STATISTICAL PROPERTIES OF TYPES

Marek Zaionc
Computer Science Department, Jagiellonian University,
Nawojki 11, 30-072 Krakow, Poland
E-mail zaionc@ii.uj.edu.pl

June 18, 1998

Abstract

We consider types and typed lambda calculus over the 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 lim or lim sup and lim inf of that fraction when $n \to \infty$. The answer to this question is equivalent with finding the "density" of inhabited types in the set of all types, or so called asymptotic probability of finding 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 $72\% \leq limsup \leq 73\%$. This means that the random type (formula) of the large size is as likely as 73% to be inhabited (tautology). We prove also that for every finite number k of ground type variables the density of inhabited types is always positive but is decreasing to 0 with k going to infinity.

1 Simple Typed λ-Calculus

We shall consider a simple typed λ-calculus with a finite ground types. The set $TYPES$ is 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 $\mu, \tau_1, \tau_2, ..., \tau_n$ 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 $\tau_1 \to ... \to \tau_n \to a$ where a is a ground type. If τ has the form $\tau_1, ..., \to \tau_n \to a$ then τ_i for $i \leq n$ are called components 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 the length of type τ we mean a 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 [?]). 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 [?]). 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 [?], [?] or more general [?]. We are going to use Curry-Howard isomorphism and we will not distinguished 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 1.1 *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 no matter if $\tau[i]$ is a ground type or not. The proof of the lemma is by induction on the complexity $\pi(T)$. Let the type τ has a form $\tau[1],...,\tau[n] \to a$ where a 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 $\tau[1],...,a,...,\tau[n] \to a$ and this is of course classically valid formula. Suppose the term T is in long normal form $\lambda x_1...x_n.x_iK_1...K_p$. 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. 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 $\tau[i] = \tau[i,1],...,\tau[i,p] \to a$. 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 a is true.

2 Types with one Ground Type

We start the discussion about the density of inhabited types from the types build up from just one atomic type a.

Theorem 2.1 *τ is a tautology iff $\tau \to a$ is not a tautology.*

Proof. $(\to)$ Since τ is a tautology then τ is a classical tautology. (see lemma ??). So $\tau \to a$ is not a classical tautology because of valuation which assign 0 to a. Again from lemma ?? it is not a tautology.

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

Theorem 2.2 *$\tau[1], ..., \tau[n] \to a$ is not a tautology iff all $\tau[i]$ are tautologies*

Proof. $(\to)$ By the same classical logic argument as in lemma ??. $(\leftarrow)$ Assume some $\tau[i]$ is not a tautology. So $\tau[i] \to a$ is a tautology and therefore $\tau[1], ..., \tau[n] \to a$ is a tautology.

Theorem 2.3 *$\tau \to \mu$ is not a tautology iff τ is a tautology and μ is not a tautology.*

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

3 Counting types (formulas)

Let us cosider types build up with the single ground type a. In this section we are going to count inhabited types of the given length n and compare their number with the total number of types.

Definition 3.1 *The total number of types over one ground type is given by so called Catalan number.*
By C_n we mean the n-th Catalan number which is given by recursion as

$$C_1 = 1$$

$$C_n = \sum_{i=1}^{n-1} C_i C_{n-i}$$

It is well known that Catalan number can be given by an equivalent formula

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

Lemma 3.2 *C_n is the number of different types of length n.*

Proof. For $n = 1$ it is obvious. We can divide the string of n ground types in to two strings at $i - th$ position. Since the number of types of the length i and $n - i$ is respectively C_i and C_{n-i} the total number of types of length n is $\sum_{i=1}^{n-1} C_i C_{n-i}$.

Lemma 3.3 *For every $n \geq 1$ and for every $k \geq 1$*

$$C_n / C_{n+k} > (1/4)^k$$

and moreover

$$\lim_{n \to \infty} C_n / C_{n+k} = (1/4)^k$$

Proof. From the observation made in the definition of Catalan numbers such that

$$C_n = \frac{1}{n} \begin{pmatrix} 2n-2 \\ n-1 \end{pmatrix}$$

we can easily calculate $C_n/C_{n+1} = \frac{1}{4} + \frac{3}{8n-4}$. So the function $n \rightarrow C_n/C_{n+1}$ is decreasin and $\lim_{n\rightarrow\infty} C_n/C_{n+1} = 1/4$. Therefore for every positive k $\lim_{n\rightarrow\infty} C_n/C_{n+k} = (1/4)^k$

Definition 3.4 *By number N_n and respectively T_n we mean the number of nontautologies and tautologies of the length n.*

Lemma 3.5 *Numbers N_n of nontautologies of the length n and T_n of tautologies of length n are given by formulas:*

$$N_1 = 1 \ ; \quad N_n = \sum_{i=1}^{n-1} T_i N_{n-i}$$

$$T_1 = 0 \qquad T_n = C_n - \sum_{i=1}^{n-1} T_i N_{n-i}$$

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

Bellow we present the list of numbers of tautologies, nontautologies and Catalan numbers and the fractions of T_n/C_n for the first few n.

n	T_n	N_n	C_n	T_n/C_n
1	0	1	1	0.0000
2	1	0	1	1.0000
3	1	1	2	0.5000
4	4	1	5	0.8000
5	9	5	14	0.6429
6	31	11	42	0.7381
7	91	41	132	0.6894
8	309	120	429	0.7203
9	1009	421	1430	0.7056
10	3481	1381	4862	0.7160
11	11956	4840	16796	0.7118
12	42065	16721	58786	0.7156
13	148655	59357	208012	0.7146
14	532039	210861	742900	0.7162
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
300	.	.	.	0.7232
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
2000				0.7236
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

Lemma 3.6 *The fraction* $\frac{T_n}{C_n}$ *has lower and upper bounds*

$$0.6 \leq \frac{T_n}{C_n} \leq 0.8$$

or equivalent

$$0.2 \leq \frac{N_n}{C_n} \leq 0.4$$

for all $n \geq 10$.

Proof. We are going to prove the second equivalent form. For the proof we will use induction which begins with $n = 10$ and in order to prove that $0.2 \leq \frac{N_n}{C_n} \leq 0.4$ we are going to use 4 inductive assumptions $0.2 \leq \frac{N_{n-1}}{C_{n-1}} \leq 0.4$ up to $0.2 \leq \frac{N_{n-4}}{C_{n-4}} \leq 0.4$ and their equivalent forms $0.6 \leq \frac{T_{n-1}}{C_{n-1}} \leq 0.8, \ldots, 0.6 \leq \frac{T_{n-4}}{C_{n-4}} \leq 0.8$. For fraction $\frac{T_5}{C_5}, \ldots, \frac{T_{10}}{C_{10}}$ we can check by hand that it belongs to the $[0.6, 0.8]$ set. (see for example raws 5 till 10 in the table included above). For proof we use formula from **??**.

upper bound for N_n for $n \geq 10$

$$
\begin{aligned}
\frac{N_n}{C_n} = \quad & [T_{n-1} + T_{n-3} + T_{n-4} + (\textstyle\sum_{i=5}^{n-5} T_i N_{n-i}) + \\
& + N_{n-2} + N_{n-3} + 4N_{n-4}]/C_n
\end{aligned}
$$

(Now we use our inductive assumptions:)

$$
\begin{aligned}
\leq \quad & [0.8C_{n-1} + 0.8C_{n-3} + 0.8C_{n-4} + 0.8 * 0.4(\textstyle\sum_{i=5}^{n-5} C_i C_{n-i}) + \\
& +0.4C_{n-2} + 0.4C_{n-3} + 0.4 * 4C_{n-4}]/C_n \\
= \quad & [0.8C_{n-1} + 0.4C_{n-2} + 1.2C_{n-3} + 2.4C_{n-4} + \\
& 0.32(\textstyle\sum_{i=5}^{n-5} C_i C_{n-i})]/C_n \\
= \quad & [0.8C_{n-1} + 0.4C_{n-2} + 1.2C_{n-3} + 2.4C_{n-4} + \\
& 0.32(\textstyle\sum_{i=1}^{n-1} C_i C_{n-i}) \\
& - 0.32[2C_{n-1} + 2C_{n-2} + 4C_{n-3} + 10C_{n-4}]]/C_n \\
= \quad & [0.32C_n + 0.16C_{n-1} - 0.24C_{n-2} - 0.08C_{n-3} - 0.8C_{n-4}]/C_n \\
= \quad & 0.32 + 0.16\tfrac{C_{n-1}}{C_n} - 0.24\tfrac{C_{n-2}}{C_n} - 0.08\tfrac{C_{n-3}}{C_n} - 0.8\tfrac{C_{n-4}}{C_n} \\
< \quad & 0.32 + 0.16\tfrac{C_9}{C_{10}} - 0.24\tfrac{1}{16} - 0.08\tfrac{1}{64} - 0.8\tfrac{1}{256} \\
= \quad & 0.32 + 0.16\tfrac{1430}{4862} - 0.24\tfrac{1}{16} - 0.08\tfrac{1}{64} - 0.8\tfrac{1}{256} = \tfrac{9457}{27200} < 0.4
\end{aligned}
$$

Lower bound for N_n for $n \geq 10$

$$
\begin{aligned}
\frac{N_n}{C_n} = \quad & [T_{n-1} + T_{n-3} + T_{n-4} + (\textstyle\sum_{i=5}^{n-5} T_i N_{n-i}) \\
& + N_{n-2} + N_{n-3} + 4N_{n-4}]/C_n
\end{aligned}
$$

Now we use our inductive assumptions:

$$
\begin{aligned}
\geq \quad & [0.6C_{n-1} + 0.6C_{n-3} + 0.6C_{n-4} + 0.6 * 0.2(\textstyle\sum_{i=5}^{n-5} C_i C_{n-i}) \\
& +0.2C_{n-2} + 0.2C_{n-3} + 0.24C_{n-4}]/C_n
\end{aligned}
$$

using the same technique as above finally we obtain:

$$
\begin{aligned}
= \quad & [0.12C_n + 0.36C_{n-1} - 0.04C_{n-2} + 0.32C_{n-3} + 0.2C_{n-4}]/C_n \\
= \quad & 0.12 + 0.36\tfrac{C_{n-1}}{C_n} - 0.04\tfrac{C_{n-2}}{C_n} + 0.32\tfrac{C_{n-3}}{C_n} + 0.2\tfrac{C_{n-4}}{C_n} \\
> \quad & 0.12 + 0.36\tfrac{1}{4} - 0.04\tfrac{C_8}{C_{10}} + 0.32\tfrac{1}{64} + 0.2\tfrac{1}{256} = \tfrac{46033}{217600} > 0.2
\end{aligned}
$$

Theorem 3.7 $0.72 \leq T_n/C_n \leq 0.73$ *or equivalently* $0.27 \leq N_n/C_n \leq 0.28$ *for* $n \geq 250$.

Proof. The proof of this theorem was done using exactly the same technique as proposed in lemma **??** but was computer assisted using *Mathematica* program. The lower and upper bounds for T_n/C_n holds from $n = 250$ but the from 251 till 500 it was checked by the program. Then starting from 501 the inductive step was checked by *Mathematica* useing formula which manipulate on 250 inductive assumptions. The details of the program together with the copy of the *Mathematica* session is enclosed in the supplement

4 Counting two letter types

In this section we are going to find the lower and upper bounds for asymptotic probabilities for the fraction of inhabited types of the types build up with two letters. We will see that the lower and upper bounds for fraction of tautologies (inhabited types) i.e the fraction T^2_n/F^2_n can be separated from the lower and upper bounds for fraction T^1_n/F^1_n for types build up with one ground type. This proves that asymptotically there is less tautologies (inhabited types) for two ground types than for one ground type, meaning that $\limsup T^2_n/F^2_n < \liminf T^1_n/F^1_n$. At the same time the fraction of two letter tautologies is asymptotically positive.

Definition 4.1 *By F^2_n we mean the number of two letter types. $F^2_n = 2^n C_n$ where C_n is the $n-th$ Catalan number. Formula for F^2_n is correct since there is 2^n sequences of the length n and each sequence can be turn in to the tape by C_n ways of putting brackets. By T^2_n and respectively N^2_n we mean the numbers of tautologies and nontautologies. Of course $T^2_n + N^2_n = F^2_n$*

Definition 4.2 *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 4.3 *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 4.4 *By ${}^{\otimes}N^2_n$ we mean the number of such two letter nontautologies which projections are one letter tautologies, and contrary ${}^{\downarrow}N^2_n$ means the number of such two letters nontautologies which projections are again one letter nontautologies. Because classes of types described by numbers ${}^{\downarrow}N^2_n$ and ${}^{\otimes}N^2_n$ are disjoin then simply ${}^{\downarrow}N^2_n + {}^{\otimes}N^2_n = N^2_n$.*

In the next theorem we will find the class of formulas which are nontautologies but their projections 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 4.5 *For every two letters tautology τ each formula in one of the forms*

$$\begin{array}{ll} \tau, a \to b & \tau, a \to a \\ a, \tau \to b & b, \tau \to a \end{array}$$

is an two letters nontautology while its projection is a tautology. Therefore ${}^{\otimes}N^2_n \geq 4T^2_{n-2}$ for $n \geq 4$.

Proof. It is easy to notice that all those formulas are nontautologies since they are even classically not valid. (see lemma **??**.) Each projection must be a tautology since letter a and b are getting unified. The inequality holds since all four categories are disjointed for $n \geq 4$.

Lemma 4.6 ${}^{\downarrow}N^2_n \geq \frac{27}{100}F^2_n$ *for* $n \geq 250$ *and* ${}^{\otimes}N^2_n \geq \frac{1}{48}F^2_n$ *for* $n \geq 4$.

Proof. The first inequality follows from the lemma **??** since every formula which projection is an nontautology must be an nontautology and from theorem **??** which shows lower and upper bounds for the fraction $\frac{N_n^1}{F_n^1}$. So for every one letter nontautology τ of the length n there are 2^n two letters nontautologies which projections are given type τ. Therefore ${}^{\downarrow}N_n^2 = 2^n N_n^1$ so ${}^{\downarrow}N_n^2 = 2^n N_n^1 \geq 2^n \frac{27}{100} C_n = \frac{27}{100} F_n^2$ (see theorem **??**) Notice that the last inequality holds for $n \geq 250$.

The second inequality follows just from the definition of the class described by the number ${}^{\otimes}N_n^2$ lemma **??** ${}^{\otimes}N_n^2 \geq 4T_{n-2}^2$ and lemma **??**). 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 lemma **??**) and also $\frac{T_{n-2}^2}{F_{n-2}^2} \geq \frac{1}{3}$ for $n-2 \geq 2$ (see lemma **??**).

Theorem 4.7 *There is less two letters tautologies than single letter tautologies and the fraction of two letter tautologies is positive i.e*

$$\frac{1}{3} < \liminf_{n\to\infty} \frac{T_n^2}{F_n^2} \leq \limsup_{n\to\infty} \frac{T_n^2}{F_n^2} < \liminf_{n\to\infty} \frac{T_n^1}{F_n^1}$$

Proof. $N_n^2 = {}^{\downarrow}N_n^2 + {}^{\otimes}N_n^2$. But from lemma **??** we get $N_n^2 \geq \frac{27}{100}F_n^2 + \frac{1}{16}F_n^2 > \frac{29}{100}F_n^2$ for $n \geq 250$. Therefore $T_n^2 < \frac{71}{100}F_n^2$ for $n \geq 250$. The lover bound of $\frac{T_n^2}{F_n^2}$ is greater then 0 and can be obtained from **??** so altogether

$$\frac{1}{3}F_n^2 < T_n^2 < \frac{71}{100}F_n^2$$

This should be contrasted with theorem **??**

$$\frac{72}{100}F_n^1 \leq T_n^1 \leq \frac{73}{100}F_n^1$$

for $n \geq 250$ from which we can complete the proof.

5 Types with k variables

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

Definition 5.1 *By F_n^k we mean the the number of types on length n and build up with k variables. $F_n^k = k^n C_n$ where C_n is the $n-th$ Catalan number. Formula for F_n^k is correct since there is k^n sequences of length n of ground types and each sequence can be turn into the tape by C_n ways of putting brackets. By T_n^k and respectively N_n^k we mean the numbers of tautologies and nontautologies. Of course $T_n^k + N_n^k = F_n^k$ and*

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

since

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

Definition 5.2 *By a simple tautology we mean a type $\tau = \tau_1, ..., \tau_n \to a$ for some ground type a, such that there is at least one component τ_i identical with a. Evidently 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 the length n build up with k ground type variables.*

Lemma 5.3 *The number G_n^k of simple tautologies is given by a recursive (on n) formula:*

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

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

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

Using formula above we can easily calculate numbers G_n^k. Below you can find G_n^k for $n = 1, 2, 3, 4$ contrasted with numbers F_n^k of all formulas with k variables.

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

Lemma 5.4 *For every $n \geq 2$ and every $k \geq 2$*

$$\frac{2}{3k} F_n^k \leq G_n^k \leq \frac{1}{k} F_n^k$$

Proof. (lower bound) k is fixed, induction on n. For $n = 2$ it is obvious (see the table above). Suppose the inequality holds for $n = 2, 3, ..., n-1$. So

$$\begin{array}{ll}
G_n^k = \sum_{i=2}^{n-1} F_{n-i}^k G_i^k + (F_{n-1}^k - G_{n-1}^k) = & \\
\sum_{i=2}^{n-2} F_{n-i}^k G_i^k + k G_{n-1}^k + (F_{n-1}^k - G_{n-1}^k) = & \\
\sum_{i=2}^{n-2} F_{n-i}^k G_i^k + F_{n-1}^k + (k-1) G_{n-1}^k \geq & \textit{induction} \\
\frac{2}{3k} \sum_{i=2}^{n-2} F_{n-i}^k F_i^k + F_{n-1}^k + (k-1) \frac{2}{3k} F_{n-1}^k = & \\
\frac{2}{3k} \sum_{i=1}^{n-1} F_{n-i}^k F_i^k - 2 \frac{2}{3k} F_1^k F_{n-1}^k + F_{n-1}^k + \frac{2(k-1)}{3k} F_{n-1}^k = & \textit{definition ??} \\
\frac{2}{3k} F_n^k - \frac{4}{3k} k F_{n-1}^k + F_{n-1}^k + \frac{2(k-1)}{3k} F_{n-1}^k = & \\
\frac{2}{3k} F_n^k - \frac{4}{3} F_{n-1}^k + F_{n-1}^k + \frac{2(k-1)}{3k} F_{n-1}^k = & \\
\frac{2}{3k} F_n^k + F_{n-1}^k [\frac{2(k-1)}{3k} - \frac{1}{3}] \geq \frac{2}{3k} F_n^k &
\end{array}$$

since $\frac{2(k-1)}{3k} - \frac{1}{3} \geq 0$ which is equivalent to $k \geq 2$.

(upper bound) k is fixed, induction on n.

$$\begin{array}{ll}
G_n^k = \sum_{i=2}^{n-1} F_{n-i}^k G_i^k + (F_{n-1}^k - G_{n-1}^k) \leq & \textit{induction} \\
\frac{1}{k} \sum_{i=2}^{n-1} F_{n-i}^k F_i^k + (F_{n-1}^k - G_{n-1}^k) & \\
\frac{1}{k} \sum_{i=1}^{n-1} F_{n-i}^k F_i^k - \frac{1}{k} F_1^k F_{n-1}^k + (F_{n-1}^k - G_{n-1}^k) = & \textit{definition ??} \\
\frac{1}{k} F_n^k - G_{n-1}^k \leq \frac{1}{k} F_n^k &
\end{array}$$

Theorem 5.5 *For every $k \geq 1$ the asymptotic probability of finding inhabited type in the set of all types build up with k variables is positive i.e.* $\liminf_{n\to\infty} \frac{T_n^k}{F_n^k} > 0$

Proof. For $k = 1$ the result was done at the chapter 3. For $k > 1$ the set of simple tautologies is big enough to show the result. $\frac{T_n^k}{F_n^k} > \frac{G_n^k}{F_n^k} \geq \frac{2}{3k}$ so from lemma **??** follows $\liminf_{n\to\infty} \frac{T_n^k}{F_n^k} \geq \frac{2}{3k}$.

Definition 5.6 *By $J_n^k(p)$ we mean the number of formulas of the length n build up with k ground type variables and having p premises i.e. formulas which are on the form: $\tau = \tau_1 \to ... \to \tau_p \to a$ where a is a ground type. Since numbers $J_n^k(p)$ described disjoined sets of formulas for different p and since there is no types of length n having more then $n-1$ premises for $n \geq 2$ we have:*

$$F_n^k = J_n^k(1) + ... + J_n^k(n-1)$$

Lemma 5.7 *The number $J_n^k(p)$ is given by recursion (on p) :*

$$J_n^k(1) = k F_{n-1}^k$$

$$J_n^k(p+1) = \sum_{i=1}^{n-p-1} F_i^k J_{n-i}^k(p)$$

Proof. It is obvious that $J^k_n(1) = kF^k_{n-1}$ is a number of types of the form $\tau \to a$ since it must be the number of types of the length $n-1$ times the number k of ground types. The formula $J^k_n(p+1) = \sum_{i=1}^{n-p-1} F^k_i J^k_{n-i}(p)$ is correct since any type $\tau \to \mu$ have to have $p+1$ premises if μ has p premises. So for every type τ of the length i and every type μ with p premises of the length $n-i$ we get one type $\tau \to \mu$ of the length n with p premises. The summation is only till $n-p-1$ since starting from $n-p$ all components $F^k_{n-p}J^k_p(p) = F^k_{n-p+1}J^k_{p-1}(p) = \ldots = 0$ because $J^k_p(p) = J^k_{p-1}(p) = \ldots = 0$.

Example 5.8 *It is easy to calculate formulas for $J^k_n(p)$ and consequently find the fraction $\frac{J^k_n(p)}{F^k_n}$ and it's asymptotic probability i.e.* $\lim_{n\to\infty} \frac{J^k_n(p)}{F^k_n}$. *For example $J^k_n(2) = kF^k_{n-1}$ so $\frac{J^k_n(2)}{F^k_n} = \frac{kF^k_{n-1}}{F^k_n} = \frac{kk^{n-1}C_{n-1}}{k^nC_n} = \frac{C_{n-1}}{C_n}$ Therefore*

$$\lim_{n\to\infty} \frac{J^k_n(2)}{F^k_n} = \lim_{n\to\infty} \frac{C_{n-1}}{C_n} = \frac{1}{4}$$

This means that statistically 25% of types are having 2 premises. For the first few $p = 1,2,3,4,5,6$ the number of formulas with p premises are given in the following table:

$J^k_n(p)$			$\lim_{n\to\infty} \frac{J^k_n(p)}{F^k_n}$
$J^k_n(1)$	$=$	kF^k_{n-1}	25.00%
$J^k_n(2)$	$=$	kF^k_{n-1}	25.00%
$J^k_n(3)$	$=$	$kF^k_{n-1} - k^2F^k_{n-2}$	18.75%
$J^k_n(4)$	$=$	$kF^k_{n-1} - 2k^2F^k_{n-2} -$	
		$k^3F^k_{n-3} + k^4F^k_{n-4}$	11.33%
$J^k_n(5)$	$=$	$kF^k_{n-1} - 3k^2F^k_{n-2}$	
		$+2k^4F^k_{n-4}$	7.03%
$J^k_n(6)$	$=$	$kF^k_{n-1} - 4k^2F^k_{n-2} +$	
		$2k^3F^k_{n-3} + 3k^4F^k_{n-4} - k^5F^k_{n-5}$	4.19%
.		.	.
.		.	.
.		.	.

From the table above for example we can see that there are 91.3 % of types with 6 or less premises and 50 % of types with 1 or 2 premises.

Lemma 5.9 *For every k and for every $n \geq 5$*

$$J^k_n(4) \leq \frac{9}{10} J^k_n(3)$$

Proof. We can obtain this inequality from the form of $J^k_n(4)$ and $J^k_n(3)$ which are in lemma **??**. The equivalent inequalities are the following:

$$
\begin{array}{rcl}
J_n^k(4) & < & \frac{9}{10} J_n^k(3) \\
kF_{n-1}^k - 2k^2 F_{n-2}^k - k^3 F_{n-3}^k + k^4 F_{n-4}^k & < & \frac{9}{10}(kF_{n-1}^k - k^2 F_{n-2}^k) \\
C_{n-1} - 2C_{n-2} - C_{n-3} + C_{n-4} & < & \frac{9}{10}(C_{n-1} - C_{n-2}) \\
10C_{n-1} - 20C_{n-2} - 10C_{n-3} + 10C_{n-4} & < & 9C_{n-1} - 9C_{n-2} \\
C_{n-1} + 10C_{n-4} & < & 11C_{n-2} + 10C_{n-3} \\
1 + 10\frac{C_{n-4}}{C_{n-1}} & < & 11\frac{C_{n-2}}{C_{n-1}} + 10\frac{C_{n-3}}{C_{n-1}}
\end{array}
$$

The last inequality holds since: $1 + 10\frac{C_{n-4}}{C_{n-1}} < 1 + 10\frac{C_1}{C_4} = 3 < \frac{54}{16} < 11\frac{1}{4} + 10\frac{1}{16} < 11\frac{C_{n-2}}{C_{n-1}} + 10\frac{C_{n-3}}{C_{n-1}}$. Remember that both functions on n $\frac{C_{n-2}}{C_{n-1}}$ and $\frac{C_{n-3}}{C_{n-1}}$ are decreasing, and limits are $\frac{1}{4}$ and $\frac{1}{16}$ respectively . Inspect lemma **??** for details and definition **??** for Catalan numbers C_1 and C_4.

Lemma 5.10 *For every k, $p \geq 4$ and $n \geq p+1$*

$$J_n^k(p) \leq \frac{9}{10} J_n^k(p-1)$$

Proof. Induction on p. The first step for $p = 4$ was done in lemma **??**. Suppose $n \geq p+1$. We have $n-p-1$ inductive assumptions for all j $n-1 \leq j \leq p+1$, namely

$$J_j^k(p) \leq \frac{9}{10} J_j^k(p-1)$$

So

$$
\begin{array}{l}
J_n^k(p+1) = \sum_{i=1}^{n-p-1} F_i^k J_{n-i}^k(p) \leq \\
\frac{9}{10} \sum_{i=1}^{n-p-1} F_i^k J_{n-i}^k(p-1) = \\
\frac{9}{10} \sum_{i=1}^{n-p-1} F_i^k J_{n-i}^k(p-1) + \frac{9}{10} F_{n-p}^k J_p^k(p-1) - \frac{9}{10} F_{n-p}^k J_p^k(p-1) = \\
\frac{9}{10} \sum_{i=1}^{n-p} F_i^k J_{n-i}^k(p-1) - \frac{9}{10} F_{n-p}^k J_p^k(p-1) = \\
\frac{9}{10} J_n^k(p) - \frac{9}{10} F_{n-p}^k J_p^k(p-1) \leq \frac{9}{10} J_n^k(p)
\end{array}
$$

Lemma 5.11 *For every $\varepsilon > 0$ there exists p such that for every k and for every sufficiently large n:*

$$J_n^k(p) \leq \varepsilon F_n^k$$

Proof. Let us take any $p \geq 4$ such that $(\frac{9}{10})^{p-3} < \varepsilon$. We are going to prove that for every $n \geq p+1$ and every k our inequality holds. By iteration of inequality from **??** we have:

$$J_n^k(p) \leq (\frac{9}{10})^{p-3} J_n^k(3).$$

By construction and using the formula for $J_n^k(3)$ (in example **??**) we get

$$J_n^k(p) \leq \varepsilon k^n (C_{n-1} - C_{n-2})$$

So finally:

$$\frac{J_j^k(p)}{F_n^k} \leq \varepsilon \frac{k^n(C_{n-1} - C_{n-2})}{k^n C_n} < \varepsilon$$

since $\frac{C_{n-1}}{C_n} - \frac{C_{n-2}}{C_n} < 1$ for $n \geq 4$. This complete the proof.

Lemma 5.12 *For every $\varepsilon > 0$ there exists p such that for every k and for every sufficiently large n:*

$$J_n^k(p) + ... + J_n^k(n-1) \leq \varepsilon F_n^k$$

Proof. From lemma **??** we can take any $p \geq 4$ such that $J_n^k(p) < \frac{\varepsilon}{10} F_n^k$ for any $n \geq p+1$ and for every k. By iteration of lemma **??** for any $s \geq p$ we obtain

$$J_n^k(s) \leq \left(\frac{9}{10}\right)^{s-p} J_n^k(p)$$

So we obtain

$$\begin{array}{l} J_n^k(p) + J_n^k(p+1) + ... + J_n^k(n-1) \leq \\ J_n^k(p) + \left(\frac{9}{10}\right)^1 J_n^k(p) + ... + \left(\frac{9}{10}\right)^{n-1-p} J_n^k(p) = \\ J_n^k(p) \left[1 + \left(\frac{9}{10}\right)^1 + ... + \left(\frac{9}{10}\right)^{n-1-p}\right] < \\ J_n^k(p) \left(\frac{1}{1-\frac{9}{10}}\right) = 10 J_n^k(p) < 10 \frac{\varepsilon}{10} F_n^k = \varepsilon F_n^k \end{array}$$

Lemma 5.13 *For every $\varepsilon > 0$ there exists p such that*

$$(1-\varepsilon) F_n^k < J_n^k(1) + ... + J_n^k(p-1)$$

for every k and every sufficiently large n.

Proof. This is just complementary to lemma **??** where p is a same as in lemma **??** and $n \geq p+1$.

Definition 5.14 *By $''a - type''$, we mean any type of the form $\tau_1, ..., \tau_n \to a$ where a is a ground type. The type a is also $''a - type''$ with $n = 0$.*

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

It is quite evident that a simple nontautology is neither an intuitionistic nor classical tautology. Suppose the type $\tau_1, ..., \tau_n \to a$ is a simple nontautology. Thus all τ_i are $''a_i - types''$ where $a_i \neq a$ for all $i \leq n$. Just put the $0,1$ valuation to τ which evaluate all a_i to 1 and a to 0. Therefore τ is not classically valid so by lemma **??** it is not a tautology.

Lemma 5.15 *The number of simple nontautologies of the length n build up with k ground type variables with p premises is exactly*

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

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

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

in which $\bigcirc$ 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 up the pattern in the way to obtain a simple untatutology. So altogether there is

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

of simple nontautologies with p premises.

Lemma 5.16 *The total number of simple nontautologies of the length n build up with k ground type variables is exactly*

$$\left(\frac{k-1}{k}\right)^1 J_n^k(1) + \left(\frac{k-1}{k}\right)^2 J_n^k(2) + ... + \left(\frac{k-1}{k}\right)^{n-1} J_n^k(n-1)$$

Proof. The formula for all simple nontautologies is obtain just by adding all simple nontautologies for all possible p since all classes of simple nontautologies described by the numbers $J_n^k(p)$ are disjoin.

Theorem 5.17 *For every $0 < \varepsilon < 1$ there exists k_0 such that for all $k \geq k_0$ and every sufficiently large n:*

$$N_n^k > (1-\varepsilon) F_n^k$$

Proof. Let us first define $\delta = 1 - \sqrt{1-\varepsilon}$. Such δ is a positive number much smaller then ε satisfying an equation $(1-\delta)^2 = 1-\varepsilon$. Now according to lemma **??** let us find a number p for δ such that

$$(1-\delta) F_n^k < J_n^k(1) + ... + J_n^k(p-1)$$

for all sufficiently large n. Now having p we can find such k_0 that $\left(\frac{k_0-1}{k_0}\right)^{p-1} > 1-\delta$. Since the fraction $\left(\frac{k_0-1}{k_0}\right) < 1$ then the same inequality holds for all numbers smaller then $p-1$, so we have:

$$\begin{array}{c}
\left(\frac{k_0-1}{k_0}\right)^1 > 1-\delta \\
\left(\frac{k_0-1}{k_0}\right)^2 > 1-\delta \\
\vdots \\
\left(\frac{k_0-1}{k_0}\right)^{p-1} > 1-\delta
\end{array}$$

Of course the same $p-1$ inequalities hold for all numbers $k > k_0$. Now we are ready for final estimation of nontautologies. Let us take the number of ground types k such that $k > k_0$ and let us take sufficiently large $n > p$ so:

$$
\begin{aligned}
&N_n^k \geq \left(\tfrac{k-1}{k}\right)^1 J_n^k(1) + \left(\tfrac{k-1}{k}\right)^2 J_n^k(2) + \ldots + \left(\tfrac{k-1}{k}\right)^{n-1} J_n^k(n-1) > \\
&\left(\tfrac{k-1}{k}\right)^1 J_n^k(1) + \left(\tfrac{k-1}{k}\right)^2 J_n^k(2) + \ldots + \left(\tfrac{k-1}{k}\right)^{p-1} J_n^k(p-1) > \\
&(1-\delta) J_n^k(1) + (1-\delta) J_n^k(2) + \ldots + (1-\delta) J_n^k(p-1) = \\
&(1-\delta)\left[J_n^k(1) + J_n^k(2) + \ldots + J_n^k(p-1)\right] > \\
&(1-\delta)(1-\delta) F_n^k = (1-\varepsilon) F_n^k
\end{aligned}
$$

Theorem 5.18 *For every $\varepsilon > 0$ there exists k_0 such that for $k \geq k_0$ and for every sufficiently large n:*

$$T_n^k < \varepsilon F_n^k$$

Proof. Since $T_n^k + N_n^k = F_n^k$, this is just a dual form of theorem **??**.

Theorem 5.19 *Let $g_k = \liminf_{n\to\infty} \frac{T_n^k}{F_n^k}$. For every number of ground variables k*

$$g_k > 0$$

$$\lim_{k\to\infty} g_k = 0$$

Proof. The first statement is guaranteed by theorem **??** i.e. $\liminf_{n\to\infty} \frac{T_n^k}{F_n^k} > \frac{2}{3k}$ for $k \geq 1$. The second one follows from theorem **??** since for every $\varepsilon > 0$ there exists k_0 starting from which the fraction of tautologies

$$\frac{T_n^k}{F_n^k} < \varepsilon$$

for all sufficiently large n. This just means that the limit of g_k is zero.

Example 5.20 *This is a list of best known to the author lower and upper bounds for the fraction of inhabited types (tautologies) in the set of all types build up with k variables for various k. Notice that as a result of chapter 4 we know that $g_2 < g_1$. It is not known to the author whether $g_{i+1} < g_i$ for all i however we strongly believe it is so.*

k		lower	upper	source
1		72.0%	73.0%	theorem **??**
2	$\frac{1}{3} \approx$	33.3%	71.0%	lemma **??** and theorem **??**
3	$\frac{2}{9} \approx$	22.2%	64.4%	lemma **??** lemma **??** for p=6
4	$\frac{1}{6} \approx$	16.6%	55.7%	lemma **??** lemma **??** for p=6

Acknowledgements
The authors wish to thank Paweł Idziak for lot of valuable discussions on the subject.

References

[1] H.P. Barendregt. The Lambda Calculus, Its Syntax and Semantics. *North-Holland*

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

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

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

6 SUPLEMENT

This is a copy of *Mathematica* program and session which proves by induction theorem ??. Variables and tables used in the program have the following meaning:

`po`	is a Newton symbol function
`Ca`	is a Catalan number function
`k`	is a number of induction assumptions needed in the current session
`n`	is a number induction starts
`p,q`	lower and upper bounds for which the proof is constructed
`Cata`	is a table filed up with 2*k+2 Catalan numbers
`Tau`	is a table filled up with numbers of tautologies
`Ntau`	is a table filled up with numbers of nontautologies
`Wspdol`	is a table for coefficients for lower bound
`Wspgora`	is a table for coefficients for upper bound
`ddol`	is a calculated lower bound

```
po[n_,k_]:=n!/((k!)*(n-k)!)
Ca[n_]:=po[2n-2,n-1]/n
nn:=1;

k:=250;
p:=72/100;
q:=73/100;
dl:=2*k+2;
n:=2*k;

Cata=Table[0,{i,1,dl}];
Ntau=Table[0,{i,1,dl}];
Tau=Table[0,{i,1,dl}];
Wspdol=Table[0,{i,1,k}];
Wspgora=Table[0,{i,1,k}];

Cata[[1]]=1; Cata[[2]]=1;
Ntau[[1]]=1; Ntau[[2]]=0;
Tau[[1]]=0; Tau[[2]]=1;
```

```
Do[Cata[[nn]]=Ca[nn];
Ntau[[nn]]=Sum[Ntau[[i]]*Tau[[nn-i]], {i,1,nn-1}];
Tau[[nn]]=Cata[[nn]]-Ntau[[nn]];,
{nn,3,dl,1}];

Do[Wspdol[[i]]=(p*Ntau[[i]]+
(1-q)*Tau[[i]]-2*p*(1-q)*Cata[[i]]);,
{i,1,k,1}];

Do[Wspgora[[i]]=q*Ntau[[i]]+(1-p)*Tau[[i]]-
2*q*(1-p)*Cata[[i]];,
{i,1,k,1}];


(*  N[Wspdol, 10]
N[Wspgora, 10]  *)

ddol:= p*(1-q)+
Sum[ If[Wspdol[[i]]>=0, ((1/4)^i)*Wspdol[[i]],
(Cata[[2*k-i]]/Cata[[2*k]])*Wspdol[[i]] ]
,{i,1,k,1}];

ggora:= q*(1-p)+
Sum[ If[ Wspgora[[i]]<=0, ((1/4)^i)*Wspgora[[i]],
(Cata[[2k-i]]/Cata[[2*k]])*Wspgora[[i]] ]
,{i,1,k,1}];

N[q,10]
N[p,10]
N[1-ddol, 10]
N[1-ggora, 10]
:[font = output; output; inactive; preserveAspect]
0.73
;[o]
0.73
:[font = output; output; inactive; preserveAspect]
0.72
;[o]
0.72
:[font = output; output; inactive; preserveAspect]
0.7260947441029422
;[o]
0.7260947441
:[font = output; output; inactive; preserveAspect; endGroup]
0.7205865315477873
;[o]
0.7205865315
```