-16/9$, which we have found.

We consider f_T as a function of $t=\sqrt{1-z}$. In particular, $z=1-t^2$.

```
> fR:=subs(sqrt(1-z)=t,z=1-t^2,fS);
```

$$fR := -\frac{1}{4}\sqrt{5+2\,t-3\,t^2}$$

```
> S:=series(fR,t=0);
```

$$S := -\frac{1}{4}\sqrt{5} - \frac{1}{20}\sqrt{5}\,t + \frac{2}{25}\sqrt{5}\,t^2 - \frac{2}{125}\sqrt{5}\,t^3 + \frac{2}{125}\sqrt{5}\,t^4 - \frac{26}{3125}\sqrt{5}\,t^5 + \mathrm{O}(t^6)$$

```
> subs(t=sqrt(1-z),S);
```

$$-\frac{1}{4}\sqrt{5}-\frac{1}{20}\sqrt{5}\sqrt{1-z}+\frac{2}{25}\sqrt{5}\,(1-z)-\frac{2}{125}\sqrt{5}\,(1-z)^{3/2}$$
$$+\frac{2}{125}\sqrt{5}\,(1-z)^{2}-\frac{26}{3125}\sqrt{5}\,(1-z)^{5/2}+\mathrm{O}\left((1-z)^{3}\right)$$

So the above is the expansion (2). And, as before, the expansion (3) becomes

$$-\frac{\sqrt{5}}{20}(-1)^n(4k)^n+O(n^{-2}).$$

Again as in the previous proof we get

$$\begin{aligned}\frac{S_n}{C_n} &= \frac{(-\frac{\sqrt{5}}{20}(-1)^n\binom{1/2}{n}+O(n^{-2}))4^n}{(2(-1)^{n-1}\binom{1/2}{n}+O(n^{-2}))4^{n-1}}\\ &= \frac{4\sqrt{5}/20}{2}(1+o(1))\\ &= \frac{\sqrt{5}}{10}(1+o(1)).\end{aligned}$$

Consequently $a=\sqrt{5}/10$ *and the limit we are looking for is* $1/2+a=\ldots$

```
> 1/2+sqrt(5)/10;
```

$$\frac{1}{2}+\frac{1}{10}\sqrt{5}$$

```
> evalf(");
```

$$.7236067978$$

□

8 Conclusions and open problems

Let us sum up the presented results:

1. There exists asymptotic probability for one-letter tautologies (one-letter inhabited types) $\lim_{n\to\infty}\frac{T_n^1}{F_n^1}=1/2+\sqrt{5}/10$ (Theorem 4.6)
2. There are fewer two-letters tautologies (two-letters inhabited types) than one-letter tautologies $\limsup_{n\to\infty}\frac{T_n^2}{F_n^2}<\lim_{n\to\infty}\frac{T_n^1}{F_n^1}$ (Theorem 5.7)
3. There exists asymptotic probability for simple tautologies for any number of ground types $\lim_{n\to\infty}\frac{G_n^k}{F_n^k}=\frac{4k+1}{(2k+1)^2}$ (Theorem 6.3)

4. There exists asymptotic probability for simple nontautologies for any number of ground types $\lim_{n\to\infty} \frac{SN_n^k}{F_n^k} = \frac{k(k-1)}{(k+1)^2}$ (Theorem 6.8)

5. Asymptotically $\frac{4k+1}{(2k+1)^2} < \frac{T^k}{F^k} < \frac{3k+1}{(k+1)^2}$ (Corollaries 6.4 and 6.10).

As we can see, the gap between asymptotic lower and upper bounds for T_n^k/F_n^k is relatively small and asymptotically the upper bound is 3 times bigger then the lower bound. The main problem is close this gap and prove the convergence of T_n^k/F_n^k for all $k \geq 2$. However the problem seems difficult since it requires a syntactic description of the notion of a tautology which is inherently semantical.

Notice that all results are valid for both classical and intuitionistic logic, i.e. we can write

$$\frac{4k+1}{(2k+1)^2} \leq \liminf_{n\to\infty} \frac{CL_n^k}{F_n^k} \leq \limsup_{n\to\infty} \frac{CL_n^k}{F_n^k} \leq \frac{3k+1}{(k+1)^2},$$

where CL_n^k is the number of classical tautologies of length n with k letters. This is because simple tautologies are classical tautologies, and simple nontautologies are neither intuitionistic nor classical tautologies. Thus we can calculate the fraction of intuitionistic tautologies among classical tautologies.

Theorem 8.1.

$$\lim_{k\to\infty} \liminf_{n\to\infty} \frac{T_n^k}{CL_n^k} \geq \frac{1}{3}$$

Proof.

$$\limsup_{n\to\infty} \frac{CL_n^k}{T_n^k} \leq \limsup_{n\to\infty} \frac{CL_n^k}{G_n^k} = \limsup_{n\to\infty} \frac{CL_n^k}{F_n^k} / \lim_{n\to\infty} \frac{G_n^k}{F_n^k}$$
$$\leq \frac{3k+1}{(k+1)^2} / \frac{4k+1}{4k^2+4k+1} \overset{k\to\infty}{\longrightarrow} 3$$

Thus

$$\lim_{k\to\infty} \liminf_{n\to\infty} \frac{T_n^k}{CL_n^k} \geq \frac{1}{3}.$$

□

The result 1/3 has been obtained using very rough estimations. However, basing on our experiences of counting the intuitionistic tautologies, we strongly believe that the limit of T_n^k/CL_n^k exists and is equal to one. Although intuitionistic logic attracts the attention of logicians and computer scientists, it is often underestimated by classical mathematicians. Such a result would prove that a random classical tautology has an intuitionistic proof with the probability 1. Moreover, we believe that the proportion of simple tautologies to intuitionistic tautologies tends to one as the length of formulas tends to infinity. It would be an even more unexpected result, because it would show that, in fact, a randomly chosen proof is a projection and statistically all true statements are the trivial ones. At the moment, these are just expectations but certainly it is worth to look in this direction.

Conjecture 1.

1. *For every* $k \geq 1$ $\lim_{n\to\infty} \frac{T_n^k}{CL_n^k}$ *exists.*

2. *Moreover,* $\lim_{k\to\infty} \lim_{n\to\infty} \frac{T_n^k}{CL_n^k} = 1$.

Conjecture 2.

1. *For every* $k \geq 1$ $\lim_{n\to\infty} \frac{G_n^k}{T_n^k}$ *exists.*

2. *Moreover,* $\lim_{k\to\infty} \lim_{n\to\infty} \frac{G_n^k}{T_n^k} = 1$.

References

[1] Barendregt, H.P. *The lambda calculus. Its syntax and semantics.* Revised edition. Studies in Logic and the Foundations of Mathematics, 103, North-Holland, Amsterdam-New York, 1984.

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

[3] Hindley, J.R. *Basic Simple Type Theory.* Cambridge Tracts in Theoretical Computer Science, **42** Cambridge University Press, 1997.

[4] Statman, R. *On the existence of closed terms in the typed λ-calculus.* In: Hindley, J.R., and Seldin, J., (eds.) Combinatory Logic, Lambda Calculus and Formalism, Academic Press, New York, 1980.

[5] Szegö, G. *Orthogonal polynomials.* Fourth edition. AMS, Colloquium Publications, Vol. 23, Providence, 1975.

[6] Wolfram, D.A. *The Clausual Theory of Types.* Cambridge Tracts in Theoretical Computer Science **21** Cambridge University Press 1993.

[7] Wilf, H.S. *Generatingfunctionology.* Second edition. Academic Press, Boston, 1994.