

Some properties of random λ -terms*

René David, Christophe Raffalli, Guillaume Theyssier[†]
(Université de Savoie)

Katarzyna Grygiel, Jakub Kozik, Marek Zaionc[‡]
(Jagiellonian University)

October 2, 2009

Abstract

We present quantitative analysis of various (syntactic and behavioral) properties of random λ -terms. Our main results are that asymptotically all the terms are strongly normalizing and that any fixed closed term almost never appears in a random term. Surprisingly, in combinatory logic (the translation of the λ -calculus into combinators) the result is exactly opposite. We show that almost all terms are *not* strongly normalizing. This due to the fact that any fixed combinator almost always appears in a random combinator.

Keywords: λ -calculus, strong normalization, randomness, combinatory logic.

1 Introduction

Since the pioneering works of Church, Turing *et al.*, more than 70 years ago, a wide range of computational models have been introduced. It turns out that they are all equivalent in what they *can* compute. However, this equivalence says nothing about what do *typical* programs or machines of each of these models.

This paper addresses the following question. Having a (theoretical) programming language and a property, what is the probability that a random program satisfies the given property? In particular, is it true that almost every random program satisfies the desired property.

We concentrate on functional programming languages and, more specifically, on the λ -calculus, the simplest such language (see [9, 11, 14] for similar works on other models of computation). The only work that we have found on this subject is some experiments made by Jue Wang (see [16]). Most interesting properties of terms are those concerning their behavior. However, to analyze them, one has to consider some syntactic properties as well.

As far as we know, no asymptotic value for the number of λ -terms of size n is known. We give (see Section 6) upper and lower bounds for this (super-exponential) number. Although the gap between the lower and the upper bound is big (exponential), these estimations are sufficient for our purpose.

*This work was supported by the research project funded by the French Rhône-Alpes region and initiated by Pierre Lescanne and by grant number N206 376137 funded by Polish Ministry of Science and Higher Education

[†]Laboratoire de Mathématiques de l'Université de Savoie, Campus Scientifique, 73376 Le Bourget-du-Lac, France, email: {rene.david, christophe.raffalli, guillaume.theyssier}@univ-savoie.fr

[‡]Theoretical Computer Science, Jagiellonian University, Łojasiewicza 6, Kraków, Poland, email: {Katarzyna.Grygiel, Jakub.Kozik, zaionc}@tcs.uj.edu.pl

We prove several results on the structural form of a λ -term. In particular, we show that almost every closed λ -term begins with “many” λ 's (the precise meaning is given in Theorem 14). Moreover, each of them binds “many” occurrences of variables (Theorems 15, 16 and 17). Finally, given any fixed closed λ -term, almost no λ -term has this term as a sub-term (Theorem 21).

We also give a result on the behavior of terms, our original motivation. We show that a random term is strongly normalizing (*SN* for short) with probability 1. Remember, that, in general, being *SN* is an undecidable question.

Combinatory logic is another programming language related to the λ -calculus. It can be seen as an encoding of λ -calculus into a language without variable binding. Moreover, there are translations, in both directions, which, for example, preserve the property of being *SN*. Surprisingly, our results concerning random combinators are very different from those for the λ -calculus. For example we show that, for every fixed term t_0 , almost every term has t_0 as sub-term and this, of course, implies that almost every term is not *SN*. The difference of results concerning strong normalization between λ -calculus and combinatory logic might come from the large increase of size induced by the coding of bound variables in combinatory logic. This is discussed in Section 9.

Our interest in statistical properties of computational objects like lambda terms or combinators is a natural extension on similar research on logical objects like formulas or proofs. This paper is a continuation of the research in which we try to estimate the properties of random formulas in various logics. Especially the probability of truth (or satisfiability) for random formulas. For the purely implicational logic with one variable, (and at the same time simple type systems) the exact value of the density of true formulas have been computed in the paper of Moczurad, Tyszkiewicz and Zaionc [13] and [15]. Quantitative relationship between intuitionistic and classical logics (based on the same language) has also been analyzed. The exact value describing how big fragment of the classical logic with one variable is intuitionistic has been determined in Kostrzycka and Zaionc [10]. For the results with more than one variable, and other logical connectives consult [4],[8],[7].

The case of and/or connectors received much attention – see Lefmann and Savický [12], Chauvin, Flajolet, Gardy and Gittenberger [2] and Gardy and Woods [6]. We refer to Gardy [5] for a survey on probability distribution on Boolean functions induced by random Boolean expressions.

2 Organization of the paper

In Sections 3 and 4 we recall basic definitions of the λ -calculus and introduce the notation used within the paper. Section 5 summarizes the basic combinatorial facts that are useful in our development. Starting from Section 6 we present our results for lambda calculus.

Section 8 contains results in combinatory logic, namely that every fixed term appears in almost every term. Section 9 discusses the question of size, gives experimental results for questions for which we have no answers. It also gives open questions and proposes future direction of research.

In the whole paper we do not aim at providing the best possible estimations for the analyzed sequences. Most of the quantitative results can be easily improved. We present estimations which are sufficient for our structural results, without sacrificing the simplicity of proofs for better accuracy.

3 Generality on the λ -calculus

Definition 1. *The set of λ -terms (or simply terms) is defined by the following grammar (where V is a countable set of variables)*

$$t, u := V \mid \lambda V.t \mid (t u)$$

We denote by Λ the set of all closed λ -terms.

As usual, λ -terms are considered modulo the α -equivalence i.e. two terms which differ only by the names of bound variables are considered equal. Note that λ -terms can be seen as trees. For every term t , if we forget about variable binding we obtain a unary-binary tree. We call it *the structure* of t . Removing from the structure the unary nodes and connecting binary ones and leaves so that to preserve the original connectivity, we obtain *binary structure* of t .

We often use (without giving the precise definition) the classical terminology about trees (e.g. path, root, leaf, etc.). The paths from the root to leaves are called branches.

Definition 2. 1. *t' is a sub-term of t (denoted as $t' \leq t$) if*

- $t = t'$,
- or $t = \lambda x.u$ and $t' \leq u$,
- or $t = (u v)$ and ($t' \leq u$ or $t' \leq v$).

2. *Let t be a term and $u = \lambda x.a$ be a sub-term of t . We say that λx is binding if x has a free occurrence in a .*

3. *The unary height of a term t is the maximal number of λ 's on a path from the root to some leaf of t .*

4. *Let t be λ -term.*

- *Two λ 's in t are called incomparable if there is no branch containing both of them.*
- *The λ -width of t is the maximal number of pairwise incomparable **binding** λ 's.*

5. *We say that a term has k head lambdas if its structure starts with k unary nodes followed by a binary node or a leaf.*

Definition 3. *The size of a term t (denoted as $size(t)$) is defined by the following rules.*

- $size(x) = 0$ if x is a variable,
- $size(\lambda x.t) = size(t) + 1$,
- $size((t u)) = size(t) + size(u) + 1$.

Definition 4. *Let n be an integer. We denote by Λ_n the set of closed terms of size n . Obviously the set Λ_n is finite. We denote by L_n its cardinality.*

As far as we know, no asymptotic analysis of the sequence L_n has been made. Moreover, typical combinatorial techniques does not seem to apply easily for this task.

4 Main notations

To attribute a precise meaning to sentences like "asymptotically all lambda terms have property P " we use the following definition of asymptotic density. For any finite set A we denote by $\#A$ the number of its elements.

Definition 5. For a set of lambda terms A , we denote by $d(A)$ the following limit (if it exists):

$$\lim_{n \rightarrow \infty} \frac{\#(A \cap \Lambda_n)}{\#\Lambda_n}.$$

If the limit exists it is called asymptotic density of A .

Note that d is not a measure (e.g. it is not countably additive).

Let P be a property of lambda terms. If $d(\{t \in \Lambda \mid P(t) \text{ holds}\}) = \alpha$, we say that the density of terms satisfying P is α . By an analogy to researches on graphs and trees, we abbreviate the sentence like "the density of terms satisfying P is 1" by "random term satisfies P ".

The idea of proofs

In Section 6, a density 0 is proved by computing bounds for the cardinalities of the sets we consider, and showing that the quotient tends to zero. The computations are quite standard. The computations have been checked by Maple. The corresponding file, together with a pdf of it, can be found at the URL:

www.lama.univ-savoie.fr/~david/ftp/limit

In Section 7 we show that a set A of terms has density 0 by defining an injective, size preserving function φ from A into Λ (we call such functions codings). Then we show that the image of φ has density 0. This is done either by using the fact that it is included in a set, which is already known to have density 0, or by computing an upper bound for the cardinality of this image.

The proofs concerning densities in the calculus of combinators are based on analysis of generating functions enumerating considered sets of combinators.

Note about the statement of the theorems

1. Many of the following sub-sections use results of the previous ones. When, in some section, we say "let t be a typical term", this implicitly mean that we restrict ourselves to terms having the properties for which we have seen, in the previous sub-sections, that it has density 1. We also assume that its size is big enough.
2. The statement of the theorems sometimes requires to give a name to the size of terms. This size is always denoted by n . Thus a statement "the density of terms satisfying $P(t, n)$ is α " means that

$$\lim_{n \rightarrow \infty} \frac{\#\{t \in \Lambda_n \mid P(t, n)\}}{\#\Lambda_n} = \alpha.$$

5 Combinatorial results

5.1 Catalan numbers

We denote by $C(n)$ the Catalan numbers i.e. the number of binary trees with n inner nodes. We use the following proposition.

Proposition 6. $C(n) \sim \frac{4^n}{n^{3/2}\sqrt{\pi}}$ and thus, for large enough n , we have $C(n) \geq C \frac{4^n}{n^{3/2}}$ for some constant $C > 0$.

Proof. This is a classical result. See for example [3]. □

5.2 Large Schröder numbers

We denote by $M(n, k)$ the number of unary-binary trees with n inner nodes and k leaves. Let $M(n) = \sum_{k \geq 1} M(n, k)$ denote the number of unary-binary trees with n inner nodes. These numbers are known as the large Schröder numbers. Note that, since in this paper the size of variable is 0, we use them instead of the so-called Motzkin numbers, which enumerate unary-binary trees with n total nodes. We use the following proposition.

Proposition 7. 1. $M(n, k) = C(k-1) \binom{n+k-1}{n-k+1}$.

$$2. M(n) \sim \left(\frac{1}{3-2\sqrt{2}} \right)^n \frac{1}{\sqrt{\pi n^{3/2}}}.$$

Proof. (1) Every unary-binary tree with n inner nodes and k leaves has $k-1$ binary and $n-k+1$ unary nodes. We have $C(k-1)$ binary trees with k leaves. Every such a tree has $2 \cdot k - 1$ nodes (inner nodes and leaves). Therefore there are $\binom{n+k-1}{n-k+1}$ possibilities of inserting $n-k+1$ unary nodes (we can put unary node above every node of a binary tree). (2) The asymptotic for $M(n)$ is obtained by using standard tools of the generating function (for this sequence it is equal to $m(x) = \frac{1-x-\sqrt{1-6x+x^2}}{2x}$). For more details see [3]. \square

6 Proofs using calculus

6.1 Lower Bound for L_n

The asymptotic inequality $f(n) \gtrsim g(n)$ means that there exists h such that $h(n) \sim g(n)$ and $f(n) \geq h(n)$.

Theorem 8. For any $\varepsilon > 0$ we have

$$L_n \gtrsim \left(\frac{(4-\varepsilon)n}{\ln(n)} \right)^{n - \frac{n}{\ln(n)}}$$

Proof. Let $LB(n, k)$ be the number of λ -terms of size n with k head λ 's and no other λ below. Since the lower part of the term is a binary tree with $n-k$ inner nodes, and each leaf can be bound by k lambdas, we have $LB(n, k) = C(n-k)k^{n-k+1}$. Clearly $LB(n, k) < L_n$. We choose $k = \lceil \frac{n}{\ln(n)} \rceil$. Then we get:

$$L_n > C(n - \lceil \frac{n}{\ln(n)} \rceil) \cdot \left(\frac{n}{\ln(n)} \right)^{n - \frac{n}{\ln(n)} - 1} \gtrsim \left(\frac{4 \cdot n}{\ln(n)} \right)^{n - \frac{n}{\ln(n)}} \cdot \frac{1}{p(n)},$$

for some positive polynomial p (the last asymptotic inequality is a consequence of Proposition 6). It is easy to see that the last formula is asymptotically greater than $\left(\frac{(4-\varepsilon)n}{\ln(n)} \right)^{n - \frac{n}{\ln(n)}}$ \square

6.2 The number of λ 's in a term

Theorem 9. 1. The density of terms having more than $\frac{3n}{\ln(n)}$ λ 's is 0.

2. The density of terms having less than $\frac{n}{3\ln(n)}$ λ 's is 0.

Proof. (1) Let $S(n, k)$ be the number of terms of size n containing more than $\frac{kn}{\ln(n)}$ λ 's. We have $S(n, k) \leq \sum_{p \geq \frac{kn}{\ln(n)}} UB(n, p)$ where $UB(n, p) = M(n, n-p+1) \cdot p^{n-p+1}$. This is because a term with p lambdas is a unary binary tree whose

$n - p + 1$ leaves can be bound by, at most, p lambdas each. For $k > 1$ the function p^{n-p+1} is decreasing for $p \geq \frac{kn}{\ln(n)}$. Thus, for every $k > 1$, we have

$$\begin{aligned} S(n, k) &\leq \left(\frac{kn}{\ln(n)} \right)^{n+1-\frac{kn}{\ln(n)}} \cdot \sum_{p \geq \frac{kn}{\ln(n)}} M(n, p) \\ &\leq M(n) \left(\frac{kn}{\ln(n)} \right)^{n+1-\frac{kn}{\ln(n)}} \end{aligned}$$

Using our lower bound for L_n , we find $\frac{S(n, k)}{L_n} \leq \Phi(n, (4 - \varepsilon))$ where

$$\Phi(n, q) = \frac{M(n) \left(\frac{kn}{\ln(n)} \right)^{n+1-\frac{kn}{\ln(n)}}}{\left(\frac{qn}{\ln(n)} \right)^{n-\frac{qn}{\ln(n)}}}$$

To get the result it remains to show that, for $k = 3$ and any $\varepsilon > 0$, $\Phi(n, 4 - \varepsilon)$ tends to 0.

Using that $M(n) \sim \left(\frac{1}{3-2\sqrt{2}} \right)^n \frac{1}{n^{\frac{3}{2}}}$ we have, for n large enough, (we introduce an extra constant $C > 1$ to compensate for the equivalent)

$$\Phi(n, q) \leq C \cdot \frac{\left(\frac{1}{3-2\sqrt{2}} \right)^n \left(\frac{kn}{\ln(n)} \right)^{n+1-\frac{kn}{\ln(n)}}}{n^{\frac{3}{2}} \left(\frac{qn}{\ln(n)} \right)^{n-\frac{qn}{\ln(n)}}}$$

We get a simpler upper bound by using the $n^{\frac{3}{2}}$ to compensate for the $+1$ exponent:

$$\begin{aligned} \Phi(n, q) &\leq \frac{\left(\frac{1}{3-2\sqrt{2}} \right)^n \left(\frac{kn}{\ln(n)} \right)^{n-\frac{kn}{\ln(n)}}}{\left(\frac{qn}{\ln(n)} \right)^{n-\frac{qn}{\ln(n)}}} \\ &= \left(\frac{k}{q(3-2\sqrt{2})} \right)^n \left(\frac{kn}{\ln(n)} \right)^{\frac{-kn}{\ln(n)}} \left(\frac{qn}{\ln(n)} \right)^{\frac{qn}{\ln(n)}} \end{aligned}$$

Remarking that $\left(\frac{kn}{\ln(n)} \right)^{\frac{-kn}{\ln(n)}} = e^{-kn} \left(\frac{k}{\ln(n)} \right)^{\frac{-kn}{\ln(n)}}$ and $\left(\frac{qn}{\ln(n)} \right)^{\frac{qn}{\ln(n)}} = e^n \left(\frac{q}{\ln(n)} \right)^{\frac{qn}{\ln(n)}}$, we have:

$$\Phi(n, q) \leq \left(\frac{ke^{1-k}}{q(3-2\sqrt{2})} \right)^n \left(\frac{qk^{-k}}{\ln^2(n)} \right)^{\frac{n}{\ln(n)}}$$

This means that $\Phi(n, q)$ converges toward zero if $\frac{ke^{1-k}}{q(3-2\sqrt{2})} < 1$. Since, ke^{1-k} reaches its maximum 1 in $k = 1$ and $0 < q(3-2\sqrt{2}) < 4(3-2\sqrt{2}) < 1$ (recall that we will use $q = 4 - \varepsilon$ with $\varepsilon > 0$), the equation $ke^{1-k} = q(3-2\sqrt{2})$ has two solutions, one for $k > 1$ the other for $k < 1$. It is easy to see that the first solution is smaller than 3 because $3e^{1-3} < 3\frac{4}{25} < 4(3-2\sqrt{2})$ and $\varepsilon = 4 - q$ can be chosen small enough.

(2) The proof of the second part of the theorem is analogous. The computation is essentially the same with $k < 1$. It is easy to check that the solution (less than 1) of the equation $ke^{1-k} = q(3-2\sqrt{2})$ is less than $\frac{1}{3}$. \square

Remark 10. *The theorem above shows that the typical proportion of unary nodes over binary nodes in lambda terms is far from the typical proportion in ordinary unary-binary trees which tends a positive constant.*

6.3 Upper Bound for L_n

Theorem 11. *For all $\varepsilon > 0$ we have*

$$L_n \lesssim \left(\frac{(12 + \varepsilon)n}{\ln(n)} \right)^{n - \frac{n}{3 \ln(n)}}$$

Proof. Let N_n be the number of terms of size n with less than $\frac{3n}{\ln(n)}$ and more than $\frac{n}{3 \ln(n)}$ λ 's. Note that, according to Theorem 9 we have $L_n \sim N_n$. We have

$$N_n \lesssim C \left(n - \frac{n}{3 \ln(n)} + 1 \right) \binom{2n+1}{\frac{3n}{\ln(n)}} \left(\frac{3n}{\ln(n)} \right)^{n+1 - \frac{n}{3 \ln(n)}}$$

where

- $C \left(n - \frac{n}{3 \ln(n)} + 1 \right)$ corresponds to the number of possibilities of choice of binary structure (which has less than $n - \frac{n}{3 \ln(n)}$ inner nodes).
- $\binom{2n+1}{\frac{3n}{\ln(n)}}$ is an asymptotic upper bound the number of possible distributions of unary nodes within binary structure.
- $\left(\frac{3n}{\ln(n)} \right)^{n+1 - \frac{n}{3 \ln(n)}}$ corresponds to the possibilities of variable bindings. Indeed, $\frac{3n}{\ln(n)}$ is an upper bound for the number of lambdas above a variable and $n + 1 - \frac{n}{3 \ln(n)}$ is an upper bound for the number of leafs.

Now it is sufficient to observe that $\binom{2n+1}{\frac{3n}{\ln(n)}}$ is subexponential. (The replacement of 12^n by $(12 + \varepsilon)^n$ compensates all factors smaller than exponential.) □

6.4 Comparison between the lower bound and the upper bound

In the ratio between our lower and upper bounds, the dominant factor is exponential. This means that we are far from having an equivalent, but still this is not too bad because L_n is super-exponential.

The following corollary shows that we know the two first terms of the asymptotic expansion of $\ln(L_n)$, but we do not know the linear factor yet.

Corollary 12. *For all $\varepsilon > 0$ and for n large enough*

$$\ln(4 - \varepsilon) - 1 \leq \frac{\ln(L_n)}{n} - \ln(n) + \ln(\ln(n)) \leq \ln(12 + \varepsilon) - \frac{1}{3}$$

6.5 Bounds on the unary height of a term

Theorem 13. *The set of terms with the unary height greater than $\frac{n}{3 \ln(n)}$ has density 1.*

Proof. The same argument as in 9.(2) applies here. □

7 Proofs using coding

7.1 The number of λ 's in head position

Theorem 14. *Let $g(n) \in o(\sqrt{n/\ln(n)})$. The density of terms having less than $g(n)$ head λ 's is 0.*

Proof. Let us denote by A_n the set of typical terms of size n with less than $g(n)$ head λ 's. We construct an injective, size-preserving function (coding) $\varphi: A \rightarrow \Lambda$ such that its image has density 0.

Let $t \in A_n$. We can write $t = \lambda x_1 \dots \lambda x_p.M$, where $p < g(n)$ and M is a term starting with an application and containing at least one λ (by Theorem 13). Let B be the maximal purely applicative prefix of M i.e. B is the term using only application nodes and variables such that $M = B[\vec{t}]$ where terms in \vec{t} start with λ and variables in B are taken from the set $\{x_1, \dots, x_p\}$ (see Figure 1).

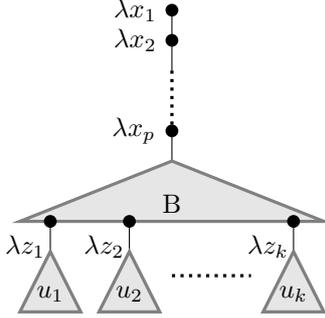


Figure 1: th. 14, the term $t \in A_n$.

Let us denote by $A(n, p, b, \vec{t})$ the set of terms from A_n having, as in the decomposition of t above, p head λ 's, then a purely applicative context (i.e. a context without any lambda) of size b , and, in that context, a sequence \vec{t} of subterms beginning with λ 's. Because $p < g(n)$ the cardinality of $A(n, p, b, \vec{t})$ is less than

$$P(b, n) = C(b+1)(g(n)+1)^{b+1}.$$

Let $t \in A(n, p, b, \vec{t})$ where $\vec{t} = [t_1, \dots, t_k]$. By hypothesis on A_n , we have $k \geq 1$. Let $t_i = \lambda z_i.u_i$. Let z be a fresh variable and $u'_i = u_i[z_i := z]$. Consider the term $T = \lambda z \lambda x_1 \dots \lambda x_p.(u'_1 (u'_2 (\dots (u'_{k-1} u'_k) \dots)))$ which is of size $n - b$. Let $\lambda y.C$ denote the term rooted at the leftmost deepest λ of term T and let Y be the set of variables introduced by the λ 's occurring on the path from the the root to λy . By Theorem 13 there are at least $\frac{n}{3 \ln(n)}$ elements in Y .

Let U be the set of purely applicative terms of size $b - 1$ whose variables are chosen from Y . For any $u \in U$, let $\rho(t, u)$ be the term obtained by substituting sub-term $\lambda y.C$ in T with $\lambda y.(u C)$.

There are at least

$$Q(b, n) = C(b-1) \left(\frac{n}{3 \ln(n)} \right)^b$$

elements in U . Since for n large enough we have $P(b, n) < Q(b, n)$ (because the limit of the quotient is 0), there exists an injective function h which assigns to any purely applicative prefix B of size b an element from U . Let $\varphi(t) = \rho(t, h(B))$ where B is the purely applicative prefix in the decomposition of t (see Figure 2). By the injectivity of h , we get that φ is injective, too.

We also define $\Psi(t) = \{\rho(t, u) : u \in U\}$. Note that for $t \in A(n, p, b, \vec{t})$ the cardinal of $\Psi(t)$ is always $Q(b, n)$. Due to the construction, the sets $\Psi(t)$ and $\Psi(t')$ are disjoint for any pair of distinct terms t and t' .

Let us denote by $\psi(b, n) = \frac{P(b, n)}{Q(b, n)}$. By the assumption on g there is a function ε such that $\varepsilon(n)$ tends to 0 and $\psi(b, n) = \frac{C(b+1)}{3^b C(b)} \left(\frac{n}{\ln(n)} \right)^{\frac{1-b}{2}} \varepsilon(n)$. For $b \geq 2$, $\left(\frac{n}{\ln(n)} \right)^{\frac{1-b}{2}}$

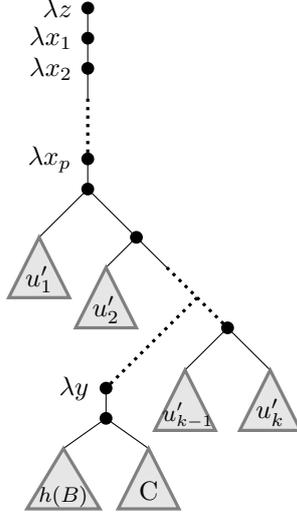


Figure 2: th. 14, the term $\varphi(t)$.

is decreasing in b , so $\frac{C(b+1)}{3^b C(b)}$ is bounded. Thus, $\psi(b, n)$ tends to 0 uniformly in b . Since the $A(n, p, b, \vec{t})$ form a partition of A_n , the result follows. \square

7.2 Head λ 's bind “many” occurrences

Theorem 15. *Let $g(n) \in o(\sqrt{n/\ln(n)})$. The density of terms in which there is at least one λ among $g(n)$ head λ 's that does not bind any variable is 0.*

Proof. Let $g(n) \in o(\sqrt{n/\ln(n)})$ and denote by T_v the set of random terms for which there exists at least one λ among first $g(n)$ head λ 's that does not bind any variable, and let $T_v^n = T_v \cap \Lambda_n$. We construct a coding function $\varphi: T_v \rightarrow \Lambda$ such that the density of its image is 0.

Let $T = \lambda x_1 \dots x_{g(n)}.A$ be a term from T_v^n and let i be the smallest integer such that the i -th head λ in T does not bind any variable. Take

$$\varphi(T) = \lambda x_1 \dots x_{i-1} x_{i+1}.(x_{i+1} (\lambda x_{i+2} \dots x_{g(n)}.A)).$$

The size of $\varphi(T)$ is n . Terms from the set $\varphi(T_v^n)$ have less than $g(n)$ head λ 's, so, by Theorem 14, the density of them in the set Λ_n is zero. Since the function φ is injective, the density of T_v^n is also zero. \square

Theorem 16. *Let $g(n) \in o(\ln(n))$. The density of terms in which the total number of occurrences of variables bound by the first three λ 's is at most $g(n)$ is 0.*

Proof. Let $g(n) \in o(\ln(n))$ and denote by $T_{n,g(n)}$ the set of random terms of size n in which the total number of occurrences of variables bound by first three λ 's is at most $g(n)$. We construct a coding functions $\varphi_n: T_{n,g(n)} \rightarrow \Lambda_n$ such that the density of the union of images of all functions in Λ is zero.

Let us define an equivalence relation \sim_n on the set of random terms of size n in the following way: $M \sim_n N$ iff M and N are equal after substituting all occurrences of variables bound by first three λ 's by the variable bound by the first λ . Let us denote by $[M]$ the equivalence class of M .

Let $T = \lambda x_1 \lambda x_2 \lambda x_3.A$ be a term from $T_{n,g(n)}$. There are at most $3^{g(n)}$ elements in the class $[T]$.

Let $T' = \lambda xy.A[x_1 := y, x_2 := y, x_3 := y]$. The size of T' is $n-1$. Let us consider $\lambda a.U$ the sub-term of T' such that λa is the leftmost deepest λ in T' . Denote by $\mathcal{B}(T)$ the set of variables bound by λ 's occurring in T' on the path from λa to λy . Note that the variable x does not occur neither in T' nor in $\mathcal{B}(T)$. By Theorem 13, there are at least $\frac{n}{3 \ln(n)} - 3$ such λ 's. Since $3 \leq \frac{n}{6 \ln(n)}$, there are at least $\frac{n}{6 \ln(n)}$ elements in $\mathcal{B}(T)$. As $g(n) \in o(\ln(n))$, we have

$$\lim_{n \rightarrow \infty} \frac{3^{g(n)}}{\left(\frac{n}{6 \ln(n)}\right)} = 0.$$

Thus, we can find for each class $[T]$ an injective function h_T from $[T]$ into the set $\mathcal{B}(T)$.

We define $\varphi(T)$ as the term obtained from T' by replacing the sub-term $\lambda a.U$ with $\lambda a.((y B) U)$, where $B = h_{[T]}(T)$.

All terms from the image $\varphi(T_{n,g(n)})$ start with a λ that binds no variable. By Theorem 15 we know that the set of such terms have density zero in Λ_n . Since f is injective, the density of $\bigcup_{n \in \mathbb{N}} T_{g(n)}$ is zero, as well. \square

Theorem 17. *For any fixed integers k and k' , the density of terms in which each of the first k λ 's binds more than k' variables is 1.*

Proof. Let us fix integers k, k' and let $g(n) = \sqrt{\ln(n)}$. We assume that $k \geq 3$. By Theorem 16, the total number of occurrences of variables bound by first k λ 's in a random term of size n is more than $g(n)$.

For each n and $q \geq g(n)$ let $A(n, q)$ be the set of typical terms of size n having exactly q leaves bound by the first k lambdas and let $B(n, q)$ be the set of terms in $A(n, q)$ for which one of the first k λ 's binds at most k' variables.

Consider the equivalence relation \sim_n defined analogously to the relation from the proof of Theorem 16, but with respect to the first k (instead of three) head λ 's. For $T \in A(n, q)$ the cardinality of $[T] \cap A(n, q)$ is k^q and the cardinality of $[T] \cap B(n, q)$ is at most $k \cdot k' \cdot q^{k'} \cdot (k-1)^{q-k'}$ and thus the quotient is less than $\psi(q) = \frac{k \cdot k' \cdot q^{k'} \cdot (k-1)^{q-k'}}{k^q}$ which, since ψ is eventually decreasing, is less than $\psi(g(n))$.

Since the $[T] \cap A(n, q)$ give a partition of $A(n, q)$ and the $A(n, q)$ give a partition of the set of typical terms of size n and since $\psi(g(n))$ has limit 0 when n tends to ∞ this finishes the proof. \square

7.3 The width of a term

Let us recall that lambda width of a term is the maximal number of incomparable binding lambdas in the term. In the following proposition we show that lambda width tends to be very low for typical lambda terms.

Theorem 18. *The density of terms having λ -width at most 2 is 1.*

Proof. Let us denote by W the set of terms with λ -width greater than 2. As usual we put $W_n = W \cap \Lambda_n$. We show that there exists an injective, size preserving function $\varphi: W \rightarrow \Lambda$ such that its image has density 0. Let t be an element of W_n and let us denote by $\lambda x, \lambda y$ and λz the three highest, pairwise incomparable binding λ 's (appearing in this order from left to right in t).

Let $\lambda x.A, \lambda y.B$ and $\lambda z.C$ be sub-terms rooted at those λ 's (see Figure 3). Let $A' = A[x := y]$, let a be a new variable, let C' be the term obtained from C by replacing the leftmost occurrence of z with a and the others (possibly none) with y . Let $\varphi(t)$ be the term obtained from t by adding λa at the root, substituting both sub-terms $\lambda x.A$ and $\lambda z.C$ with a and replacing the leftmost occurrence of y in B with term $(A' C')$. We have $size(\varphi(t)) = size(t)$. Also note that since we chose the

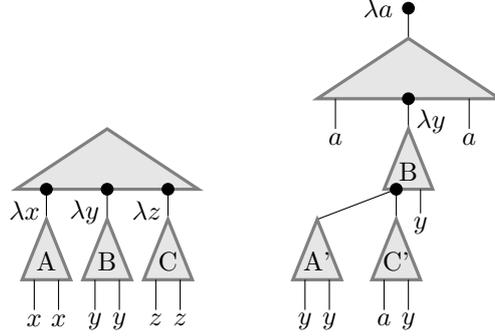


Figure 3: th. 18, the terms t and $\varphi(t)$.

highest three incomparable λ 's no variable becomes free in the constructed term. The injectivity of φ comes from the fact that both λy and the sub-term $(A' C')$ of $\varphi(t)$ are uniquely identifiable (see Figure 3):

- Let v_l (resp. v_r) be the deepest node above the two left-most (resp. right-most) occurrences of a . Remark that since there is exactly 3 occurrences of a , one of these two nodes is above the other. Let v be the deepest one. λy is the first binding λ on the path from the node v to the middle occurrence of a ;
- then, the application node $(A' C')$ is the deepest node above the middle occurrence of a and all the occurrences of y on the left of this middle occurrence of a .

Since the image of φ contains only terms starting with a λ which binds only 3 occurrences of the corresponding variable, by Theorem 17, the density of $\varphi(W_n)$ is equal to zero. The injectivity of φ finishes the proof. \square

7.4 A random term avoids any fixed closed term

Definition 19. Let t_0 be a term. We denote by Λ^{t_0} the set of terms having t_0 as a sub-term and by $\Lambda_n^{t_0}$ the set $\Lambda^{t_0} \cap \Lambda_n$

Theorem 20. Let t_0 be a term of size k' with k occurrences of free variables. Assume $k' \geq k + 1$. Then the density of $\Lambda_n^{t_0}$ is 0.

Proof. We construct a size preserving coding $\varphi: \Lambda^{t_0} \rightarrow \Lambda$ such that its image is of density 0.

There are at most $k' - k + 1$ occurrences of λ 's and at most $k' + 1$ leaves in t_0 , so there are at most

$$K = M(k')(k' + 1)^{k'+1}$$

such terms and we can enumerate them in a fixed way. Let m be the number of t_0 . The tree t_0 contains at least one occurrence of λ , since otherwise we would have $k' < k$. Let $g \in o(\frac{n}{3 \ln(n)})$ be such that $g(n) \xrightarrow{n \rightarrow \infty} \infty$. Let n be an integer satisfying $g(n) > K$.

Let $t \in \Lambda_n^{t_0}$ be a random term. By Theorem 14 the term t has more than m head λ 's since $m \leq K$ (see Figure 4).

Let us consider the term T which is obtained from the term t by adding an additional unary node (labelled with λx) at depth m . Let us define $\varphi(t)$ as the term T' obtained by replacing the left-most deepest sub-term t_0 in T by the term $t_1 = (U B)$ of size $k' - 1$ (see Figure 4), where U is a binary tree such that

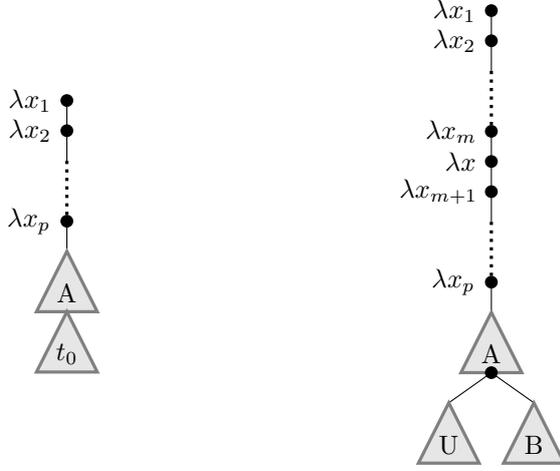


Figure 4: th. 20, the terms $t \in \Lambda_n^{t_0}$ and T' .

$U = (x (x (\dots (x x) \dots)))$ and $B = (x_1 (x_2 (\dots (x_{k-1} x_k) \dots)))$ (in case where t_0 has no free variables we put $t_1 = U$). Thus, the size of T' is equal to n . The variable x is bound by the m -th λ in the tree T' . Since m is the number of the tree t_0 , the function φ is injective.

By Theorem 17, each of K head λ 's in a random tree of size n binds more than k' variables. Trees from the image $f(\Lambda_n \cap \Lambda^{t_0})$ do not have this property, since the m -th λ binds only k' variables. Thus, those trees are negligible among all trees of size n . \square

Corollary 21. *Let t_0 be a term. If t_0 is closed or if there are at least two λ 's in t_0 , the density of Λ^{t_0} is 0.*

Proof. These are special cases of the previous theorem. \square

7.5 The density of strongly normalizable terms

From theorem 18, we know that almost all terms are of width at most 2. In this section, we introduce a notion of 'safeness' for terms of width 2 with the two following properties:

- safeness and width at most 2 implies strong normalisation (proposition 23);
- the set of unsafe terms of width 2 has density 0 (proposition 30).

The first part of this section is devoted to proposition 23 and is pure λ -calculus. We tried to write the proofs to be accessible for non specialist in λ -calculus. Nevertheless, For the basic fact 1 below, see [1] and for similar proofs techniques, see [18, 19]).

Definition 22 (fair and safe terms).

1. Let t be a term of width 1. We say that t is fair if there is no binding λ on the left branch of t (this includes the root node of t).
2. Let t be a term of width 2 and let $(u v)$ be the smallest sub-term of t of width 2. By definition, u and v have width 1. We say that t is safe if at least one of the term u or v is fair.

Proposition 23. *Let t be a safe term of width at most 2, then $t \in SN$.*

Definitions and notation 1.

- Let t be a term. If t is a term we denote by $\eta(t)$ the length of the longest reduction starting from t and $+\infty$ if t is not SN.
- Let σ be a substitution (that is a partial map from variables to terms). We write $t[\sigma]$ the capture free application of the substitution to t .
- A context, is a λ -term with a unique hole denoted \square . Traditionally context are defined by a BNF. If E is an arbitrary context, it is given by the following BNF:

$$E := \square \mid \lambda x.E \mid (E \ \Lambda) \mid (\Lambda \ E) \text{ where } \Lambda \text{ denote arbitrary terms.}$$

- When E is a context and t is a term, $E[t]$ denotes the replacement of the hole in E by t allowing capture: the λ 's in E can bind variables in t .
- For a context E , $\eta(E) = \eta(E[x])$ and $\text{size}(E) = \text{size}(E[x])$ where x is an arbitrary variable not captured by E .

Fact 1 (Basic fact on λ -terms and strong normalisation). *For some proofs in this section, we use the fact that a λ -term can be written in one of the following forms:*

- $t = (x \ t_1 \ \dots \ t_n)$ with $n \geq 0$ in which case $\eta(t) = \eta(t_1) + \dots + \eta(t_n)$ and t is SN if and only if t_1, \dots, t_n are SN.
- $t = \lambda x.u$ in which case $\eta(t) = \eta(u)$ and t is SN if and only if u is SN.
- $t = ((\lambda x.u) \ v \ t_1 \ \dots \ t_n)$ with $n \geq 0$ in which case $\eta(t) < \eta(u[x := v] \ t_1 \ \dots \ t_n)$ and t is SN if and only if $(u[x := v] \ t_1 \ \dots \ t_n)$ is SN.

Moreover, if t is a term and x is a variable, then t is SN if and only $(t \ x)$ is SN. This can be shown by induction on the size of t using the above case analysis.

Lemma 24.

1. The set of terms of width 0 (resp. of width at most 1) is closed by reduction.
2. If t is a term of width at most 1 then $t \in SN$.

Proof. (1) for width 0 is easy because substitution and reduction can not bind variables and width 0 means that all variables are free. For width 1, we first remark that width 1 means that all binding λ 's occur on the same branch. We consider such a term t and a β -reduction: $t = E[(\lambda x.u) \ v] \triangleright E[u[x := v]] = t'$. There are two cases: either x is not bound in u and $t' = E[u]$ or it is bound in u and v must have width 0 which means that all the free variables of v are free or bound by the context. In both cases, it is clear the t' is still of width 1 because the binding λ 's remain on one branch.

(2) follows from the fact that a reduction decreases the pair $\langle N_1(t), N_0(t) \rangle$ for the lexicographic ordering, where $N_1(t)$ (resp. $N_0(t)$) is the number of binding (resp. non binding) λ 's. To prove this, We consider again such a term t and a β -reduction: $t = E[(\lambda x.u) \ v] \triangleright E[u[x := v]] = t'$. If x is not bound in u , $N_1(t)$ is non increasing (it is decreasing if v contains some binding λ 's) and $N_0(t)$ is decreasing (we erase at least one non binding λ). If x is bound in u , we know that v is of width 0 and contain no binding λ , which means that we create no binding λ and therefore $N_1(t)$ is decreasing. \square

Lemma 25. *If u has width 0 and t_1, \dots, t_n are SN terms then the term $(u t_1 \dots t_n)$ is SN.*

Proof. By induction on the size of u , we distinguish three cases:

- If $u = x$, the result is trivial by the fact 1.
- If $u = (u' v)$, because of lemma 24, v is SN and we conclude by induction on u' .
- If $u = \lambda x.u'$, we use lemma 24 if $n = 0$ and we get that $(u' t_2 \dots t_n)$ is SN by induction otherwise. \square

Lemma 26. *Let $t \in SN$ be a term and σ be a substitution such that, for each x , there is k such that $\sigma(x) = (u v_1 \dots v_k)$ where u has width 0 and $v_1 \dots v_k$ are SN. Then, $t[\sigma] \in SN$.*

Proof. By induction on $\langle \eta(t), \text{size}(t) \rangle$. We consider the following cases:

- If $t = (x t_1 \dots t_n)$ and x is not in the domain of σ or $t = \lambda x.t_1$. In this case, it is enough to prove that for all i , $t_i[\sigma]$ is SN. This follows from the induction hypothesis because $\eta(t_i) \leq \eta(t)$ and $\text{size}(t_i) < \text{size}(t)$.
- If $t = ((\lambda x.u) v t_1 \dots t_n)$ we have to show that $(u[x := v] t_1 \dots t_n)[\sigma]$ is SN which follows from the induction hypothesis because $\eta(u[x := v] t_1 \dots t_n) < \eta(t)$.
- If $t = (x t_1 \dots t_n)$ and x is in the domain of σ . Then, $t[\sigma] = (\sigma(x) t_1[\sigma] \dots t_n[\sigma])$ which is SN by lemma 25 because $t_1[\sigma], \dots, t_n[\sigma]$ are SN by induction hypothesis and $\sigma(x) = (u v_1 \dots v_k)$ where u has width 0 and $v_1 \dots v_k$ are SN. \square

Definition 27. *We define the set of context of width 1 by the following BNF (where Λ_0 denotes the set of λ -terms of width 0):*

$$E := [] \mid \lambda x.E \mid (E \Lambda_0) \mid (\Lambda_0 E)$$

This definition means that all the binding λ 's are on the path from the root to the hole of the context. This justifies the name of such a context.

Lemma 28. *Let E be a context of width 1 and $u \in SN$ be a term. Then $E[u] \in SN$*

Proof. By induction on $\text{size}(E)$. The cases $E = []$ or $E = \lambda x.E_1$ are trivial (by induction in the second case because $\text{size}(E_1) < \text{size}(E)$).

If $E = (E_1 v)$ where $v \in \Lambda_0$, then $E[u] = (E_1[u] x)[x := v]$ where x is a fresh variable and $E_1[u]$ is SN by induction hypothesis because $\text{size}(E_1) < \text{size}(E)$. Therefore $(E_1[u] x)$ is SN by fact 1 and finally $(E_1[u] x)[x := v]$ is SN by Lemma 26.

The case $E = (v E_1)$ is symmetric. \square

Proof of proposition 23. If t has width at most one, this is lemma 24. If t has width 2, let $(t_1 t_2)$ be the smallest sub-term of t of width 2. This means that t can be written $E[t_1 t_2]$ where E is a context of width 1 and t_1 and t_2 have width 1. By lemma 28, it is therefore enough to show that $(t_1 t_2)$ is SN.

We know that t is safe. This means that either t_1 or t_2 is fair. If t_i is fair, it can be written $F[u v]$ where u has width 0, v has width 1 and F is a context using the following BNF:

$$F := [] \mid \lambda_.F \mid (F \Lambda_0) \text{ where } \lambda_. \text{ denotes non binding } \lambda\text{'s and } \Lambda_0 \text{ terms of width 0.}$$

The context F is defined precisely to denote the beginning of the left branch until we reach an application node whose argument is of width 1. The definition of fair term of width 1 ensures the existence of such an application node on the left branch.

This means that $(t_1 t_2)$ can be written $(F[u v] t_2)$ (resp. $(t_1 F[u v])$). Let us define $t' = (F[x] t_2)$ (resp. $t' = (t_1 F[x])$).

Therefore in both cases, $t = t'[x := (uv)]$, because the context F can not bind variables. Thus, we can conclude by lemma 26 because u has width 0 and because t' and v are SN (by lemma 24, since they have width 1). \square

To end this section and the proof of the main result (corollary 31), we establish a density result about unsafe terms of width 2.

Lemma 29. *The density of the set A of terms containing two consecutive non-binding λ is 0.*

Proof. We define an injective and size-preserving coding from A to the set of terms whose leading λ binds only once and the proof follows from theorem 17. The coding is as follows: in a term $t \in A$ of size n , we replace the subterm t_1 rooted at an occurrence of 2 non-binding lambdas, $t_1 = \lambda a. \lambda b. u$, by the term $x u$ where x is a fresh variable. We get a term t' of size $n - 1$ and the final coded term is $\lambda x. t'$ which is of size n . \square

Proposition 30. *The set of unsafe lambda terms with lambda width 2 has density 0.*

Proof. For every such a term, the root of the minimal subterm of width 2 is called the branching node and is always binary.

Let us divide the set of unsafe terms of width 2 into two parts:

S_1 : the set of terms such that both the lengths of paths from the branching node to the two highest independent binding lambdas is not greater than $\ln(n)$

S_2 : the set of remaining unsafe terms with lambda width 2.

The set S_1 can be encoded to the set of all terms in the following two steps. First, remove two highest independent binding lambdas and put one lambda, binding their variables, at the root of the whole term. The resulting size is smaller by 1 and the branching node is uniquely determined. Second, insert one lambda that binds nothing between the head lambdas of a term. According to Theorem 14 we have more than $\ln(n)^2$ head lambdas. Therefore we can encode the lengths of the paths from the branching node to the two highest binding lambdas by the position of this new lambda. Theorem 15 grants that the image of such transformation have density 0.

For the set S_2 proceed as follows: First, choose the path that is longer than $\ln(n)$. Let t_0 be the subterm rooted at the binding lambda at the end of this path (we assume it is the left path, the case of the right one is analogous). By lemma 29 we can suppose that at least half of the nodes on this path are binary. Let t_1, \dots, t_k be the right subtrees of the consecutive binary nodes on the path (the path goes always to the left since the term is unsafe). Second, chose some leaf v belonging to some subtree t_1, \dots, t_k and exchange it with the subterm t_0 . Independently of a choice of the leaf, the encoding can be reversed since the position of t_0 is uniquely identifiable as a highest binding lambda of the fair subtree of the branching node. The encoding is size preserving and the number of possibilities for the choice of a leaf v exceeds $\ln(n)/2$ therefore S_2 has density 0. \square

Corollary 31. *The set of strongly normalizable terms has density 1.*

Proof. First, by theorem 18, we can focus on terms of width at most 2. Proposition 30 shows in addition that we can restrict to the following types of terms:

- terms of width at most 1,
- safe terms of width 2.

Proposition 23 shows that they are all strongly normalizable. \square

8 Combinatory logic

Definition 32. 1. The set C of combinators is defined by the following grammar

$$C := K \mid S \mid I \mid (C C)$$

2. The size of a combinator is defined by the following rules: $\text{size}(S) = \text{size}(K) = \text{size}(I) = 1$ and $\text{size}((u v)) = \text{size}(u) + \text{size}(v)$.

3. The reduction on combinators is the closure by contexts of the following rules.

$$(K u v) \triangleright u \quad (S u v w) \triangleright (u w (v w)) \quad (I u) \triangleright u$$

Remark: It is easy to see that the number of internal nodes in a binary tree represented by a combinator is smaller by 1 than its size. Therefore, all the results concerning densities would be the same if we had defined the size as a number of internal nodes (like we have for λ -terms).

Proposition 33. 1. The generating function f enumerating the set of combinators is $f(z) = \frac{1 - \sqrt{1 - 12z}}{2}$.

2. The generating function f_{t_0} enumerating set of all combinators having t_0 as a sub-term is $f_{t_0}(z) = -\frac{\sqrt{1 - 12z}}{2} + \frac{\sqrt{1 - 12z + 4z^{n_0}}}{2}$.

Proof. 1. The function f thus satisfies

$$f(z) = 3z + f(z)^2.$$

Solving the equation and choosing between the two possibilities ($f(0) = 0$) gives the solution.

2. Assume that $n_0 = \text{size}(t_0)$. Using the fact that every combinator t having t_0 as a sub-term is either t_0 or has the form $t = (t_1 t_2)$ where either t_0 is a sub-term of t_1 but not of t_2 or t_0 is sub-term of t_2 but not of t_1 or finally t_0 is sub-term of both t_1 and t_2 we get the following equation.

$$f_{t_0}(z) = z^{n_0} + 2f_{t_0}(z) (f(z) - f_{t_0}(z)) + (f_{t_0}(z))^2$$

which can be simplified to

$$f_{t_0}(z) = z^{n_0} + 2 \cdot f_{t_0}(z) \cdot f(z) - (f_{t_0}(z))^2$$

Solving the quadratic equation with unknown function f_{t_0} and choosing between the two possibilities ($f_{t_0}(0) = 0$) gives the solution. \square

In the next theorem, the symbol $[z^n]\{F\}$ represents the coefficient of z^n in the series expansion of the generating function F .

Theorem 34. Let v, w be functions satisfying the following hypotheses:

- v, w are analytic in $|z| < 1$ with $z = 1$ being the only singularity at the circle $|z| = 1$.
- $v(z), w(z)$ in the vicinity of $z = 1$ have expansions of the form

$$v(z) = \sum_{p \geq 0} v_p (1-z)^{p/2}, \quad w(z) = \sum_{p \geq 0} w_p (1-z)^{p/2}.$$

Let \tilde{v} and \tilde{w} be defined by $\tilde{v}(\sqrt{1-z}) = v(z)$ and $\tilde{w}(\sqrt{1-z}) = w(z)$. Then

$$\lim_{n \rightarrow \infty} \frac{[z^n]\{v(z)\}}{[z^n]\{w(z)\}} = \frac{(\tilde{v})'(0)}{(\tilde{w})'(0)}.$$

Proof. This is a standard result in the theory of generating functions. For example, see [17]. \square

Theorem 35. *Let t_0 be a combinator. The density of combinatorics having t_0 as a sub-term is 1.*

Proof. The proof uses standard tools on generating functions. It follows from Proposition 33 and Theorem 34 below and some easy computations.

In order to satisfy assumptions of Theorem 34 we normalize functions in such a way to have the closest to the origin singularity located in $|z| \leq 1$ at the position in $z = 1$. So, we define functions $\overline{f_{t_0}}(z) = f_{t_0}(z/12)$ and $\overline{f}(z) = f(z/12)$. Therefore we have:

$$\begin{aligned} \overline{f_{t_0}}(z) &= -\frac{\sqrt{1-z}}{2} + \frac{\sqrt{1-z+4\left(\frac{z}{12}\right)^{n_0}}}{2} \\ \overline{f}(z) &= \frac{1}{2} - \frac{1}{2}\sqrt{1-z} \end{aligned}$$

This representation reveals that the closest singularity of $\overline{f_{t_0}}(z)$ and $\overline{f}(z)$ located in $|z| \leq 1$ is indeed $z = 1$. We have to remember that change of a caliber of the radius of convergence for functions f_{t_0} and f effects accordingly sequences represented by the new functions. Therefore those new functions enumerate two sequences $(12)^n ([z^n]\{f_{t_0}\}(z))$ and $(12)^n ([z^n]\{f(z)\})$. Now let us define functions \tilde{f} and \tilde{f}_{t_0} so as to satisfy the following equations: $\tilde{f}(\sqrt{1-z}) = \overline{f}(z)$ and $\tilde{f}_{t_0}(\sqrt{1-z}) = \overline{f_{t_0}}(z)$. Functions \tilde{f} and \tilde{f}_{t_0} are defined in the following way:

$$\begin{aligned} \tilde{f}_{t_0}(z) &= -\frac{z}{2} + \frac{\sqrt{z^2+4\left(\frac{1-z^2}{12}\right)^{n_0}}}{2} \\ \tilde{f}(z) &= \frac{1}{2} - \frac{1}{2}z \end{aligned}$$

The derivatives $(\tilde{f}_{t_0})'$ and $(\tilde{f})'$ are the following:

$$\begin{aligned} (\tilde{f}_{t_0})'(z) &= -\frac{1}{2} + \frac{\left(2z - 8\left(\frac{1-z^2}{12}\right)^{n_0} n_0 \cdot z\right)}{4(1-z^2)\sqrt{z^2+4\left(\frac{1-z^2}{12}\right)^{n_0}}} \\ (\tilde{f})'(z) &= -\frac{1}{2} \end{aligned}$$

Finally derivatives $(\tilde{f}_{t_0})'(0) = -\frac{1}{2}$ and $(\tilde{f})'(0) = -\frac{1}{2}$. To conclude the proof we use accordingly Theorem 34 so: $\lim_{n \rightarrow \infty} \frac{[z^n]\{f_{t_0}(z)\}}{[z^n]\{f(z)\}} = \lim_{n \rightarrow \infty} \frac{(12)^n [z^n]\{f_{t_0}(z)\}}{(12)^n [z^n]\{f(z)\}} = \frac{(\tilde{f}_{t_0})'(0)}{(\tilde{f})'(0)} = 1$. \square

Theorem 36. *The density of non strongly normalizing combinators is 1.*

Proof. Let $\Omega = (S I I (S I I))$. Then Ω reduces to itself and is thus not strongly normalizing. The theorem is thus an immediate consequence of the theorem 35. \square

9 Discussion

9.1 Other notions of size

The difference between Theorem 21 in the λ -calculus and Theorem 35 in combinatory logic may be surprising since there are translations between these systems which respect many properties (including strong normalization). However, these translations do not preserve the size.

The usual translation, which we denote by T_1 , from combinatory logic to λ -calculus is linear, i.e. there is a constant k such that, for all terms, $size(T_1(t)) \leq k * size(t)$. Note that this translation is far from being surjective: its image has density 0. The usual translation T_2 in the other direction (see [1]) is not linear. As far as we know, there is no known bound on the size of $T_2(t)$ but it is not difficult to find examples where $size(T_2(t))$ is of order $size(t)^3$.

The point is that T_2 has to code the variable binding in some way and this takes place. The difference between the two theorems comes probably from the definition of size that we have used for the variables in the λ -calculus. The usual way to implement coding of variables is to replace the names of variables by their de Bruijn indices: a variable is replaced by the number of λ 's that occur, on the path from the variable to the λ that binds it. Note that, in this case, different occurrences of the same variable may be represented by different indices.

Choosing the way in which we code de Bruijn indices gives different ways of defining the size of a term. This can be done in the following ways:

- using unary notation, i.e. the size of the index n is simply n itself;
- using binary notation, i.e. the size of the index n is $\lceil \log_2(n) \rceil$, i.e. the logarithm of n in base 2.

9.2 Some experiments

Although the results we proved concern only the model where the size of a variable is 0, we did some experiments on the other models. There is an easy algorithm (polynomial time in n) to compute L_n for each model of size. This algorithm can be sometimes adapted to compute (still in polynomial time) the number of terms of size n having a given property P . We did this for several simple syntactical properties until size 1000. It is always a strange exercise to guess the limit of a sequence from its first values, but our results, at least, suggest the following:

- Almost all terms start with several λ 's for model with constant size variables (99.99% start with at least one λ for size 1000), whereas it is not clear that terms starting with an application are negligible for other models;
- Identity almost always (exceptions represent a fraction of terms less than 10^{-5} for size 500) occur for models with non-constant size of variables, whereas at least 80% of terms don't contain identity for model with variables of constant size (for variables of size 0, we now that it goes toward 100%).

9.3 Future work and open questions

We give here some questions for which it will be desirable to have an answer.

- Give an asymptotic equivalence for L_n or, at least, better upper and lower bounds.

- Give the density of typable terms. Numerical experiments done by Jue Wang (see [16]) seem to show that this density is 0.
- Compute the densities of strongly normalizing terms with other notions of size (mainly by changing the size of variables, and eventually making it non constant). If we can not simplify the present proof of density 1 of SN terms (corollary 31), it seems very difficult to extend this result if only for variables having size 1: most encoding techniques really use the fact that variables have size 0. However, we believe that proving theorem 14 is an achievable goal for variables of size 1.

References

- [1] H. Barendregt, *The Lambda Calculus. Its Syntax and Semantics*. Studies in Logic and The Foundations of Mathematics, vol 103, North-Holland.
- [2] B. Chauvin, P. Flajolet, D. Gardy and B. Gittenberger. And/Or trees revisited, *Combinatorics, Probability and Computing*, 13(4-5):475-497, 2004.
- [3] P. Flajolet and R. Sedgewick, *Analytic combinatorics*. Cambridge University Press, 2008.
- [4] H. Fournier, D. Gardy, A. Genitrini, M. Zaionc *Classical and intuitionistic logic are asymptotically identical*, Computer Science Logic 2007, Lecture Notes in Computer Science 4646, pp 177-193.
- [5] D. Gardy. Random Boolean expressions, *Colloquium on Computational Logic and Applications*, Chambéry (France), June 2005. In *Discrete Mathematics and Theoretical Computer Science* proceedings AF, pp 1-36, 2006.
- [6] D. Gardy and A. Woods. And/or tree probabilities of Boolean function, *Discrete Mathematics and Theoretical Computer Science*, pp 139-146, 2005.
- [7] Antoine Genitrini, Jakub Kozik, *Quantitative comparison of Intuitionistic and Classical logics - full propositional system*, LFCS09, Lecture Notes in Computer Science 5407 (2009) pp. 280-294.
- [8] A. Genitrini, J. Kozik, M. Zaionc, *Intuitionistic vs. Classical Tautologies, Quantitative Comparison*, Lecture Notes in Computer Science 4941, (2008) pp. 100-109.
- [9] J.D. Hamkins and A. Miasnikov, *The halting problem is decidable on a set of asymptotic probability one*, Notre Dame J. Formal Logic 47 (2006) (4), pp. 515-524.
- [10] Z. Kostrzycka and M. Zaionc. Statistics of intuitionistic versus classical logic, *Studia Logica*, 76(3):307-328, 2004.
- [11] A. Rybalov, *On the strongly generic undecidability of the Halting Problem*, Theoretical Computer Science, Volume 377, Issues 1-3, 31 May 2007, Pages 268-270.
- [12] H. Lefmann and P. Savický. Some typical properties of large And/Or Boolean formulas, *Random Structures and Algorithms*, vol 10, pp 337-351, 1997.
- [13] M. Moczurad, J. Tyszkiewicz and M. Zaionc. Statistical properties of simple types, *Mathematical Structures in Computer Science*, 10(5):575-594, 2000.
- [14] L. Boyer and G. Theyssier, *On Local Symmetries and Universality in Cellular Automata* 26th International Symposium on Theoretical Aspects of Computer Science (STACS), 2009. Dagstuhl Seminar Proceedings, <http://stacs2009.informatik.uni-freiburg.de/proceedings.php>.

- [15] M. Zaionc. On the asymptotic density of tautologies in logic of implication and negation, *Reports on Mathematical Logic*, vol 39, pp 67-87, 2005.
- [16] Jue Wang, *Generating Random Lambda Calculus Terms*. <http://cs-people.bu.edu/juewang/research.html>
- [17] M. Zaionc, *Probability distribution for simple tautologies*. *Theoretical Computer Science*, 355(2):243-260, 2006.
- [18] R. David, *Normalization without reducibility*. APAL n 107 (2001) p 121-130.
- [19] R. David, *A short proof of the strong normalization of the simply typed lambda calculus*. <http://www.lama.univ-savoie.fr/~david/>