Some properties of random lambda terms

René David, Christophe Raffalli, Guillaume Theyssier, Katarzyna Grygiel, Jakub Kozic, Marek Zaionc

Some properties of random λ-terms

René David, Christophe Raffalli, Guillaume Theyssier*
(Université de Savoie)
Katarzyna Grygiel, Jakub Kozik, Marek Zaionc†
(Jagiellonian University)

Abstract

We show various (syntactic and behavioral) properties of random λ-terms. Our main results are that at least $3/4$ of 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 different. We show that almost all terms are not *strongly normalizing, because any fixed term almost always appears in a random term.*

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 i.e. the probability is 1? The notion of random program is precisely defined in Section 5.

We concentrate on functional programming languages and, more specifically, on the λ-calculus, the simplest such language (see [5, 6, 10] 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 [11]). 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 4) 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.

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 16). Moreover, each of them binds "many" occurrences of variables (Theorems 17, 18 and 19). Finally, given any fixed closed λ-term, almost no λ-term has this term as a sub-term (Theorem 23).

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 at least $3/4$ and we conjecture it is 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 which is fair for questions we are concerned with. This means that there are translations, in both directions, which, for example, preserve the property of being SN. Surprisingly, our results concerning random programs 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 different results for SN 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 7.

The organization of the paper is as follows. In Section 2 we recall basic definitions of the λ-calculus. Section 3 gives combinatorial results we need in our proofs. The lower and upper bounds for the number of λ-terms of size n are given

*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

in Section 4. Our main results appear in Section 5. Section 6 contains results in combinatory logic, namely that every fixed term appears in almost every term. Section 7 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. Appendix at the end gives missing details for some proofs.

Note that, in sections 2 to 5, we do not aim at providing the best possible estimations for the analyzed sequences. Most of them can be easily improved. We present results which are sufficient for our structural results, without sacrifying the simplicity of proofs for better estimations (unless the result we get is tight).

Acknowledgments

This research has been partially supported by a project funded by the French Rhône-Alpes region and initiated by P. Lescanne.

2 Generality on the λ-calculus

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

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

As usual, λ-terms are considered modulo the α-equivalence i.e. two terms equal modulo a change in their bound variables are considered equal. Note that λ-terms can be seen as trees. Thus, we often use (without giving the precise definition) the classical terminology about trees (e.g. branch, path, root, leaf, etc.)

Notation 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 really occurs free in a.*
   - *We say that λx is in head position if $t = \lambda \overrightarrow{y}.u$ for some, possibly empty, sequence of abstractions $\lambda \overrightarrow{y}$.*
3. *The unary height of a term t is the maximal number of λ's in a branch of t.*
4. *The binary height of a term t is the maximal number of applications in a branch of t.*

Definition 3. *The size (denoted as $size(t)$) of a term 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 terms of size n and by L_n its cardinality.*

Remark

This is not the usual definition of size. Most often the size of a variable is 1. We choose such definition (by analogy to the research on trees), hoping that our results can be easily translated to the usual case. The translation turned out not to be trivial and will be a subject of our future work. From the point of view of implementation, both definitions are not realistic (see the discussion in Section 7).

3 Classical combinatorial results

3.1 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 5. *$C(n) \sim \frac{4^n}{n^{3/2}\sqrt{\pi}}$ 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 [4]. □

3.2 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 6.

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) A tree enumerated by $M(n, k)$ has $n + k$ nodes, among which $k - 1$ are binary and $n - k + 1$ are unary. (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 [4]. □

3.3 Lambert W function

The Lambert function $W(x)$ is defined by the equation $x = W(x)e^{W(x)}$ which has a unique solution in $\mathbb{R}_+$. We use the following proposition.

Proposition 7. *1. Let a, b, c be positive constants and $M = \frac{b}{cW(\frac{abe}{c})}$. The function f defined by $f(x) = (ax)^{b-cx}$ is increasing on $]0, M]$ and decreasing on $[M, +\infty[$.*

2. *For x large enough, we have* $\ln(x) - \ln(\ln(x)) \leq W(x) \leq \ln(x) - \ln(\ln(x)) + 1$

Proof. The proof is given in Appendix. □

4 Bounds for L_n

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.

4.1 Lower Bound for L_n

Theorem 8. *For any $\varepsilon > 0$ and for n large enough we have*

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

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 with k possibilities for each leaf, we have $LB(n,k) = C(n-k)k^{n-k+1}$. Since $n+1 \geq n-k+1$, by Proposition 5 we get $LB(n,k) \geq K\frac{(4k)^{n-k+1}}{n^{\frac{3}{2}}}$ for some constant K. The result follows from Lemma 9 below. □

Lemma 9. *For any $K, \varepsilon > 0$ and for n large enough we have*

$$\sum_{k=1}^{n} K\frac{(4k)^{n-k+1}}{n^{\frac{3}{2}}} \geq \left(\frac{(4-\varepsilon)n}{\ln(n)}\right)^{n-\frac{n}{\ln(n)}}$$

Proof. The proof is given in Appendix. □

4.2 Upper Bound for L_n

The computation of an upper bound uses Theorem 13. Since this result is not used in section 5 the proof is only given in the appendix.

Theorem 10. *For all $\varepsilon > 0$ and for n large enough, we have*

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

4.3 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 11. *For all $\varepsilon > 0$ and for n large enough*

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

5 Main results

By an analogy to researches on graphs and trees, we define a notion of a "random" term using asymptotic densities.

Definition 12. *Let A be a set of terms.*
1) We denote by $\#(A)$ the cardinality of A.
2) We denote by $d(A)$ the following limit (if it exists):

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

Remarks and notation

- Note that d is not a measure, in the usual sense, since it is not countably additive. Moreover, note that $d(A)$ may be undefined if the previous limit does not exist.
- Let P be a property of terms. If $d(\{t \mid P(t) holds\}) = \alpha$, we say that the density of terms satisfying P is α. If $\alpha = 1$, we say that a random term satisfies P.

The idea of proofs

We use two different technics.

In Section 5.1, a density 0 is proved by computing an upper bound on the cardinality of the set we are considering, by using the lower bound given in Theorem 8 and by showing that the quotient tends to zero. The computation of these limits is quite standard. A sketch is given in Appendix. 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 5.2 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.

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 random 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

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

5.1 Proofs using calculus

5.1.1 The number of λ's in a term

Theorem 13. *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 $UB(n,p) = M(n, n-p+1)p^{n-p+1}$. This is because a term with p λ's is a unary binary tree whose $n-p+1$ leaves are bounded by, at most, p nodes. By Proposition 7.(1) the function p^{n-p+1} is decreasing on $[\frac{n+1}{W(e(n+1))}, +\infty[$. Observing that, for $k>1$, $\frac{n+1}{W(e(n+1))} < \frac{kn}{\ln(n)}$ for n large enough (this is because the quotient tends to k) we thus have

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

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

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

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

(2) The proof is similar. □

5.1.2 Bounds on the unary height of a term

Theorem 14. *The set of terms with the unary height greater than $\frac{n}{3\ln(n)}$ has density 1.*

Proof. Let M_n be the number terms of size n with more than $\frac{n}{3\ln(n)}$ λ's and with the unary height less than $\frac{n}{K\ln(n)}$ Because a variable can only be bound by λ above and because such a term has at most $n - \frac{n}{3\ln(n)} + 1$ leaves, we have

$$M_n \leq M(n)\left(\frac{n}{K\ln(n)}\right)^{n-\frac{n}{3\ln(n)}+1}$$

Dividing by our lower bound for the number of terms (Theorem 8) we get $\frac{M_n}{L_n} \leq \Psi(n,4-\varepsilon)$ where

$$\Psi(n,q) = M(n)\frac{\left(\frac{n}{K\ln(n)}\right)^{n-\frac{n}{3\ln(n)}+1}}{\left(\frac{qn}{\ln n}\right)^{n-\frac{n}{\ln(n)}}}$$

It remains to show that, for $K=3$ and any $\varepsilon>0$, $\Psi(n,4-\varepsilon)$ tends to 0. This is done in Appendix. □

5.1.3 Bounds on the binary height of a term

Since the following theorem is not used later, the proof is given in the appendix.

Theorem 15. *The density of terms having binary height greater than $\frac{n}{\sqrt{\ln(\ln(n))}}$ equals 0.*

5.2 Proofs using coding

5.2.1 The number of λ's in head position

Theorem 16. *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 random terms of size n with less than $g(n)$ head λ's. We construct an injective, size-preserving function (coding) $\varphi\colon A_n \to \Lambda_n$ such that its image has density 0.

Let $t \in A_n$. We can write $t = \lambda x_1 \ldots \lambda x_p.M$, where $p < g(n)$ and M is a term starting with an application and containing at least one λ (by Theorem 14). Let B be the maximal purely applicative prefix of M i.e. B is the term using only application nodes and variables such that $M = B[\overrightarrow{t}]$ where terms in $\overrightarrow{t}$ start with λ and variables in B are taken from the set $\{x_1,\ldots,x_p\}$ (see Figure 1). Let us denote by $A(n,p,b,\overrightarrow{t})$ the set of terms in A_n having, as in the decomposition of t above, p head λ's, then a purely applicative context of size b, and, in that context, a sequence $\overrightarrow{t}$ of subterms beginning by λ. Because $p < g(n)$ the cardinal of $A(n,p,b,\overrightarrow{t})$ is less than

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

Let $t \in A(n,p,b,\overrightarrow{t})$ where $\overrightarrow{t} = [t_1,\ldots,t_k]$. 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 $T = \lambda z \lambda x_1 \ldots \lambda x_p.(u'_1\ (u'_2\ (\ldots(u'_{k-1}\ u'_k)\ldots)$ which is

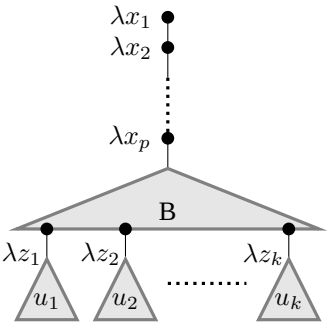

Figure 1. th. 16, the term $t \in A_n$.

of size $n - b$. Let $\lambda y.C$ denote the term rooted at the left-most 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 14 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 $P(b,n) < Q(b,n)$ (this is 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 $t \in A(n,p,b,\overrightarrow{t})$ the cardinal of $\Psi(t)$ is always $Q(b,n)$. By construction, for any pair fo distinct terms t and t', the sets $\Psi(t)$ and $\Psi(t')$ are disjoint.

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 $\psi(b,n) = \frac{C(b+1)}{3^b C(b)}(\frac{n}{\ln(n)})^{\frac{1-b}{2}}\varepsilon(n)$. Since $\frac{C(b+1)}{3^b C(b)}$ is bounded, since, for $b \geq 2$, $(\frac{n}{\ln(n)})^{\frac{1-b}{2}}$ is decreasing in b, it follows that $\psi(b,n)$ tends to 0 uniformly in b. Since the $A(n,p,b,\overrightarrow{t})$ form a partition of A_n, the result follows. □

5.2.2 Head λ's bind "many" occurrences

Theorem 17. *Let $g(n) \in o\big(\sqrt{n/\ln(n)}\big)$. 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.*

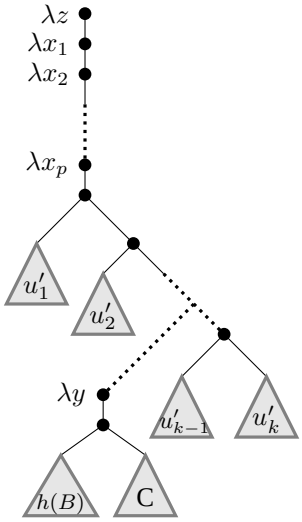

Figure 2. th. 16, the term $\varphi(t)$.

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

Let $T = \lambda x_1 \ldots x_{g(n)}.A$ 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

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

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 16, 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. □

Theorem 18. *Let $g(n) \in o\big(\ln(n)\big)$. 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 $g(n) \in o\big(\ln(n)\big)$ and denote by $T_{g(n)}$ 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 function $\varphi\colon T_{g(n)} \to \Lambda_n$ such that the image of f is of density zero in Λ_n.

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 $T_{g(n)}$. 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 14, 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 $\left(\frac{n}{6\ln(n)}\right)$ elements in $\mathcal{B}(T)$. As $g(n) \in o(\ln(n))$, we have

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

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 $\lambda a.\big((y\ B)\ U\big)$, where $B = h_{[T]}(T)$.

All terms from the image $\varphi(T_{g(n)})$ start with λ that binds no variable. By Theorem 17 we know that the set of such terms have density zero in Λ_n. Since f is injective, the density of $T_{g(n)}$ is zero, as well. □

Theorem 19. *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 18, 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 random terms of size n having exactly q λ's in head position 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. Finally let ψ be the function defined by $\psi(q) = \frac{k\cdot k'\cdot q^{k'}\cdot(k-1)^{q-k'}}{k^q}$. Consider the equivalence relation $\sim_n$ defined analogously to the relation from the proof of Theorem 18, 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 at least $k^{g(n)}$ 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 $\psi(q)$ which, since ψ is decreasing, is less than $\psi(g(n))$.

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

5.2.3 The width of a term

Definition 20. *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.*

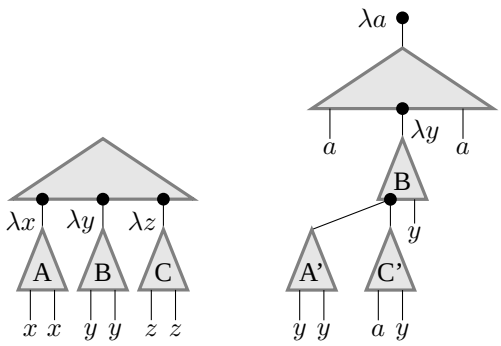

Figure 3. th. 21, the terms t and $\varphi(t)$.

Theorem 21. *The density of terms having λ-width at most 2 is 1.*

Proof. Let us denote by W_n the set of terms of size n with λ-width greater than 2. We show that there exists an injective function $\varphi\colon W_n \to \Lambda_n$ such that its image has density 0. Let t be an element of W_n and let us denote by λx, λy 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 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 leftmost (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 19, the density of $\varphi(W_n)$ is equal to zero. The injectivity of φ finishes the proof. □

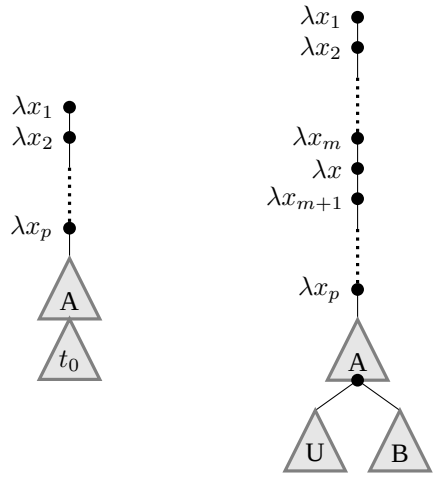

Figure 4. th. 24, the terms $t \in \Lambda_n^{t_0}$ and T'.

5.2.4 A random term avoids a fixed closed term

Definition 22. *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 23. *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 next theorem. □

Theorem 24. *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 coding $\varphi\colon \Lambda_n^{t_0} \to \Lambda_n$ 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 $g \in o\big(\frac{n}{3\ln(n)}\big)$ be such that $g(n) \xrightarrow{n\to\infty} \infty$. Let n be an integer satisfying $g(n) > K$.

Let $t \in \Lambda_n^{t_0}$ be a random term. By Theorem 16 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 $U = (x\ (x\ (\ldots (x\ x)\ \ldots)))$ and $B = (x_1\ (x_2\ (\ldots (x_{k-1}\ x_k)\ \ldots)))$ (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 19, 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. □

5.2.5 The density of strongly normalizable terms

Definition 25.
- *The reduction (denoted as $t \rhd t'$) on terms is the contextual closure of the β- reduction i.e. $(\lambda x.a\ b) \rhd a[x := b]$*
- *A term t is strongly normalizable (denoted as $t \in SN$) if every sequence of reduction starting from t is finite.*

Theorem 26. *The density of strongly normalizing terms is (if it exists) at least $\frac{3}{4}$.*

Proof. This is done as follows. We use Theorem 28 below (usually known as the Ω-theorem) which states that a syntactic property (that we call safety) implies strong normalization. We then give 4 encodings of unsafe terms in such a way that the images of these encodings are disjoint and have the same cardinality (see Theorem 29). This implies that, if the set of unsafe terms has a density, this density is less than $\frac{1}{4}$. □

Definition 27. *Say that a term t is unsafe if there are contexts H, E_l, E_r and terms P_l, P_r such that $t = H[(E_l[\lambda x_l.P_l]\ E_r[\lambda x_r.P_r])]$ and, for $i \in \{l, r\}$, x_i occurs at least twice in P_i.*

Theorem 28 (Ω-theorem)**.** *If $t \notin SN$ then t is unsafe.*

Proof. This is theorem 1.4.13 p34 in [7]. □

Theorem 29. *If the set of unsafe terms has a density, this density is less than $\frac{1}{4}$.*

Proof. Since this proof is rather involved, we give in the appendix, a simpler one where we show that the density of the set of unsafe terms is less than $\frac{1}{2}$.

By definition 27 write $t = H[(M_l\ M_r)]$ where, for $i \in \{l, r\}$, $M_i = E_i[\lambda x_i.P_i]$ and x_i has at least two occurrences in P_i. Assume that $(M_l\ M_r)$ is the left-most highest pattern in t with the given property and that, for $i \in \{l, r\}$, $\lambda x_i.P_i$ is the left-most highest pattern in M_i with the given property. We partition the set of unsafe terms in four disjoint subsets.

1. Let $U_{n,0}$ be the set of unsafe terms of size n for which, for both i in $\{l, r\}$, E_i binds no variable in $\lambda x_i.P_i$.
2. Let $U_{n,l}$ (resp. $U_{n,r}$) be the set of unsafe terms of size n for which E_l (resp. E_r) binds at least one variable in $\lambda x_l.P_l$ (resp. $\lambda x_r.P_r$) and E_r (resp. E_l) binds no variable in $\lambda x_r.P_r$ (resp. $\lambda x_l.P_l$).

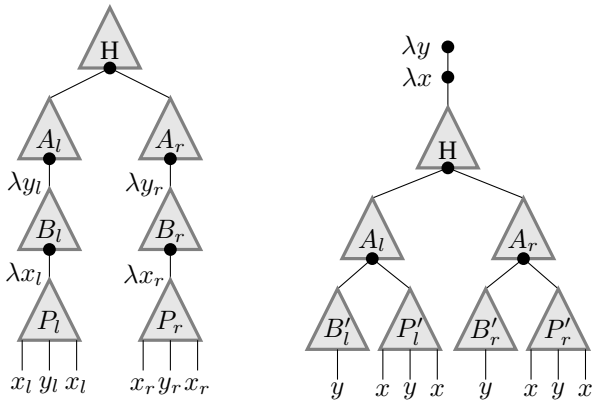

Figure 5. th. 29, the terms t and $F(t)$.

3. Let $U_{n,1}$ be the set of unsafe terms of size n for which E_l and E_r bind respectively at least one variable in $\lambda x_l.P_l$ and $\lambda x_r.P_r$

Note that, by these definitions, when E_i binds a variable y_i in $\lambda x_i.P_i$ there is exactly one occurrence of y_i in P_i.

We first show that that $U_{n,1}$ is negligible. This is done by giving an injective function F from $U_{n,1}$ to a negligible sub-set of Λ_n (see figure 5). For $i \in \{l, r\}$ write $E_i[] = A_i[\lambda y_i.B_i[]]$ where λy_i is the first λ in the path of M_i (down-up) from $\lambda x_i.P_i$ to the root such that y_i has an occurrence in P_i. Let x, y be fresh variables. Let $B'_i = B_i[y]$, $P'_i = P_i[x_i := x, y_i := y]$. Finally let $F(t) = \lambda y, x.H[(A_l[(B'_l\ P'_l)]\ A_r[(B'_r\ P'_r)])]$. It is clear that $F(t)$ is closed, has size n and that F is injective. Since $F(t)$ has exactly 4 occurrences of y, by Theorem 19, this finishes the proof that $U_{n,1}$ is negligible.

The encodings are defined in the following way. Let x be a fresh variable, $P'_i = P_i[x_i := x]$ and $E'_i = E_i[x]$.

1. If $t \in U_{n,0}$ (see figure 6) let
   $F_1(t) = \lambda x.H[(E'_l\ E_r[(P'_l\ P'_r)])]$,
   $F_2(t) = \lambda x.H[(E_l[(P'_l\ P'_r)]\ E'_r)]$,
   $F_3(t) = \lambda x.H[(P'_l\ E_r[(E'_l\ P'_r)])]$ and
   $F_4(t) = \lambda x.H[(P'_r\ E_r[(P'_l\ E'_l)])]$.
   Note that, since $t \in U_{n,0}$, the $F_j(t)$ are closed.

2. For $t \in U_{n,r}$ (see figure 7) write $E_r[] = A_r[\lambda y_r.B_r[]]$ where λy_r is the first λ in the path of M_r (down-up) from $\lambda x_r.P_r$ to the root such that y_r has an occurrence in P_r. Let x, y be fresh variables. For $j \in \{1, 2, 3, 4\}$, let $F_j(t) = \lambda x.H[(E_l[x]\ =\ A_r[\lambda y_r.B_r[G_j(t)]])]$ where $G_1(t) = (P'_l\ P'_r)$, $G_2(t) = (P'_r\ P'_l)$, $G_3(t) = (P'_l\ Q'_r)$ where Q'_r is obtained from P'_r by replacing the left-most occurrence of x by y and $G_4(t) = (P'_l\ Q''_r)$ where Q''_r is obtained from P'_r by replacing the right-most occurrence of x by y. Since $t \in U_{n,r}$, the $F_j(t)$ are closed.

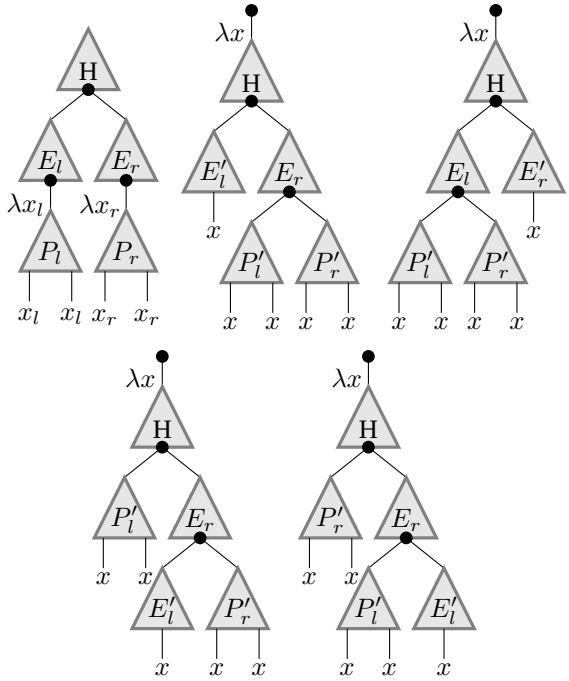

Figure 6. th. 29, $t \in U_{n,0}$, $F_1(t)$, $F_2(t)$, $F_3(t)$, $F_4(t)$.

3. For $t \in U_{n,l}$ let $F_j(t) = \lambda x.H[(E_r[G'_j(t)]\ E_l[x])]$ where the G'_j are computed as the G_j by permuting the roles of the indices l, r.

To finish the proof, it remains to check that

- $F_i \restriction U_{n,j}$ is injective for each i, j. This is done (as in previous proofs) by showing that each new node can be recovered.

- Each F_i preserves the size and, for each j, the cardinality of the images of the $F_i \restriction U_{n,j}$ are identical . This is immediate.

- The images of the $F_i \restriction U_{n,j}$ are all disjoint. This is quite fastidious but easy. Most often, looking at the number of occurrences of x in some branches is enough but, for some cases, we must also look at other variables. For example to show that $F_1(t) \neq F_1(t')$ for $t \in U_{n,0}$ and $t' \in U_{n,r}$ we use the fact that in some part of $F_1(t)$ there is no binding λ whereas there is one in $F_1(t')$. □

6 Combinatory logic

Definition 30. *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

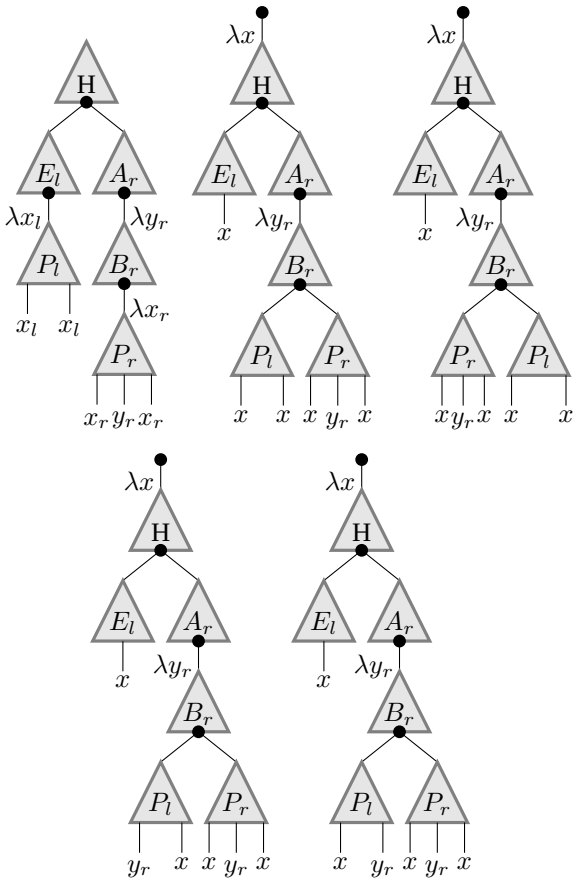

Figure 7. th. 29, $t \in U_{n,r}$, $F_1(t)$, $F_2(t)$, $F_3(t)$, $F_4(t)$.

rules: $size(S) = size(K) = size(I) = 1$ *and* $size((u\ v)) = size(u) + size(v)$.

3. *The reduction on combinators is the closure by contexts of the following rules.* $(K\ \ u\ \ v) \rhd u$ $(S\ u\ v\ w) \rhd (u\ w\ (v\ w))$ $\qquad$ $(I\ u) \rhd u$

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).

Theorem 31. *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 next theorem. □

Theorem 32. *Let t_0 be a combinator. The density of combinators 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. The details are given in Appendix. □

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 $f_{t_0}(z) = -\frac{\sqrt{1-12\,z}}{2} + \frac{\sqrt{1-12\,z+4\,z^{n_0}}}{2}$.

Proof. The proof is given in Appendix. □

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

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

Proof. This is a standard result in the theory of generating functions. For example, see [13]. □

7 Conclusion

7.1 Other notions of size

The difference between Theorem 23 in the λ-calculus and Theorem 32 in combinatory logic may be surprising since there are translations between these systems which respect many properties (for example the one of being terminating). 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 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 one we have used (or the one with the size of a variable being 1) are, for the implementation point of view, not realistic because, in case a term has a lot of distinct variables, it is not realistic to use a constant number of bits to code them. The usual way to implement this coding 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 root, between the variable and 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.

7.2 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%).

7.3 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 strongly normalizable terms. We conjecture that it is 1.

- Give the density of typable terms. Numerical experiments done by Jue Wang (see [11]) seem to show that this density is 0.

- Compute the densities of sets studied in Section 5 with other notions of size.

References

[1] H. Barendregt, *The Lambda Calculus. Its Syntax and Semantics*. Studies in Logic and The Foundations of Mathematics, vol 103, North-Holland.

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

[3] Philippe Flajolet and Zhicheng Gao and Andrew M. Odlyzko and L. Bruce Richmond, *The Distribution of Heights of Binary Trees and Other Simple Trees*, volume 2 of *Combinatorics, Probability & Computing*. Cambridge University Press, 1993.

[4] P. Flajolet and R. Sedgewick, *Analytic combinatorics*. Cambridge University Press, 2008.

[5] 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.

[6] 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.

[7] M.H.B Sørensen, *Normalization in λ-calculus and Type theory*, PhD thesis University of Copenhagen. ftp://ftp.diku.dk/diku/semantics/papers/D-367.ps.gz

[8] M. Sørensen and P. Urzyczyn. *Lectures on the Curry-Howard Isomorphism*, volume 149 of *Studies in Logic and the Foundations of Mathematics*. Elsevier Science, 2006.

[9] G. Szegö, *Orthogonal polynomials*, fourth ed., AMS, Colloquium Publications, 23, Providence, 1975.

[10] G. Theyssier, *How common can be universality in cellular automata?* 22nd Symposium on Theoretical Aspects of Computer Science. Volume 3404 of Lecture Notes in Computer Science, pp 121-132, 2005.

[11] Jue Wang, *Generating Random Lambda Calculus Terms*. http://cs-people.bu.edu/juewang/research.html

[12] H.S. Wilf, *generatingfunctionology*, second ed., Academic Press, Boston. 1994.

[13] M. Zaionc, *Probability distribution for simple tautologies*. Theoretical Computer Science, 355(2):243-260, 2006.

8 Appendix

8.1 Proof of Proposition 7

The first point is immediate. The second is as follows. Let $A(x) = \ln(x) - \ln(\ln(x))$ and $f(y) = ye^y$.

Since f is increasing, to show that $A(x) \leq W(x)$ it is enough to show that $f(A(x)) \leq x$ which is trivial since this means:

$$(\ln(x) - \ln(\ln(x))e^{\ln(x)-\ln(\ln(x))} = x\left(1 - \tfrac{\ln(\ln(x))}{\ln(x)}\right) \leq x$$

which is true for $x \geq e$.

Using the fact that f is convex we get (see Figure 8):

$$W(x) \leq A(x) - \tfrac{f(A(x))-x}{f'(A(x))} = A(x) + \tfrac{\ln(\ln(x))}{\ln(x)-\ln(\ln(x))+1}$$

which is less that $A(x) + 1$ for x large enough.

8.2 A useful lemma

Lemma 35.

$$\binom{2n+1}{\frac{3n}{\ln(n)}} = o\left(e^{4n\frac{\ln(\ln(n))}{\ln(n)}}\right)$$

Proof. Let $v_n = \binom{2n+1}{\frac{3n}{\ln(n)}}$. Use Stirling approximation to replace p ! by $Cp^{p+\frac{1}{2}}e^{-p}$ and get $v_n \sim w_n$. By developing $\ln(w_n)$ we get $\ln(w_n) = 3n\frac{\ln(\ln(n))}{\ln(n)} + o(n\frac{\ln(\ln(n))}{\ln(n)})$ which gives the desired result. □

8.3 Proof of Lemma 9

Let $\alpha = k/n$. By Proposition 7(1) the maximum of $f(\alpha) = (4n\alpha)^{n(1-\alpha)+1}$ is obtained for $\alpha = k_1 = \frac{n+1}{W(4e(n+1))}$. Let $k_0 = \frac{n+1}{\ln(4e(n+1))}$. By Proposition 7(2) we have

$$k_1 - k_0 = \frac{n(\ln(4e(n+1)) - W(4e(n+1)))}{W(4e(n+1))\ln(4e(n+1))}$$

$$\geq \frac{n(\ln(\ln(4e(n+1))) - 1)}{\ln^2(4e(n+1))}$$

Thus, $\Phi(n) \geq \Psi(n)$ where

$$\Phi(n) = \sum_{k=1}^{n} K\frac{(4k)^{n-k+1}}{n^{\frac{3}{2}}}$$

$$\Psi(n) = K\left\lfloor \frac{n(\ln(\ln(4e(n+1))) - 1)}{\ln^2(4e(n+1))} \right\rfloor \frac{(4k_0)^{n-k_0+1}}{n^{\frac{3}{2}}}$$

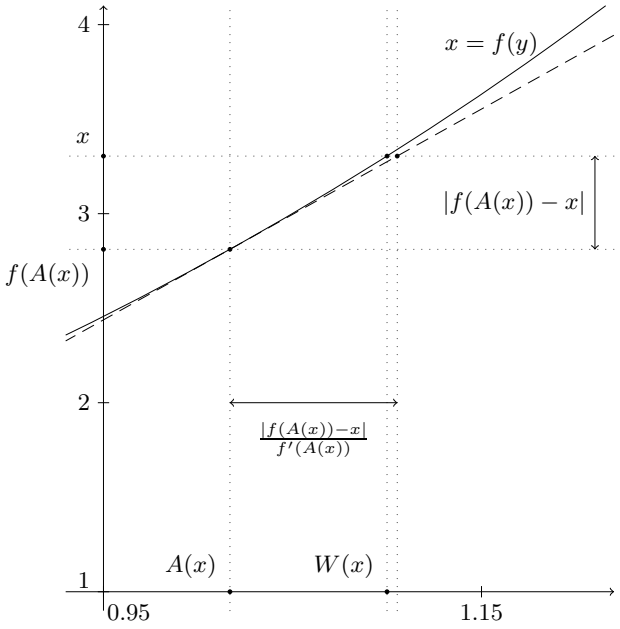

Figure 8. Bound for Lambert W function

The result follows then immediately from the following fact:

Fact $\lim_{n\to+\infty} \frac{\Psi(n)}{\left(\frac{(4-\varepsilon)n}{\ln(n)}\right)^{n-\frac{n}{\ln(n)}}} = +\infty$ for any $\varepsilon > 0$.

Proof Replacing K and $\ln(\ln(4e(n+1))) - 1$ by any constant C we get the following inequalities for n large enough:

$$\begin{aligned}
\Psi(n) &\geq C\tfrac{n}{\ln^2(4e(n+1))}\tfrac{(4k_0)^{n-k_0+1}}{n^{\frac{3}{2}}} \\
&\geq C\tfrac{\sqrt{n}}{\ln^3(4e(n+1))}\left(\tfrac{4(n+1)}{\ln(4e(n+1))}\right)^{n-\frac{n+1}{\ln(4e(n+1))}} \\
&\quad \text{Using } \frac{n+1}{\ln(4e(n+1))} \leq \frac{n}{\ln(n)}\text{, we get} \\
&\geq C\tfrac{\sqrt{n}}{\ln^3(4e(n+1))}\left(\tfrac{4(n+1)}{\ln(4e(n+1))}\right)^{n-\frac{n}{\ln(n)}} \\
&\quad \text{Then, we use } \lim_{n\to+\infty}\frac{\ln(n)}{\ln(4e(n+1))} = 1 \\
&\geq C\tfrac{\sqrt{n}}{\ln^3(n)}\left(\tfrac{4(n+1)}{\ln(4e(n+1))}\right)^{n-\frac{n}{\ln(n)}} \\
&\quad \forall\epsilon > 0,\ \lim_{n\to+\infty}\left(\frac{n+1}{n}\right)^n = e \text{ gives:} \\
&\geq C\tfrac{\sqrt{n}}{\ln^3(n)}\left(\tfrac{4n}{\ln(4e(n+1))}\right)^{n-\frac{n}{\ln(n)}} \\
\forall\epsilon > 0,\ \lim_{n\to+\infty}&\left(\frac{4\ln(n)}{(4-\epsilon)\ln(4e(n+1))}\right)^n = +\infty \text{ gives:} \\
&\geq C\tfrac{\sqrt{n}}{\ln^3(n)}\left(\tfrac{(4-\epsilon)n}{\ln(n)}\right)^{n-\frac{n}{\ln(n)}}
\end{aligned}$$

This ends the proof because $\lim_{n\to+\infty}\frac{\sqrt{n}}{\ln^3(n)} = +\infty$.

8.4 Proof of theorem 10

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, here, we use Theorem 13 and that the proportion of unary nodes over binary nodes in such trees is far from the typical proportion in ordinary unary-binary trees which tends to some nonzero constant. We have

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

where

- $C(n - \frac{n}{3\ln(n))})$ corresponds to the binary structure (which has at most $n - \frac{n}{3\ln(n))}$ binary nodes).
- $\binom{2n+1}{3\frac{n}{\ln(n)}}$ is an upper bound the number of possible distributions of unary nodes within binary structure. This is because, in general, the number of possible ways of inserting q unary nodes in a tree of size p is $\binom{p+q}{q}$ (including leaves in the size and the order of insertion having no importance). Here (since the tree has n inner nodes and $n+1$ leaves) we have $p+q \leq 2n+1$ and $q \leq \frac{3n}{\ln(n)}$. We also need to remark that $\frac{3n}{\ln(n)} < \frac{2n+1}{2}$ for n large enough.
- $\left(\frac{3n}{\ln(n)}\right)^{n+1-\frac{n}{3\ln(n)}}$ corresponds to the possibilities of bindings. Indeed, $\frac{3n}{\ln(n)}$ is an upper bound for the number of λ's above a variable and $n+1-\frac{n}{3\ln(n)}$ is an upper bound for the number of leafs.

Lemma 35 and Proposition 5 allow to conclude. Note that the replacement of 4^n by $(4+\varepsilon)^n$ compensates all factors smaller than exponential.

8.5 End of proof of Theorem 13

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)

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

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

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

$$= \left(\frac{k}{4(3-2\sqrt{2})}\right)^n \left(\frac{kn}{\ln(n)}\right)^{\frac{-kn}{\ln(n)}} \left(\frac{qn}{\ln(n)}\right)^{\frac{n}{\ln(n)}}$$

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 $\left(\frac{qn}{\ln(n)}\right)^{\frac{n}{\ln(n)}} = e^n \left(\frac{q}{\ln(n)}\right)^{\frac{n}{\ln(n)}}$, 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, by Proposition 7.(1) 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 $q = 4-\epsilon$ with $\epsilon > 0$), 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 $\epsilon = 4-q$ can be chosen small enough.

(2) 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 $\frac{1}{3}$.

8.6 End of proof of Theorem 14

$$\begin{aligned}
\Psi(n,q) &= M(n)\frac{\left(\frac{n}{K\ln(n)}\right)^{n-\frac{n}{3\ln(n)}+1}}{\left(\frac{qn}{\ln n}\right)^{n-\frac{n}{\ln(n)}}} \\
&\quad \text{we compensate the exponent } +1 \text{ with the } n^{-\frac{3}{2}} \\
&\quad \text{in the asymptotic for } M(n) \\
&\leq \left(\frac{1}{qK(3-2\sqrt{2})}\right)^n (qK^{\frac{1}{3}})^{\frac{n}{\ln(n)}} \left(\frac{n}{\ln(n)}\right)^{\frac{2n}{3\ln(n)}} \\
&\quad \text{using } n^{\frac{n}{\ln(n)}} = e^n, \text{ we find:} \\
&\leq \left(\frac{e^{\frac{2}{3}}}{qK(3-2\sqrt{2})}\right)^n (qK^{\frac{1}{3}})^{\frac{n}{\ln(n)}} \left(\frac{1}{\ln(n)}\right)^{\frac{2n}{3\ln(n)}}
\end{aligned}$$

Thus it is clear that if $\frac{e^{\frac{2}{3}}}{(4-\epsilon)K(3-2\sqrt{2})} < 1$, that is if $K > \frac{e^{\frac{2}{3}}}{(4-\epsilon)(3-2\sqrt{2})} \simeq 2.84 < 3$, then we have:

$$\lim_{n\to+\infty} \Psi(n, 4-\epsilon) = 0.$$

8.7 Proof of theorem 15

We will use the following theorem.

Theorem 36. *Let $C_n^{[>h]}$ denote the number of binary trees of size n with height greater than h. For $1 \leq h \leq n$ we have:*

$$C_n^{[>h]} = O(C_n n^{3/2} e^{-h^2/(4n)})$$

Proof. It is a part of the Theorem 1.3 from [3]. Although, as the authors say, this upper bound is rather poor for big heights, it is sufficient for our needs. □

Let A_n be the number of terms of size n having at most $3\frac{n}{\ln(n)}$ λ's and having binary-height greater than $\frac{n}{\sqrt{\ln(\ln(n))}}$. Let $u_n = \frac{n+1}{W(e(n+1))}$ and $b_n = n + 1 - u_n$. By Proposition 7.(1), the pair (u_n, b_n) gives, for fixed n, the "optimal" proportion between numbers of leaves and of unary nodes. Then,

$$A_n \leq C_n^{[>h]} \binom{2n+1}{3\frac{n}{\ln(n)}} u_n^{b_n},$$

where

- $C_n^{[>h]}$ bounds the number of all possible binary structures of terms from A.
- $\binom{2n+1}{3\frac{n}{\ln(n)}}$ is justified as in the proof of theorem 10.
- $u_n^{b_n}$ bounds the number of all possible variable bindings.

Using our lower bound for L_n and Theorem 36 it remains to show that the following quotient tends to zero:

$$\frac{C(n)n^{3/2}e^{-h^2/(4n)}\binom{2n+1}{3\frac{n}{\ln(n)}}u_n^{b_n}}{\left(\frac{(4-\varepsilon)n}{\ln(n)}\right)^{n-\frac{n}{\ln(n)}}}$$

This is as follows. By Proposition 7.(2), $u_n \sim \frac{n}{\ln(n)}$. In fact we use only that $u_n + 1 \leq \frac{3n}{\ln(n)}$ for n big enough.

Since we know that $n - b_n = u_n + 1 \sim \frac{n}{\ln(n)}$, we can use $\frac{3n}{\ln(n)}$ as an asymptotic upper bound for it, we get

$$\lim_{n\to\infty} \frac{A_n}{L_n} \leq \lim_{n\to\infty} cn^2 \left(4^{\frac{3n}{\ln(n)}} e^{-\left(\frac{n}{4\ln(\ln(n))}\right)+\frac{4n\ln(\ln(n))}{\ln(n)}}\right)$$

$$= \lim_{n\to\infty} c \cdot e^{\left(2\ln(n)+\ln(4)\frac{3n}{\ln(n)}-\frac{n}{4\ln(\ln(n))}+\frac{4n\ln(\ln(n))}{\ln(n)}\right)}$$

It is easy to verify that the term $-\frac{n}{4\ln(\ln(n))}$ asymptotically dominates all the other terms in the exponent. It means that the exponent goes to $-\infty$, as n increases, and consequently

$$\lim_{n\to\infty} \frac{A_n}{L_n} = 0.$$

8.8 Proof of $d(SN) \geq \frac{1}{2}$

Proposition 37. *If the set of unsafe terms has a density, this density is less than* $\frac{1}{2}$.

Proof. By definition 27, write $t = A[(L[\lambda x_1.a_1]\ B[\lambda x_2.a_2])]$ where x_i has at least one occurrence in a_i (we only use a weak version of the definition since we do not ask for at least two occurrences). See Figure 9. We assume that, each time there is a possible choice for a pattern, we choose the leftmost deepest occurrence of the pattern. By taking the highest binding λ in $L[\lambda x_1.a_1]$ we may assume that in the branch of L from the root to the node λx_1 there is no binding λ.

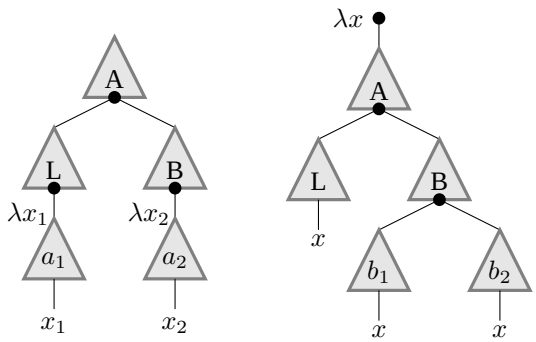

Figure 9. The terms t and $\varphi(t)$.

Let $\varphi(t) = \lambda x.A[(L[x]\ B[(b_1\ b_2)])]$ where $b_i = a_i[x_i := x]$ (see Figure 9). Note that $size(\varphi(t)) = size(t)$ and that, since the branch of L to $\lambda x_1.a_1$ has no binding λ, $\varphi(t)$, no variable become bound or free, during the encoding. Since the occurrence of x in $L[x]$ is the leftmost deepest occurrence of x in $\varphi(t)$ and since the node $(b_1\ b_2)$ is the leftmost deepest node having all the occurrences of x in t, φ is injective.

Let $\psi(t) = \lambda x.A[(B[(b_1\ b_2)]\ L[x])]$. For the same reasons as before ψ is injective. It is clear that the set of images of non safe terms by φ and ψ are disjoint and have the same cardinalities. This gives the desired result. □

8.9 Proof of Theorem 32

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 $\widetilde{f}$ and $\widetilde{f_{t_0}}$ so as to satisfy the following equations: $\widetilde{f}(\sqrt{1-z}) = \overline{f}(z)$ and $\widetilde{f_{t_0}}(\sqrt{1-z}) = \overline{f_{t_0}}(z)$. Functions $\widetilde{f}$ and $\widetilde{f_{t_0}}$ are

defined in the following way:

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

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

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

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

8.10 Proof of Proposition 33

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 $n_0 = size(t_0)$. 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.