Asymptotically almost all λ-terms are strongly normalizing

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

Asymptotically almost all λ-terms are strongly normalizing *

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

September 27, 2010

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 is 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 has been introduced. It turns out that they are all equivalent in sense of computational power. However, this equivalence says nothing about what *typical* programs or machines of each of these models do.

*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

†LAMA, Université de Savoie, 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

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 [10, 12, 16] 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 [19]). 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 5) 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 37). Moreover, each of them binds "many" occurrences of variables (Theorems 39, 41 and 43). Finally, given any fixed closed λ-term, almost no λ-term has this term as a subterm (Theorem 47).

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 asymptotic 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 subterm 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 8.

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 [17]. 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 [11]. For the results with more then one variable, and other logical connectives consult [7],[9],[8].

2 λ-calculus and combinatory logic

2.1 λ-calculus

We start with presenting some fundamental concepts of the λ-calculus, as well as with some new definitions used in this paper.

Definition 1. Let V be a countable set of variables. The set $\overline{\Lambda}$ of λ-terms is defined by the following grammar:

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

We denote by Λ the set of all closed λ-terms. We write $t_1 \ t_2 \ldots t_n$ without parenthesis for $(\ldots (t_1 \ t_2) \ldots t_n)$.

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.

Let us observe that λ-terms can be seen as rooted unary-binary trees.

Definition 2. By lambda tree we mean the following rooted tree. There are two kinds of internal nodes labeled by @ and by λ. Nodes labeled by @ have two successors left and right. Nodes labeled with λ have only one successor. Leaves of the tree are either labeled by variables or are connected with the one of λ nodes above it.

Definition 3. With every lambda term t we associate the lambda tree $G(t)$ in the following way:

- If x is a variable then $G(x)$ is a single node labeled with x.
- Tree $G(PQ)$ is a tree with the new root labeled with @ and two subtrees left $G(P)$ and right $G(Q)$.
- Tree $G(\lambda x.P)$ is obtained from $G(P)$ by four steps:
  - Add new root node labeled with λ
  - Connect new root with $G(P)$
  - Connect all leaves of $G(P)$ labeled with x with the new root.
  - Remove all labels x.

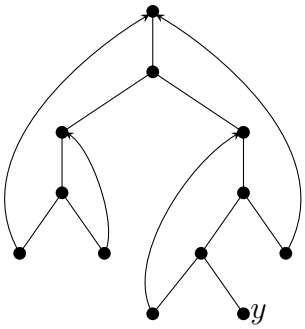

Figure 1: The lambda tree representing the term $\lambda z.(\lambda u.zu)((\lambda u.uy)z)$

Observation 4. *If T is a lambda tree then $T = G(M)$ for some lambda term M. Terms M and N are α convertible iff $G(M) = G(N)$.*

We often use (without giving the precise definition) the classical terminology about trees (e.g. path, root, leaf, etc.). A path from the root to a leaf is called a branch.

Definition 5. Let t be a λ-term.

1. A term t' is a *subterm* of t (denoted as $t' \leq t$) if
   - either $t = t'$,
   - or $t = \lambda x.u$ and $t' \leq u$,
   - or $t = (u\ v)$ and ($t' \leq u$ or $t' \leq v$).
2. Let $u = \lambda x.a$ be a subterm 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 in lambda tree of t.
4. Two λ's in t are called *incomparable* if there is no branch in the lambda tree containing both of them. The λ-*width* of t is the maximal number of pairwise incomparable binding λ's.
5. We say that t has k *head* λ's if its lambda tree starts with k unary nodes.

Definition 6.
- A term of the form $(\lambda x.P)Q$ is called a β redex. A lambda term is in normal form if it does not contain β redex sub-terms. The least relation $\triangleright$ on terms satisfying $(\lambda x.P)Q \triangleright P[x := Q]$ and closed by context is called β reduction.
- A term M is (weakly) normalizing if there is reduction sequence starting from M and ending in a normal form N.

- A term M is strongly normalizing if all reduction sequences are finite. By SN we mean all terms which are strongly normalizing. We denote by $\eta(M)$ the length of the longest reduction starting from M if M is SN. The fact that such a longest reduction exists follows from Konig's lemma. If M is not SN, $\eta(M) = +\infty$.

As an example we can see lambda tree representation of redex which is a subterm of some lambda tree. Therefore β conversion can be seen as a operation on lambda trees.

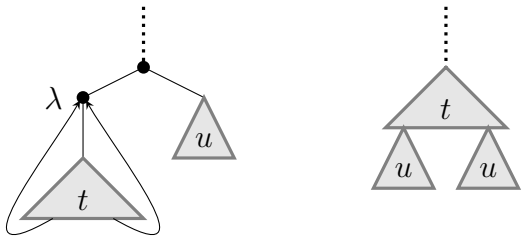

Figure 2: β-reduction scheme

Definition 7. The size of a term (denoted by size$(\cdot)$) is defined recursively as follows:

(i) size$(x) = 0$, if x is a variable,

(ii) size$(\lambda x.t) = 1 + \text{size}(t)$,

(iii) size$(t\ u) = 1 + \text{size}(t) + \text{size}(u)$.

As we can see, $size(t)$ is the number of internal nodes of lambda tree $G(t)$.

Notation 8. 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 its cardinality by L_n.

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

2.2 Fair and safe λ terms

This part is devoted to Proposition 18 and concerns only λ-calculus. Nevertheless, for Fact 11 see [1] and for similar proof techniques, see [4, 5].

Definition 9.

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 subterm 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 terms u and v is fair.

Definition 10.

- A substitution σ is a partial map from variables to terms such that the domain of σ is finite. Let t be a term and let σ be a substitution such that all variables from the domain of σ are free in t. By $t[\sigma]$ we denote the term obtained from t by replacing all free occurrences of variables x by $\sigma(x)$.

- A context is a λ-term with a unique hole denoted []. Traditionally, contexts are defined by a BNF grammar. If E is an arbitrary context, it is given by the following BNF grammar:

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

- When E is a context and t is a term, $E[t]$ denotes the result of 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.

- In a few cases, we need contexts with multiple holes. In this case, we write $E[t_1, \ldots, t_n]$ when E is a context with exactly n holes and $E[t_1, \ldots, t_n]$ denotes the term where the holes of E are substituted from the leftmost to the rightmost by $t_1, \ldots, t_n$ in this order.

In some proofs in this section we use the following basic fact on lambda terms and strong normalisation:

Fact 11. *A lambda term can be written in one of the following forms:*

- $t = (x\ t_1\ \ldots\ t_n)$ *with* $n \geq 0$. *In this case* $\eta(t) = \eta(t_1) + \cdots + \eta(t_n)$ *and* t *is SN if and only if* $t_1, \ldots, t_n$ *are SN.*

- $t = \lambda x.u$. *In this case* $\eta(t) = \eta(u)$ *and* t *is SN if and only if* u *is SN.*

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

Remark

The previous fact is well known (see for example [1]). Also note that, if t is a term and x is a variable, then t is SN if and only if $(t\ x)$ is SN.

Lemma 12. *The set of terms of width at most* 1 *is closed under* β*-reduction.*

Proof. If a term is of width 0, then no substitution or reduction can change the width, since all variables in the term are free.

Let t be a term be of width 1. First, let us remark that all binding λ's in t occur on the same branch. We consider 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 variables of v are free in t or bound by the context E. In both cases, it is clear the t' is still of width 1 because the binding λ's remain on one branch. □

Lemma 13. *If* t *is a term of lambda width at most* 1*, then* t *is SN.*

Proof. Let $N_0(t)$ and $N_1(t)$ denote the numbers of, respectively, non-binding and binding lambdas in term t. Let us introduce the lexicographic order on pairs $\langle N_1(t), N_0(t)\rangle$. Let t be of width at most 1. Then, performing a β-reduction on t decreases the pair $\langle N_1(t), N_0(t)\rangle$ while keeping the width at most 1 by Lemma 12. To prove this, we consider a β-reduction: $t = E[(\lambda x.u)\ v] \triangleright E[u[x := v]] = t'$. If x is not bound in u, then $N_1(t)$ 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, then v is of width 0 and contains no binding λ, which means that we erase one binding λ and only duplicate non-binding λ's. Therefore, $N_1(t)$ is decreasing. □

Lemma 14. *If* u *has width* 0 *and* $t_1, \ldots, t_n$ *are SN terms, then the term* $t = (u\ t_1\ \ldots\ t_n)$ *is SN.*

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

- If $u = x$, the result is trivial by Fact 11.
- If $u = (u'\ v)$, v has width 0 and is SN because of Lemma 13. We conclude by induction on u'.
- If $u = \lambda x.u'$. If $n = 0$, the result follows from Lemma 13. Otherwise, by Fact 11, it is enough to show that the head reduct of t is SN. But, since u has width 0, this reduct is $(u'\ t_2\ \ldots\ t_n)$ and the result follows by induction. □

Lemma 15. *Let* $t \in SN$ *be a term and* σ *be a substitution such that, for each* x*, there is* k *such that* $\sigma(x) = (u\ v_1\ \ldots\ v_k)$ *where* u *has width* 0 *and* $v_1 \ldots v_k$ *are SN. Then,* $t[\sigma] \in SN$*.*

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

- If $t = \lambda x.t_1$ or if $t = (x\ t_1\ \ldots\ t_n)$ with x not in the domain of σ. In this cases, 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 $\mathrm{size}(t_i) < \mathrm{size}(t)$.

- If $t = ((\lambda x.u)\, v\, t_1\ \ldots\ t_n)$ we show that $v[\sigma]$ and $(u[x := v]\, t_1\ \ldots\ t_n)[\sigma]$ are SN. This follows from the induction hypothesis because $\eta(v) \leq \eta(t)$ and $size(v) < size(t)$ for the first point and because $\eta(u[x := v]\, t_1\ \ldots\ t_n) < \eta(t)$ for the second.

- If $t = (x\, t_1\ \ldots\ t_n)$ where x is in the domain of σ. Then, we have $t[\sigma] = (\sigma(x)\, t_1[\sigma]\ \ldots\ t_n[\sigma])$ which is SN by Lemma 14 because $t_1[\sigma], \ldots, t_n[\sigma]$ are SN by induction hypothesis and $\sigma(x) = (u\, v_1\ \ldots\ v_k)$ where u has width 0 and $v_1 \ldots v_k$ are SN. □

Definition 16. We define the set of contexts of width 1 by the following BNF grammar (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 17. *Let E be a context of width 1 and $u \in SN$ be a term. Then $E[u] \in SN$.*

Proof. By induction on size(E). Cases $E = []$ or $E = \lambda x.E_1$ are trivial (in the second case, since size(E_1) < size(E), the proof goes by induction).

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 size(E_1) < size(E). Therefore $(E_1[u]\ x)$ is SN by Fact 11 and finally $(E_1[u]\ x)[x := v]$ is SN by Lemma 15.

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

Proposition 18. *Let t be a safe term of width at most 2. Then t is SN.*

Proof. If t has width at most one, the result follows directly from Lemma 13. If t has width 2, let $(t_1\ t_2)$ be the smallest subterm of t of width 2. This means that t can be written $E[(t_1\ t_2)]$ where E is a context of width at most 1 and t_1 and t_2 are both of width 1. By Lemma 17, 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 defined by the following BNF grammar:

$F := [] \mid \lambda\_.F \mid (F\ \Lambda_0)$ where $\lambda\_$ denotes non-binding λ's and Λ_0 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 together with the definition 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, $(t_1\ t_2) = t'[x := (u\ v)]$, because the context F cannot bind variables. Thus, we can conclude by Lemma 15 because u has width 0 and because t' and v are SN (by Lemma 13, since they have width 1). □

2.3 Combinatory logic

Combinatory logic is a theoretical model of computation introduced by Moses Schönfinkel in [14] and many years later rediscovered and deeply studied by Haskell Curry in [3]. For the main reference on the subject we propose Barendregt [1]. A very intelligible approach towards this subject can be found in [15]. It is a well known fact that both models, the lambda calculus and the combinatory logic, are equivalent in sense of expression power. It turns out, however, that those two models differ radically as concerns the behavior of random terms.

Definition 19. Combinatory logic

1. The set $\mathcal{F}$ of combinatory logic terms, *combinators*, is defined by the following grammar:
$$\mathcal{F} := K \mid S \mid I \mid (\mathcal{F}\ \mathcal{F}).$$
The notational conventions concerning parentheses are the same as for lambda terms that is $t_1\ t_2 \dots t_n$ without parenthesis for $(\dots(t_1\ t_2)\dots t_n)$.

2. The *reduction* on combinators is performed according to the following rules:
$$K\ u\ v \rhd u \qquad S\ u\ v\ w \rhd u\ w\ (v\ w) \qquad I\ u \rhd u.$$

Let us observe that combinatory logic term can be seen as rooted binary tree. Leaves are labeled by combinators K, S and I and binary internal nodes are labeled by an application operation. Therefore every combinatory logic term t can be associated uniquely with the combinatory logic tree $G(t)$. Accordingly every reduction rule can be seen as a transformation of combinatory logic tree.

Definition 20. The least relation $\rhd$ on combinatory logic terms satisfying reduction rules is called reduction. A combinatory logic term is in normal form if no reduction can be performed. A term M is normalizing if there is reduction sequence starting from M and ending in a normal form N. A term M is strongly normalizing if all reduction sequences are finite. By SN we mean all terms which are strongly normalizing.

Definition 21. Subterm and size

1. A combinator u is a *subterm* of v if either $u = v$ or v is of the form $v_1\ v_2$ and u is a subterm of v_1 or v_2.

2. The *size* of a combinator is defined by the following rules:

$$size(S) = size(K) = size(I) = 0 \quad \text{and} \quad size(u\ v) = 1 + size(u) + size(v).$$

As we can see $size(t)$ is the number of internal nodes of combinatory logic tree $G(t)$.

Notation 22. Let n be an integer. We denote by $\mathcal{F}_n$ the set of combinatory logic terms of size n. Obviously the set $\mathcal{F}_n$ is finite. We denote its cardinality by F_n.

3 Combinatorial results

The following standard notions will be used throughout the whole paper.

Definition 23. Let $f, g\colon \mathbb{N} \to \mathbb{R}$.

(i) Functions f and g are said to be *asymptotically equal* iff $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$. We denote it by $f \sim g$.

(ii) The *asymptotic inequality* $f \gtrsim g$ holds iff there exists a function $h\colon \mathbb{N} \to \mathbb{R}$ such that $h \sim g$ and $f(n) \geq h(n)$ for any n.

(iii) A function f is said to be of the *smaller order* than g iff $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$. We denote it by $f \in o(g)$.

(iv) A function f is said to be *subexponential* in n iff there exists $h\colon \mathbb{N} \to \mathbb{R}$ such that $h \in o(n)$ and $f(n) = 2^{h(n)}$.

(v) If x is a real number we denote by $\lfloor x \rfloor$ (resp. $\lceil x \rceil$) the largest (resp. smallest) integer n such that $n \leq x$ (resp. $x \leq n$)

Notation

When an unknown function f is, for example, asymptotically equal to an explicit function (say for example $n \ln(n)$) we will write $f \sim n \ln(n)$ or sometimes $f(n) \sim n \ln(n)$.

3.1 Generating function method

Many questions concerning the asymptotic behavior of sequences of real positive numbers can be efficiently resolved by analyzing the behavior of the generating function of the sequence. This is the approach we take to determine the asymptotic fraction of certain combinatory logic trees of the given size.

The main tool for finding limits of the fraction $\frac{a_n}{b_n}$, when generating functions for sequences a_n and b_n satisfy certain conditions of simplified Szegö Lemma can

be found in Theorem 22 of [18]. From this theorem we can derive techniques summarized in Theorem 24 which is useful for finding limits of the proportion between two sequences of known generating functions. In this part of the section we present the method of finding asymptotic densities for the classes of lambda trees or combinatory logic trees for which the generating functions are already calculated. Also the generating functions method will be used to calculate asymptotics in Proposition 26. The next theorem is a well-known result in the theory of generating functions. Its proof can be found e.g. in [18]. We denote by $[z^n]\{v(z)\}$ the coefficient of z^n in the expansion of v.

Theorem 24. *Let v, w be functions satisfying the following conditions:*

(i) *v, w are analytic in $|z| < 1$ with $z = 1$ being the only singularity on the circle $|z| = 1$,*

(ii) *v, w have the following expansions in the vicinity of $z = 1$:*

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

where $w_1 \neq 0$.

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

$$\lim_{n\to\infty} \frac{[z^n]\{v(z)\}}{[z^n]\{w(z)\}} = \frac{v_1}{w_1} = \frac{(\widetilde{v})'(0)}{(\widetilde{w})'(0)}.$$

3.2 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 classical result (see, for example, [6]).

Proposition 25. *$C(n) \sim \frac{4^n}{n^{3/2}\sqrt{\pi}}$ and thus, for n large enough, we have $C(n) \geq \gamma\frac{4^n}{n^{3/2}}$ for some constant $0 < \gamma < 1$.*

3.3 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 inner and outer nodes. We use the following proposition.

Proposition 26. *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 asymptotics 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 [6]. □

4 Densities

4.1 Main notations

To attribute a precise meaning to sentences like "asymptotically all lambda terms have a 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 27. Let $B \subset \Lambda$ be a set containing terms of every size. For $A \subseteq B$ if the following limit

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

exists, then we call it the asymptotic density of A in B and denote it by $d_B(A)$.

Remarks and notations

- The number $d_B(A)$ if it exists is an asymptotic probability of finding a lambda term from the class A among all lambda terms from B or it can be interpreted as the asymptotic density of the set A in the set B.
- It can be seen immediately that the density d_B is finitely additive so if A_1 and A_2 are disjoint classes of lambda terms such that $d_B(A_1)$ and $d_B(A_2)$ exist then $d_B(A_1 \cup A_2)$ also exists and $d_B(A_1 \cup A_2) = d_B(A_1) + d_B(A_2)$.
- It is straightforward to observe that for any infinite B and finite set A the density $d_B(A)$ exists and is 0. Dually for co-finite sets A the density $d_B(A) = 1$.
- The density d_B is not countably additive so in general the formula

$$d_B\left(\bigcup_{i=0}^{\infty} A_i\right) = \sum_{i=0}^{\infty} d_B\left(A_i\right)$$

is not true for all pairwise disjoint classes of sets $\{A_i\}_{i\in\mathbb{N}}$. A good counterexample for the equation is to take $B = \Lambda$ and A_i the singleton of i-th lambda term from our language under any natural order of terms. On the left hand side of the equation we get $d_\Lambda(\Lambda)$ which is 1 but on right hand side $d_\Lambda(A_i) = 0$ for all $i \in \mathbb{N}$ and so the sum is 0.

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

5 Proofs using calculus

In this section we state a few theorems which provide bounds for L_n (the number of lambda terms of size n). We also find a lower bound for the unary height in a random term.

5.1 Lower bound for L_n

The estimation for L_n we provide is very rough but sufficient in further considerations.

Theorem 28. *For any $\varepsilon \in (0, 4)$ we have*

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

Proof. Let $LB(n, k)$ denote 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, $L_n > LB(n, k)$ for all $k = 1, \ldots, n$. Let $k = \left\lceil \frac{n}{\ln(n)} \right\rceil$. Then we get:

$$\begin{aligned}
L_n &> C\left(n - \left\lceil \frac{n}{\ln(n)} \right\rceil\right) \cdot \left(\left\lceil \frac{n}{\ln(n)} \right\rceil\right)^{n-\left\lceil \frac{n}{\ln(n)} \right\rceil - 1} \\
&\sim \frac{4^{n-\frac{n}{\ln(n)}}}{\left(n - \frac{n}{\ln(n)}\right)^{3/2}\sqrt{\pi}} \cdot \left(\left\lceil \frac{n}{\ln(n)} \right\rceil\right)^{n-\left\lceil \frac{n}{\ln(n)} \right\rceil - 1} && \text{by Proposition 25} \\
&\gtrsim \left(\frac{4 \cdot n}{\ln(n)}\right)^{n-\frac{n}{\ln(n)}} \cdot \frac{1}{p(n)} && \text{for some positive polynomial } p \\
&\gtrsim \left(\frac{(4-\varepsilon)n}{\ln(n)}\right)^{n-\frac{n}{\ln(n)}} && \text{since } \left(\frac{4}{4-\varepsilon}\right)^{n-\frac{n}{\ln n}} \gtrsim p(n).
\end{aligned}$$

□

5.2 Number of λ's in a term

In this part we focus on the number of unary and binary nodes in random lambda terms. We need the following lemma:

Lemma 29. *Let $n > 1$. The function $f(p) = p^{n-p+1}$ is*

(i) decreasing on $[\frac{3n}{\ln(n)}; n]$,

(ii) increasing on $[0; \frac{n}{3\ln(n)}]$.

Proof. Let us start with computing the derivative of the function f:

$$f'(p) = \left(p^{n-p+1}\right)' = \left(e^{(n-p+1)\ln(p)}\right)' = e^{(n-p+1)\ln(p)}\left(\frac{n-p+1}{p} - \ln(p)\right).$$

(i) We want to show that $f'(p) < 0$ for any $p \in \left[\frac{3n}{\ln(n)}, n\right]$. This is equivalent to the following inequality: $n + 1 < p(\ln(p) + 1)$. The expression on the right reaches the minimum in the considered interval at $p = \frac{3n}{\ln(n)}$, thus it is sufficient to prove that

$$n + 1 < \frac{3n}{\ln(n)}\left(\ln\left(\frac{3n}{\ln(n)}\right) + 1\right).$$

But the right expression is equal to

$$\begin{aligned}\frac{3n}{\ln(n)}&(\ln(n) - \ln(\ln(n)) + \ln 3 + 1) \\ &= 2n + \frac{n}{\ln(n)}(\ln(n) - 3\ln(\ln(n)) + 3\ln 3 + 3) \\ &> n + 1,\end{aligned}$$

which finishes the proof.

(ii) We want to show that $f'(p) > 0$ for any $p \in \left[0, \frac{n}{3\ln(n)}\right]$. This is equivalent to the following inequality: $n + 1 > p(\ln(p) + 1)$. The expression on the right reaches the maximum in the considered interval at $p = \frac{n}{3\ln(n)}$, thus it is sufficient to prove that

$$n + 1 > \frac{n}{3\ln(n)}\left(\ln\left(\frac{n}{3\ln(n)}\right) + 1\right).$$

But the right expression is equal to

$$\begin{aligned}\tfrac{n}{3\ln(n)}(\ln(n) - \ln(\ln(n)) - \ln 3 + 1) &= n - \tfrac{n}{3\ln(n)}(\ln(\ln(n)) + 2\ln 3 - 1) \\ &< n + 1,\end{aligned}$$

which finishes the proof. □

The next theorem shows that the typical proportion of unary nodes over binary ones in lambda terms is far from the typical proportion in ordinary unary-binary trees, in which case it tends to a positive constant.

Notation 30. Let $\mathcal{A}$ denote the class of closed terms $t \in \mathcal{A}$ that satisfy all the following conditions:

(i) the number of lambdas in t is at most $\frac{3\text{size}(t)}{\ln(\text{size}(t))}$,

(ii) the number of lambdas in t is at least $\frac{\text{size}(t)}{3\ln(\text{size}(t))}$,

(iii) the unary height of t is at least $\frac{\text{size}(t)}{3\ln(\text{size}(t))}$.

Theorem 31. *The density of $\mathcal{A}$ in Λ is 1.*

Proof. Let us consider terms of size n with exactly p lambdas. Such terms have exactly $n-p+1$ leaves and each of them can be bound by at most p lambdas. Since the number of unary-binary trees of size n and with $n-p+1$ leaves is equal to $M(n, n-p+1)$ (see 3.3), we obtain the following upper bound for considered terms: $M(n, n-p+1) \cdot p^{n-p+1}$.

Now, we show that each of properties (i)–(iii) characterizing the class $\mathcal{A}$ is valid for random terms.

(i) Let P_n denote the number of terms of size n containing more than $\frac{3n}{\ln(n)}$ λ's. We have $P_n \leq \sum_{p \geq \frac{3n}{\ln(n)}} M(n, n-p+1) \cdot p^{n-p+1}$.

By Lemma 29 the function $p \mapsto p^{n-p+1}$ is decreasing in the interval $\left[\frac{3n}{\ln(n)}, n\right]$. Thus,

$$P_n \leq \sum_{p \geq \frac{3n}{\ln(n)}} M(n, n-p+1) \cdot \left(\frac{3n}{\ln(n)}\right)^{n+1-\frac{3n}{\ln(n)}} \leq M(n) \left(\frac{3n}{\ln(n)}\right)^{n+1-\frac{3n}{\ln(n)}}.$$

By the lower bound for L_n from 5.1 and the computations above, we get

$$\frac{P_n}{L_n} \lesssim \frac{M(n) \left(\frac{3n}{\ln(n)}\right)^{n+1-\frac{3n}{\ln(n)}}}{\left(\frac{(4-\varepsilon)n}{\ln(n)}\right)^{n-\frac{n}{\ln(n)}}}.$$

To get the result it remains to show that for some $\varepsilon \in (0, 4)$ this expression tends to 0.

By Proposition 26, $M(n) \sim \left(\frac{1}{3-2\sqrt{2}}\right)^n \frac{1}{\sqrt{\pi} n^{\frac{3}{2}}}$. Then for n large enough and for some constant $\gamma > 1$, we obtain

$$\begin{aligned}
\frac{P_n}{L_n} &\lesssim \gamma \cdot \frac{\left(\frac{1}{3-2\sqrt{2}}\right)^n \left(\frac{3n}{\ln(n)}\right)^{n+1-\frac{3n}{\ln(n)}}}{n^{\frac{3}{2}} \left(\frac{(4-\varepsilon)n}{\ln(n)}\right)^{n-\frac{n}{\ln(n)}}} \\
&\leq \frac{\left(\frac{1}{3-2\sqrt{2}}\right)^n \left(\frac{3n}{\ln(n)}\right)^{n-\frac{3n}{\ln(n)}}}{\left(\frac{(4-\varepsilon)n}{\ln(n)}\right)^{n-\frac{n}{\ln(n)}}} \qquad \text{since } \frac{3\gamma n}{\ln(n)} < n^{\frac{3}{2}} \\
&= \left(\frac{3}{(4-\varepsilon)(3-2\sqrt{2})}\right)^n \left(\frac{3n}{\ln(n)}\right)^{\frac{-3n}{\ln(n)}} \left(\frac{(4-\varepsilon)n}{\ln(n)}\right)^{\frac{n}{\ln(n)}} \\
&= \left(\frac{3}{(4-\varepsilon)(3-2\sqrt{2})}\right)^n \left(\frac{3^{-3}(4-\varepsilon)\ln^2(n)}{n^2}\right)^{\frac{n}{\ln(n)}}
\end{aligned}$$

Notice that for any α, $(n^{2-\alpha})^{n/\ln(n)} = e^{\ln(n)(2-\alpha)\frac{n}{\ln(n)}} = e^{(2-\alpha)n}$. Thus, we obtain

$$\frac{P_n}{L_n} \lesssim \left(\frac{3}{(4-\varepsilon)(3-2\sqrt{2})e^{2-\alpha}}\right)^n \left(3^{-3}(4-\varepsilon)\frac{\ln^2(n)}{n^\alpha}\right)^{\frac{n}{\ln(n)}}.$$

Let α and ε be positive and small enough so that $3 < (4-\varepsilon)(3-2\sqrt{2})e^{2-\alpha}$. Then the whole expression tends to 0 as n tends to infinity.

(ii) Let R_n denote the number of terms of size n containing less than $\frac{n}{3\ln(n)}$ lambdas. We have $R_n \leq \sum_{p \leq \frac{n}{3\ln(n)}} M(n, n-p+1) \cdot p^{n-p+1}$.

By Lemma 29 the function $p \mapsto p^{n-p+1}$ is increasing in the interval $\left[0, \frac{n}{3\ln(n)}\right]$. Thus,

$$\begin{aligned}
R_n &\leq \sum_{p \leq \frac{n}{3\ln(n)}} M(n, n-p+1) \cdot \left(\frac{n}{3\ln(n)}\right)^{n+1-\frac{n}{3\ln(n)}} \\
&\leq M(n) \left(\frac{n}{3\ln(n)}\right)^{n+1-\frac{n}{3\ln(n)}}.
\end{aligned}$$

By the lower bound for L_n from 5.1 and the computations above, we get

$$\begin{aligned}
\frac{R_n}{L_n} &\lesssim \frac{M(n)\left(\frac{n}{3\ln(n)}\right)^{n+1-\frac{n}{3\ln(n)}}}{\left(\frac{(4-\varepsilon)n}{\ln(n)}\right)^{n-\frac{n}{\ln(n)}}} \\
&\leq \gamma \cdot \frac{\left(\frac{1}{3-2\sqrt{2}}\right)^n \left(\frac{n}{3\ln(n)}\right)^{n+1-\frac{n}{3\ln(n)}}}{n^{\frac{3}{2}}\left(\frac{(4-\varepsilon)n}{\ln(n)}\right)^{n-\frac{n}{\ln(n)}}} && \text{for some } \gamma > 1 \\
&\leq \frac{\left(\frac{1}{3-2\sqrt{2}}\right)^n \left(\frac{n}{3\ln(n)}\right)^{n-\frac{n}{3\ln(n)}}}{\left(\frac{(4-\varepsilon)n}{\ln(n)}\right)^{n-\frac{n}{\ln(n)}}} && \text{since } \frac{\gamma n}{3\ln(n)} < n^{\frac{3}{2}} \\
&= \left(\frac{1}{3(4-\varepsilon)(3-2\sqrt{2})}\right)^n \left(\frac{3(4-\varepsilon)^3 n^2}{(\ln(n))^2}\right)^{\frac{n}{3\ln(n)}} \\
&= \left(\frac{e^{2/3}}{3(4-\varepsilon)(3-2\sqrt{2})}\right)^n \left(\frac{3(4-\varepsilon)^3}{(\ln(n))^2}\right)^{\frac{n}{3\ln(n)}} && \text{since } n^{\frac{2n}{3\ln(n)}} = e^{\frac{2}{3}n}.
\end{aligned}$$

For $\varepsilon > 0$ small enough the whole expression tends to 0, what finishes the proof.

(iii) Let S_n be the number of terms of size n with more than $\frac{n}{3\ln(n)}$ lambdas and with the unary height less than $\frac{n}{3\ln(n)}$ λ's.

Such a term has at most $n - \frac{n}{3\ln(n)} + 1$ leaves and each of them can be bound by at most $\frac{n}{3\ln(n)}$ lambdas. Therefore, we have

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

Dividing it by the lower bound for L_n and performing exactly the same calculations as in the proof of (ii), we obtain the desired result. □

5.3 Upper bound for L_n

Now we are ready to provide the upper bound for L_n. Once again, this estimation is very rough, however, it turns out to be sufficient for our main goal.

Lemma 32. *The function $n \mapsto \binom{2n+1}{\frac{3n}{\ln(n)}}$ is subexponential.*

Proof. By applying the Stirling formula

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

we obtain, for some positive constant γ,

$$\begin{aligned}\binom{2n+1}{\frac{3n}{\ln(n)}} &= \frac{2n+1}{2n+1-\frac{3n}{\ln(n)}} \cdot \frac{(2n)!}{\left(2n-\frac{3n}{\ln(n)}\right)!\left(\frac{3n}{\ln(n)}\right)!} \\ &\lesssim \gamma\sqrt{\frac{\ln(n)}{n}}\frac{(2n)^{2n}}{\left(2n-\frac{3n}{\ln(n)}\right)^{2n-\frac{3n}{\ln(n)}}\left(\frac{3n}{\ln(n)}\right)^{\frac{3n}{\ln(n)}}} \\ &\lesssim \frac{2^{2n}}{\left(2-\frac{3}{\ln(n)}\right)^{2n-\frac{3n}{\ln(n)}}\left(\frac{3}{\ln(n)}\right)^{\frac{3n}{\ln(n)}}}\end{aligned}$$

Let us compute the logarithm of the expression above:

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

After some simplifications we obtain that $\ln\binom{2n+1}{\frac{3n}{\ln(n)}} \lesssim 3n\frac{\ln\ln(n)}{\ln(n)} + o\left(n\frac{\ln(\ln(n))}{\ln(n)}\right)$, so

$$\binom{2n+1}{\frac{3n}{\ln(n)}} \lesssim \delta e^{3n\frac{\ln\ln(n)}{\ln(n)}} \quad \text{for some } \delta > 0. \qquad \square$$

Theorem 33. *For any $\varepsilon > 0$ we have*

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

Proof. Let T_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. According to Theorem 31 we have $L_n \sim T_n$. In λ-terms enumerated by T_n the number of binary nodes is at most $n - \frac{n}{3\ln(n)}$ and the number of leaves is at most by one greater. We compute the upper bound for T_n in the following way:

- first, we consider binary trees built on at most $n - \frac{n}{3\ln(n)}$ binary nodes — their number does not exceed Catalan number $C\left(n - \left\lfloor\frac{n}{3\ln(n)}\right\rfloor + 1\right)$,
- then, we insert in such trees at most $\frac{3n}{\ln(n)}$ (the maximal number of lambdas) unary nodes — this can be done in less than $\binom{2n+1}{3\frac{n}{\ln(n)}}$ ways ($2n+1$ is an upper bound for number of possible places for insertions in a binary tree of size $n - \frac{n}{3\ln(n)} + 1$),
- finally, we have at most $n + 1 - \frac{n}{3\ln(n)}$ leaves in such trees and each of them can by bound by at most $\frac{3n}{\ln(n)}$ lambdas — thus the number of possible labelings is not greater than $\left(\frac{3n}{\ln(n)}\right)^{n+1-\frac{n}{3\ln(n)}}$.

Thus, we get

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

Using the asymptotic expansion of Catalan numbers (Proposition 25), we obtain

$$\begin{aligned} T_n &\lesssim \binom{2n+1}{\frac{3n}{\ln(n)}} \frac{4^{n-\frac{n}{3\ln(n)}+1}}{\sqrt{\pi}\left(n - \frac{n}{3\ln(n)} + 1\right)^{3/2}} \left(\frac{3n}{\ln(n)}\right)^{n+1-\frac{n}{3\ln(n)}} \\ &\lesssim \binom{2n+1}{\frac{3n}{\ln(n)}} \left(\frac{12n}{\ln(n)}\right)^{n-\frac{n}{3\ln(n)}} \\ &\lesssim \left(\frac{(12+\varepsilon)n}{\ln(n)}\right)^{n-\frac{n}{3\ln(n)}}. \end{aligned}$$

The last line follows from the fact that $\binom{2n+1}{\frac{3n}{\ln(n)}}$ is subexponential (by Lemma 32). □

5.4 Comparison between the lower and the upper bounds

The ratio between the lower and upper bounds, which is equal to

$$\left(\frac{4-\varepsilon}{12+\varepsilon}\right)^{n-\frac{n}{3\ln(n)}},$$

is exponential. But, since L_n is super-exponential itself, our estimations are not too bad.

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

Corollary 34. *For any $\varepsilon \in (0,4)$ and for n large enough*

$$\begin{aligned} n\ln(n) - n\ln(\ln(n)) + n\ln(4-\varepsilon) - n \\ &\leq \ln(L_n) \\ &\leq n\ln(n) - n\ln(\ln(n)) + n\ln(12+\varepsilon) - \frac{n}{3}. \end{aligned}$$

6 Proofs using coding

In this section we prove theorems about random lambda terms using the following scheme. First, we consider a set $\Lambda_n(\mathcal{P})$ of terms of size n satisfying some property $\mathcal{P}$ and we aim at proving that this property is not satisfied by random terms. Next, we define an injective and size-preserving function $\varphi_n^{\mathcal{P}} \colon \Lambda_n(\mathcal{P}) \to \Lambda_n$ (called a *coding*) such that its image has density 0 among all closed lambda terms.

Lemma 35. *Let A_n be a sequence of set of terms and P be a property. Let $(A_{n,i})_{i\in I_n}$ be a partition of A_n. Let B_n (resp. $B_{n,i}$) be the set of elements of A_n (resp. $A_{n,i}$) satisfying P. Let a_n (resp. b_n, $a_{n,i}$, $b_{n,i}$) be the cardinality of A_n (resp. B_n, $A_{n,i}$, $B_{n,i}$). Assume $\frac{b_{n,i}}{a_{n,i}}$ tends to 0 uniformly in i as n tends to infinity, formally:*

$$\forall \varepsilon > 0, \exists N, \forall n \geq N, \forall i \in I_n : \frac{b_{n,i}}{a_{n,i}} < \varepsilon.$$

Then $\frac{b_n}{a_n}$ tends to 0 as n tends to infinity.

Proof. Let $\varepsilon > 0$. Let N be the corresponding integer guaranteed by the uniform convergence and let n be any integer with $n \geq N$. We have:

$$\frac{b_n}{a_n} = \frac{\sum_{i\in I_n} b_{n,i}}{a_n} = \sum_{i\in I_n} \frac{b_{n,i}}{a_{n,i}} \cdot \frac{a_{n,i}}{a_n} \leq \sum_{i\in I_n} \varepsilon \cdot \frac{a_{n,i}}{a_n} \leq \varepsilon.$$

We have shown $\lim_{n\to\infty} \dfrac{b_n}{a_n} = 0$. □

6.1 The number of λ's in head position

We start with showing that a random term starts with a long chain of lambdas.

Notation 36. Let $g\colon \mathbb{N} \to \mathbb{N}$ be a function such that $g \in o\big(\sqrt{n/\ln(n)}\big)$. Let us define $\mathcal{B}^g$ as the class of terms t such that

1. $t \in \mathcal{A}$,

2. t has at least $g(\mathrm{size}(t))$ head lambdas.

Additionally, we denote by $\overline{\mathcal{B}^g} = \mathcal{A} \setminus \mathcal{B}^g$ the complement of the set $\mathcal{B}^g$ in $\mathcal{A}$ and by $\overline{\mathcal{B}^g_n}$ the set of terms from $\overline{\mathcal{B}^g}$ of size n.

Theorem 37. *Let $g\colon \mathbb{N} \to \mathbb{N}$ be a function such that $g \in o\big(\sqrt{n/\ln(n)}\big)$. The density of $\mathcal{B}^g$ in Λ is 1.*

Proof. Let us fix $g \in o\big(\sqrt{n/\ln(n)}\big)$. Our aim is to construct a family of injective and size-preserving functions (codings) $\varphi^{\mathcal{B}}_n\colon \overline{\mathcal{B}^g_n} \to \Lambda_n$ such that the fraction $\#\varphi^{\mathcal{B}}_n\left(\overline{\mathcal{B}^g_n}\right)/L_n$ tends to 0 as n tends to infinity.

We first define a partition $\left(\overline{\mathcal{B}^g_n}(\overrightarrow{t}, \ell)\right)_{\overrightarrow{t},\ell}$ on $\overline{\mathcal{B}^g_n}$ as follows. Let $\overrightarrow{t}$ be a non empty sequence of terms (not necessarily closed) such that each of the elements of $\overrightarrow{t}$ starts with a λ and let $\ell \geq 1$ be an integer such that $0 \leq n-\ell-\mathrm{size}(\overrightarrow{t}) \leq g(n)$,

where size($\overrightarrow{t}$) denotes the sum of sizes of its components. We define $\overline{\mathcal{B}_n^g}(\overrightarrow{t}, \ell)$ as the set of terms u of the form:

$$u = \lambda x_1 \ldots \lambda x_p . v[t_1, \ldots, t_k]$$

where v is a purely applicative context with multiple holes, $\overrightarrow{t} = (t_1, \ldots, t_k)$ and $p = n - \ell - \text{size}(\overrightarrow{t})$.

First, it is clear that sets $\overline{\mathcal{B}_n^g}(\overrightarrow{t}, \ell)$ form a partition of $\overline{\mathcal{B}_n^g}$: they are pairwise disjoint by definition and any $u \in \overline{\mathcal{B}_n^g}$ belongs to $\mathcal{A}$ so it contains some λ not in the chain of head lambdas, therefore it belongs to some $\overline{\mathcal{B}_n^g}(\overrightarrow{t}, \ell)$ for some non-empty $\overrightarrow{t}$ and some $\ell \geq 1$.

Terms from $\overline{\mathcal{B}_n^g}(\overrightarrow{t}, \ell)$ differ only by applicative contexts, so the cardinality of $\overline{\mathcal{B}_n^g}(\overrightarrow{t}, \ell)$ is less than the number of all binary trees of size ℓ in which each leaf is either labeled with a variable (for which we have at most $g(n) - 1$ possibilities) or is an empty place where some sub-term can be plugged. Thus, we have:

$$\#\overline{\mathcal{B}_n^g}(\overrightarrow{t}, \ell) \leq P(\ell, n) := C(\ell)(g(n))^{\ell+1}.$$

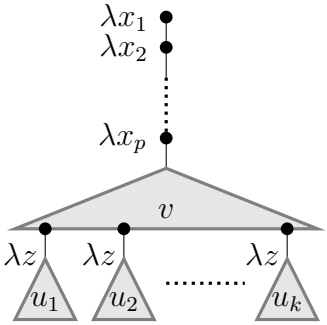

Figure 3: A term from $\overline{\mathcal{B}_n^g}(\overrightarrow{t}, \ell)$ where $\overrightarrow{t} = (\lambda z.u_1, \ldots, \lambda z.u_k)$

For $t \in \overline{\mathcal{B}_n^g}(\overrightarrow{t}, \ell)$ we define $\varphi_{n,\overrightarrow{t},\ell}(t)$ as follows. Let $\overrightarrow{t} = (t_1, \ldots, t_k)$ for some $k \geq 1$ and v be the the purely applicative context in the decomposition of t. We can write $t_i = \lambda z.u_i$ (see Figure 3). 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 of size

$$n - \ell = n \underbrace{- \ell}_{v \text{ removed}} \underbrace{- k}_{\lambda z \text{ removed}} \underbrace{+ 1}_{\text{head } \lambda z} \underbrace{+ k - 1}_{\text{applicative nodes}} .$$

Let $\lambda y.s$ denote the term rooted at the leftmost deepest λ of term t' and let V be the set of variables introduced by the lambdas occurring on the path from the

the root to λy. Since the unary height of t is the same as of t', by Theorem 31(iii) there are at least $\frac{n}{3\ln(n)}$ elements in V.

Let U be the set of purely applicative (therefore not closed) terms of size $\ell-1$ whose variables are chosen from V.There are at least

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

elements in U.

We can see that for all $\ell \geq 1$ the expression $\frac{P_{\ell,n}}{Q_{\ell,n}}$ tends to 0 as n tends to infinity. Therefore, for n large enough, there exists an injective function h which assigns to any purely applicative context v of size ℓ an element from U.

For any $u \in U$, let $\rho(t',u)$ be the term obtained by substituting the subterm $\lambda y.s$ in t' with $\lambda y.(u\ s)$.

Let $\varphi_{n,\overrightarrow{t},\ell}(t) = \rho(t', h(v))$ (see Figure 4). It is easy to check that the size of $\varphi_{n,\overrightarrow{t},\ell}$ is n and that, by the injectivity of h, $\varphi_{n,\overrightarrow{t},\ell}$ is injective, too.

Let $\varphi^{\mathcal{B}}_n := \bigcup_{\ell,\overrightarrow{t}} \varphi_{n,\overrightarrow{t},\ell}$. The function $\varphi^{\mathcal{B}}_n$ is an injection because co-domains of the $\varphi_{n,\overrightarrow{t},\ell}$ are all disjoint by the construction. Since the sets $\overline{\mathcal{B}^g_n}(\overrightarrow{t},\ell)$ form a partition of $\overline{\mathcal{B}^g_n}$, therefore it is enough to show, by means of Lemma 35, that $\frac{P_{n,\ell}}{Q_{n,\ell}}$ tends uniformly in l to 0 as n tends to infinity.

Let $\psi(n,\ell) = \frac{P(n,\ell)}{Q(n,\ell)}$. By the assumption on g, there is a function ε such that $\lim_{n\to\infty} \varepsilon(n) = 0$ and $P(n,\ell) \leq C(\ell)\left(\sqrt{\frac{n}{\ln(n)}}\varepsilon(n)\right)^{\ell+1}$. Therefore, we have

$$\psi(n,\ell) \leq \frac{C(\ell)}{3^{\ell}C(\ell-1)}\left(\frac{n}{\ln(n)}\right)^{\frac{1-\ell}{2}}(\varepsilon(n))^{\ell+1}.$$

For $\ell \geq 1$, $\left(\frac{n}{\ln(n)}\right)^{\frac{1-\ell}{2}}$ is decreasing in ℓ and since $\frac{C(\ell)}{3^{\ell}C(\ell-1)}$ is bounded it follows that $\psi(\ell,n)$ tends to 0 uniformly in ℓ.

□

6.2 Head λ's bind "many" occurrences

Now we are ready to present some theorems showing that in a random term head lambdas are "useful", i.e. they really bind some variables. The first result shows that in a random term many of head lambdas are binding.

Notation 38. Let $g\colon \mathbb{N} \to \mathbb{N}$ be a function such that $g \in o\big(\sqrt{n/\ln(n)}\big)$. By $\mathcal{D}^g$ we denote the class of terms such that $t \in \mathcal{D}^g$ iff

1. $t \in \mathcal{B}^{g+1}$, where $g+1$ is the function $n \to g(n)+1$,
2. each of first $g(\text{size}(t))$ head lambdas in t is binding.

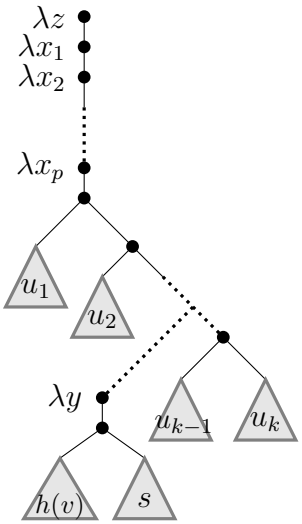

Figure 4: The term $\varphi_n^{\mathcal{B}}(t)$ from Theorem 37

Additionally, we denote by $\overline{\mathcal{D}_n^g} = \mathcal{B}^{g+1} \setminus \mathcal{D}^g$ the completion of the class $\mathcal{D}^g$ in $\mathcal{B}^{g+1}$ and by $\overline{\mathcal{D}_n^g}$ the set of terms from $\overline{\mathcal{D}^g}$ of size n.

Theorem 39. *Let $g\colon \mathbb{N} \to \mathbb{N}$ be a function such that $g \in o\big(\sqrt{n/\ln(n)}\big)$. The density of $\mathcal{D}^g$ in Λ is 1.*

Proof. Let us fix $g \in o\big(\sqrt{n/\ln(n)}\big)$. We construct a family of codings $\varphi_n^{\mathcal{D}}\colon \overline{\mathcal{D}_n^g} \to \Lambda_n$ such that their images are negligible in Λ_n, so that the fraction $\varphi_n^{\mathcal{D}}(\mathcal{D}_n^g)/L_n$ tends to 0 as n tends to infinity.

Let $t = \lambda x_1 \dots x_{g(n)+1}.u$ be a term from $\overline{\mathcal{D}_n^g}$ and let $i \leq g(n)$ be the smallest integer such that the i-th head λ in t does not bind any variable. Take

$$\varphi_n^{\mathcal{D}}(t) := \lambda x_1 \dots x_{i-1} x_{i+1}.\big(x_{i+1}\ (\lambda x_{i+2} \dots x_{g(n)+1}.u)\big).$$

The size of $\varphi_n^{\mathcal{D}}(t)$ is n. Terms from the set $\varphi_n^{\mathcal{D}}(\overline{\mathcal{D}_n^g})$ have less than $g(n)$ head λ's. By Theorem 37, the density of such terms in Λ is zero. Since the function $\varphi_n^{\mathcal{D}}$ is injective, the density of $\overline{\mathcal{D}_n^g}$ is also zero. □

Notation 40. Let $g, h\colon \mathbb{N} \to \mathbb{N}$ be functions such that $g \in o\big(\sqrt{n/\ln(n)}\big)$, $g(n) \geq 3$ for all n and $h \in o\left(\log_3\left(\frac{n}{\ln(n)}\right)\right)$. By $\mathcal{E}^{g,h}$ we denote the class of closed terms such that $t \in \mathcal{E}^{g,h}$ iff

1. $t \in \mathcal{D}^g$,
2. the total number of occurrences of variables bound by first three lambdas in t is greater than $h(\mathrm{size}(t))$.

Additionally, we denote by $\overline{\mathcal{E}^{g,h}} = \mathcal{D}^g \setminus \mathcal{E}^{g,h}$ the completion of the class $\mathcal{E}^{g,h}$ in $\mathcal{D}^g$ and by $\overline{\mathcal{E}^{g,h}_n}$ the set of terms from $\overline{\mathcal{E}^{g,h}}$ of size n.

Theorem 41. *Let $g, h\colon \mathbb{N} \to \mathbb{N}$ be functions such that $g \in o\big(\sqrt{n/\ln(n)}\big)$, $g(n) \geq 3$ for all n and $h \in o\left(\log_3\left(\frac{n}{\ln(n)}\right)\right)$. The density of $\mathcal{E}^{g,h}$ in Λ is 1.*

Proof. Let g and h be functions as in the assumptions of the theorem. We construct a family of codings $\varphi^{\mathcal{E}}_n\colon \overline{\mathcal{E}^{g,h}_n} \to \Lambda_n$ such that their images are negligible in Λ_n as n tends to infinity.

Let us define an equivalence relation $\sim_n$ on the set of terms of size n in the following way: $u \sim_n v$ iff u and v 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 $[u]$ the equivalence class of u.

Let $t = \lambda x_1 \lambda x_2 \lambda x_3 . u$ be a term from $\overline{\mathcal{E}^{g,h}_n}$. There are at most $3^{h(n)}$ elements in the class $[t]$.

Let $t' = \lambda xy.u[x_1 := y, x_2 := y, x_3 := y]$. The size of t' is $n-1$. Let us consider $\lambda a.v$ the subterm of t' such that λa is the leftmost deepest λ in t'. Denote by $V(t)$ the set of variables introduced 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 $V(t)$.

By Theorem 31(iii), there are at least $\frac{n}{3\ln(n)} - 2$ such λ's. Since $2 \leq \frac{n}{6\ln(n)}$, there are at least $\frac{n}{6\ln(n)}$ elements in $V(t)$. As $h \in o\left(\log_3\left(\frac{n}{\ln(n)}\right)\right)$, we have

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

Thus, we can find for each class $[t]$ an injective function $f_{[t]}$ from $[t]$ into the set $V(t)$.

We define $\varphi^{\mathcal{E}}_n(t)$ as the term obtained from t' by replacing the subterm $\lambda a.v$ with $\lambda a.(w\ v)$, where $w = f_{[t]}(t)$.

All terms from the image $\varphi^{\mathcal{E}}_n(\overline{\mathcal{E}^{g,h}_n})$ start with a λ that binds no variable. By Theorem 39 we know that such terms are negligible in Λ_n. Since $\varphi^{\mathcal{E}}_n$ is injective, the density of $\overline{\mathcal{E}^{g,h}}$ is zero, as well. □

Notation 42. Let k and ℓ be integers. Let $g\colon \mathbb{N} \to \mathbb{N}$ be functions such that $g \in o\big(\sqrt{n/\ln(n)}\big)$, $g(n) \geq 3$ for all n and let $h(n) = \left\lfloor \sqrt{\log_3\left(\frac{n}{\ln(n)}\right)} \right\rfloor$. Notice that $h \in o\left(\log_3\left(\frac{n}{\ln(n)}\right)\right)$. By $\mathcal{G}^{g,k,\ell}$ we denote the class of closed terms such that $t \in \mathcal{G}^{g,k,\ell}$ iff

1. $t \in \mathcal{E}^{g,h}$,

2. each of first k lambdas in t binds more than ℓ variables.

Additionally, we denote by $\overline{\mathcal{G}^{g,k,\ell}} = \mathcal{E}^{g,h} \setminus \mathcal{G}^{g,k,\ell}$ the completion of the class $\mathcal{G}^{g,k,\ell}$ in $\mathcal{E}^{g,h}$ and by $\overline{\mathcal{G}_n^{g,k,\ell}}$ the set of terms from $\overline{\mathcal{G}^{g,k,\ell}}$ of size n.

Theorem 43. *Let k and ℓ be integers. Let $g\colon \mathbb{N} \to \mathbb{N}$ be a function such that $g \in o\big(\sqrt{n/\ln(n)}\big)$ and $g(n) \geq 3$ for all n. The density of $\mathcal{G}^{g,k,\ell}$ in Λ is 1.*

Proof. Let g be a function as in the assumptions of the theorem and let us fix integers k and ℓ. Without loss of generality we can assume that $k \geq 3$. By Theorem 41, the total number of occurrences of variables bound by first k lambdas in terms from $\overline{\mathcal{G}_n^{g,k,\ell}}$ is greater than $h(n) = \left\lfloor \sqrt{\log_3\left(\frac{n}{\ln(n)}\right)} \right\rfloor$.

For $m \geq h(n)$ let us denote by $\mathcal{E}_n^{g,h}(m,k)$ be the set of terms from $\mathcal{E}_n^{g,h}$ with exactly m leaves bound by first k lambdas and let $\overline{\mathcal{G}_n^{g,k,\ell}}(m) = \overline{\mathcal{G}_n^{g,k,\ell}} \cap \mathcal{E}_n^{g,h}(m,k)$. By the definition, terms from $\overline{\mathcal{G}_n^{g,k,\ell}}(m)$ have exactly m leaves bound by first k lambdas and at least one of those lambdas binds at most ℓ variables.

Consider the equivalence relation $\sim_n$ on $\mathcal{E}_n^{g,h}(m,k)$ defined analogously as the relation from the proof of Theorem 41, but with respect to the first k (instead of three) head λ's.

Let $t \in \mathcal{E}_n^{g,h}(m,k)$. By Theorem 39, each of first k head lambdas of t are binding. Let us label k leftmost leaves (out of m) differently. For each of $m-k$ other leaves we have k possibilities. Thus, we know that the cardinality of $[t]$ is greater than k^{m-k}.

Now, let us estimate the upper bound for the cardinality of $[t] \cap \overline{\mathcal{G}_n^{g,k,\ell}}(m)$. In such terms there exists at least one lambda among first k ones which binds at most ℓ leaves (we can chose them out of m ones) and the other leaves (their number is at most equal to $m-1$) can be bound by $k-1$ lambdas. Thus, we obtain the upper bound equal to $k \cdot \binom{m}{l} \cdot (k-1)^{m-1} \leq k \cdot m^l \cdot (k-1)^{m-1}$.

Therefore, the quotient of the two cardinalities is less than

$$\frac{k \cdot m^\ell \cdot (k-1)^{m-1}}{k^{m-k}} = k^k m^\ell \left(\frac{k-1}{k}\right)^{m-1} \leq k^k (h(n))^\ell \left(\frac{k-1}{k}\right)^{h(n)-1} \text{ for all } m \geq h(n).$$

As n tends to infinity, the above quotient tends to 0 uniformly in m.

For $t \in \Lambda_n$ and $m \geq h(n)$ sets $[t] \cap \overline{\mathcal{G}_n^{g,k,\ell}}(m)$ form a partition of $\overline{\mathcal{G}_n^{g,k,\ell}}$. Now Lemma 35 finishes the proof. □

As a simple corollary of the above theorem, we obtain the following result.

Notation 44. Let k and ℓ be positive integers. Let $g\colon \mathbb{N} \to \mathbb{N}$ be a function such that $g \in o\big(\sqrt{n/\ln(n)}\big)$ and $g(n) \geq 3$ for all n. By $\mathcal{H}^{g,k,\ell}$ we denote the class of terms such that $t \in \mathcal{H}^{g,k,\ell}$ iff

1. $t \in \mathcal{G}^{g,k,\ell}$,

2. there are no two consecutive non-binding lambdas in t.

Additionally, we denote by $\overline{\mathcal{H}^{g,k,\ell}} = \mathcal{G}^{g,k,l} \setminus \mathcal{H}^{g,k,\ell}$ the completion of the class $\mathcal{H}^{g,k,l}$ in $\mathcal{G}^{g,k,\ell}$ and by $\overline{\mathcal{H}^{g,k,\ell}_n}$ the set of terms from $\overline{\mathcal{H}^{g,k,\ell}}$ of size n.

Lemma 45. *Let k and ℓ be positive integers. Let $g\colon \mathbb{N} \to \mathbb{N}$ be a function such that $g \in o\big(\sqrt{n/\ln(n)}\big)$ and $g(n) \geq 3$ for all n. The density of $\mathcal{H}^{g,k,\ell}$ in Λ is 1.*

Proof. We define a family of injective and size-preserving functions $\varphi^{\mathcal{H}}_n$ from $\overline{\mathcal{H}^{g,k,\ell}_n}$ into the set of terms whose leading λ binds only one occurrence of variable.

Let t be a term from $\overline{\mathcal{H}}^{g,k,\ell}_n$. Let t_1 be a subterm rooted at the highest occurrence of two non-binding lambdas, $t_1 = \lambda x.\lambda y.u$. We replace this subterm by the application $(z\ u)$, where z is a fresh variable. We obtain the term t' of size $n-1$ and, finally, we define $\varphi^{\mathcal{H}}_n(t) = \lambda z.t'$. The thesis follows from Theorem 43. □

6.3 A random term avoids any fixed closed term

Notation 46. Let j be a positive integer and $k(j) = \sum_{i\leq j} L_i$ (let us recall that L_i denotes the number of closed terms of size i). Let $g\colon \mathbb{N} \to \mathbb{N}$ be a function such that $g \in o\big(\sqrt{n/\ln(n)}\big)$, $g(n) \geq 3$ for all n and $\lim_{n\to\infty} g(n) = \infty$. By $\mathcal{I}^{g,j}$ we denote the class of closed terms such that $t \in \mathcal{I}^{g,j}$ iff

1. $t \in \mathcal{H}^{g,k(j),k(j)}$,

2. t does not contain any term from $\bigcup_{i\leq j} \Lambda_j$ as a subterm.

Additionally, we denote by $\overline{\mathcal{I}^{g,j}} = \mathcal{H}^{g,k(j),k(j)} \setminus \mathcal{I}^{g,j}$ the completion of the class $\mathcal{I}^{g,j}$ in $\mathcal{H}^{g,k(j),k(j)}$ and by $\overline{\mathcal{I}^{g,j}_n}$ the set of terms from $\overline{\mathcal{I}^{g,j}}$ of size n.

Theorem 47. *Let j be a positive integer and let $g\colon \mathbb{N} \to \mathbb{N}$ be a function such that $g \in o\big(\sqrt{n/\ln(n)}\big)$, $g(n) \geq 3$ for all n and $\lim_{n\to\infty} g(n) = \infty$. The density of $\mathcal{I}^{g,j}$ in Λ is 1.*

Proof. Let us fix a positive integer j and a function g as in the assumptions of the theorem. We construct a family of codings $\varphi^{\mathcal{I}}_n\colon \overline{\mathcal{I}^{g,j}_n} \to \Lambda_n$ such that their images are negligible in Λ_n.

There are $k(j) = \sum_{i\leq j} L_i$ elements in $\bigcup_{i\leq j} \Lambda_i$. Thus, there is a bijection f between $\bigcup_{i\leq j} \Lambda_i$ and $\{1, \ldots, k(j)\}$. Let $t_0 \in \bigcup_{i\leq j} \Lambda_i$ and let $f(t_0) = m$, where $m \leq k(j)$.

Let n be an integer satisfying $g(n) > k(j)$ and $n > k(j) + j$. Let $t \in \overline{\mathcal{I}^{g,j}_n}$ be a term containing t_0 as a subterm. By Theorem 37 the term t has more than $k(j)$ head λ's since $k(j) < g(n)$ (see Figure 5).

Let us consider the term s which is obtained from the term t by adding an additional unary node (labeled with λx) at depth m. Let us define $\varphi^{\mathcal{I}}_n(t)$ as the term t' obtained by replacing the leftmost deepest subterm t_0 in s by the term $u = (x\ (x\ (\ldots (x\ x)\ \ldots)))$ of size $j-1$ (see Figure 5). Thus, the size of t' is equal

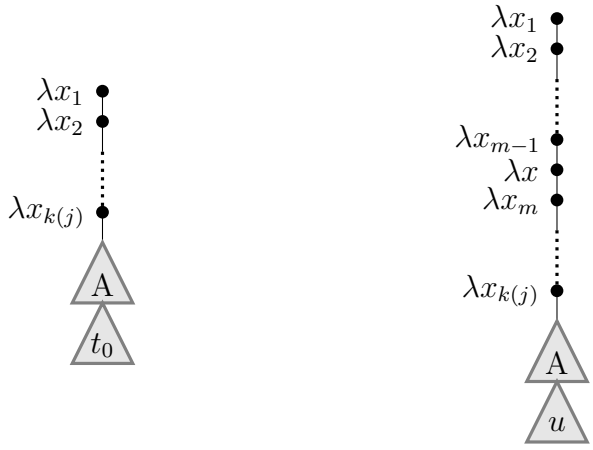

Figure 5: Terms $t \in \overline{\mathcal{I}_n^{g,j}}$ and t' from Theorem 47

to n. The variable x is bound by the m-th λ in the tree t'. Since $f(t_0) = m$, the function $\varphi_n^{\mathcal{I}}$ is injective.

By Theorem 43, each of $k(j)$ head λ's in term from $\mathcal{H}^{g,k(j),k(j)}$ of size n binds more than $k(j)$ variables. Terms from the image $\varphi_n^{\mathcal{I}}(\overline{\mathcal{I}_n^{g,j}})$ do not have this property, since the m-th λ binds only $j < k(j)$ variables. Thus, those terms are negligible among all terms of size n. □

6.4 The λ-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 of typical lambda terms is small.

Notation 48. Let j be an integer greater than 3. Let $g\colon \mathbb{N} \to \mathbb{N}$ be a function such that $g \in o\big(\sqrt{n/\ln(n)}\big)$, $g(n) \geq 3$ for all n and $\lim_{n\to\infty} g(n) = \infty$. By $\mathcal{J}^{g,j}$ we denote the class of closed terms such that $t \in \mathcal{J}^{g,j}$ iff

1. $t \in \mathcal{I}^{g,j}$,

2. the λ-width of t is at most than 2.

Additionally, we denote by $\overline{\mathcal{J}^{g,j}} = \mathcal{I}^{g,j} \setminus \mathcal{J}^{g,j}$ the completion of the class $\mathcal{J}^{g,j}$ in $\mathcal{I}^{g,j}$ and by $\overline{\mathcal{J}_n^{g,j}}$ the set of terms from $\overline{\mathcal{J}^{g,j}}$ of size n.

Theorem 49. *Let j be an integer greater than 3. Let $g\colon \mathbb{N} \to \mathbb{N}$ be a function such that $g \in o\big(\sqrt{n/\ln(n)}\big)$, $g(n) \geq 3$ for all n and $\lim_{n\to\infty} g(n) = \infty$. The density of $\mathcal{J}^{g,j}$ in Λ is 1.*

Proof. Let us fix an integer $j > 3$ and a function g as in the assumptions of the theorem. We construct a family of codings $\varphi_n^{\mathcal{J}} : \overline{\mathcal{J}_n^{g,j}} \to \Lambda_n$ such that their images are negligible in Λ_n. Let t be an element of $\overline{\mathcal{J}_n^{g,j}}$, therefore the λ-width of t is at least 3. 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).

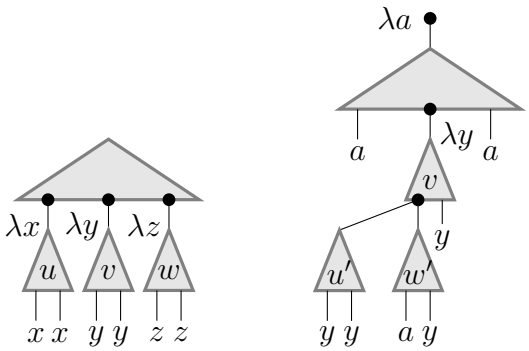

Figure 6: The terms t and $\varphi_n^{\mathcal{J}}(t)$ from Theorem 49

Let $\lambda x.u$, $\lambda y.v$ and $\lambda z.w$ be subterms rooted at those λ's (see Figure 6). Let $u' = u[x := y]$, let a be a new variable, let w' be the term obtained from w by replacing the leftmost occurrence of z with a and the others (possibly none) with y. Let $\varphi_n^{\mathcal{J}}(t)$ be the term obtained from t by adding λa at the root, substituting both subterms $\lambda x.u$ and $\lambda z.w$ with a and replacing the leftmost occurrence of y in v with term $(u'\ w')$. We have $\mathrm{size}(\varphi_n^{\mathcal{J}}(t)) = \mathrm{size}(t)$. Also note that since we chose the highest three incomparable λ's no variable becomes free in the constructed term.

We can reconstruct the term t from $\varphi_n^{\mathcal{J}}(t)$ by indicating places for λy and the subterm $(u'\ w')$:

- Let v_l (resp. v_r) be the deepest node above the two leftmost (resp. rightmost) 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 $(u'\ w')$ 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 $\varphi_n^{\mathcal{J}}$ contains only terms starting with a λ which binds only 3 occurrences of the corresponding variable, by Theorem 43, the density of $\varphi_n^{\mathcal{J}}(\overline{\mathcal{J}_n^{g,j}})$ is equal to zero. The injectivity of $\varphi_n^{\mathcal{J}}$ finishes the proof. □

6.5 The density of strongly normalizable terms

From Theorem 49 we know that almost all terms are of width at most 2. In Section 2 we introduced the notion of 'safeness' for terms of width 2 with the following property: safeness and λ-width at most 2 imply strong normalisation (Proposition 18).

Now we prove that the set of unsafe terms of width 2 has density 0.

Notation 50. Let j be an integer greater than 3. Let $g\colon \mathbb{N} \to \mathbb{N}$ be a function such that $g \in o\big(\sqrt{n/\ln(n)}\big)$, $g(n) \geq 3$ for all n and $\lim_{n\to\infty} g(n) = \infty$. By $\mathcal{K}^{g,j}$ we denote the class of closed terms such that $t \in \mathcal{K}^{g,j}$ iff

1. $t \in \mathcal{J}^{g,j}$,

2. t is safe.

Additionally, we denote by $\overline{\mathcal{K}^{g,j}} = \mathcal{J}^{g,j} \setminus \mathcal{K}^{g,j}$ the completion of the class $\mathcal{K}^{g,j}$ in $\mathcal{J}^{g,j}$ and by $\overline{\mathcal{K}^{g,j}_n}$ the set of terms from $\overline{\mathcal{K}^{g,j}}$ of size n. Note that terms from $\overline{\mathcal{K}^{g,j}}$ are of λ-width at most 2 and are unsafe.

Proposition 51. *Let j be an integer greater than* 3. *Let $g\colon \mathbb{N} \to \mathbb{N}$ be a function such that $g \in o\big(\sqrt{n/\ln(n)}\big)$, $g(n) \geq 3$ for all n and $\lim_{n\to\infty} g(n) = \infty$. The density of $\mathcal{K}^{g,j}$ in Λ is* 1.

Proof. The root of the minimal subterm of width 2 of a term is called the branching node and is always binary. We show that the density of $\overline{\mathcal{K}^{g,j}}$ in Λ is 0. Let us divide the set $\overline{\mathcal{K}^{g,j}_n}$ into two disjoint subsets:

$\overline{\mathcal{K}^{g,j,1}_n}$: the set of terms of size n such that both the lengths of paths from the branching node to the two highest incomparable binding lambdas is not greater than $\ln(n)$,

$\overline{\mathcal{K}^{g,j,2}_n}$: the set of remaining terms of size n.

We can construct a family of codings from the set $\overline{\mathcal{K}^{g,j,1}_n}$ into Λ_n in the following two steps. First, remove two highest pairwise incomparable binding lambdas and put one lambda, binding their variables, at the root of the whole term. The size of the obtained term is smaller by 1 and the branching node is uniquely determined. Second, insert one non-binding lambda between head lambdas of the term. According to Theorem 37 we have more than $g(n) > \ln(n)^2$ head lambdas. Therefore we can encode the lengths of the paths from the branching node to the two highest binding lambdas as the position of this new lambda. By Theorem 39 the image of such transformation have density 0.

For the set $\overline{\mathcal{K}^{g,j,2}_n}$ proceed as follows. First, choose the path (without loss of generality we can assume it is the left path, the case of the right one is analogous) that is longer than $\ln(n)$ and connects the branching node and one of

two highest binding lambdas. Let t_0 be the subterm rooted at this lambda. By Lemma 45 at least half of the nodes on this path are binary (since there are no two consecutive lambdas in the tree). Let $t_1, \ldots, t_k$ be the right subtrees rooted at the consecutive binary nodes on the path (the path goes always to the left since the term is unsafe). Secondly, chose some leaf v belonging to some subtree $t_1, \ldots, t_k$ 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 the highest binding lambda of the fair subtree below the branching node. The encoding preserves size and the number of possibilities for the choice of a leaf v exceeds $\ln(n)/2$. Therefore, terms from $\overline{\mathcal{K}_n^{g,j,2}}$ are negligible in Λ_n as n tends to infinity. □

Main Theorem 52. *The set of strongly normalizable terms has density* 1.

Proof. First, by Theorem 49, we can focus on terms of width at most 2. Proposition 51 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 18 shows that they are all strongly normalizable. □

7 Combinatory logic

In this section we show that our main result about strong normalization of random lambda calculus terms does not have an analog in the world of random combinatory logic terms. On the contrary, a random combinatory logic term is not strongly normalizing. The main technique used in this section is the theory of generating functions.

As stated in section 2 we can look at combinatory logic terms as at rooted binary trees whose leaves are labeled with combinators K, S and I. We denote by $\mathcal{F}_n$ the number of such trees with n inner nodes (see section 2.3). Obviously the set $\mathcal{F}_n$ is finite. We denote its cardinality by F_n. It is trivial to notice that $F_n = C(n)3^{n+1}$ where $C(n)$ is $n-th$ Catalan number (see proposition 25).

Proposition 53. *1. The generating function f enumerating the set of combinators (sequence F_n) is given by*

$$f(z) = \frac{1 - \sqrt{1-12z}}{2z}.$$

2. Let $t_0 \in \mathcal{F}_{n_0}$ be a combinator of size $n_0 \geq 1$. The generating function f_{t_0} enumerating the set of all combinators having t_0 as a subterm is given by

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

Proof. 1. F_n denotes the number of combinators of size n. Since there are three combinators of size 0, we have $F_0 = 3$. Combinators of size $n \geq 1$ are built of two combinators of sizes, respectively, i and $n-i-1$ ($i = 0, \ldots, n-1$), thus $F_n = \sum_{i=0}^{n-1} F_i F_{n-i-1}$. From this recurrence relation we obtain that the generating function f for the sequence (F_n) satisfies the equation

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

Solving this equation in $f(z)$ we get two solutions:

$$\frac{1-\sqrt{1-12z}}{2z} \quad \text{and} \quad \frac{1+\sqrt{1-12z}}{2z}.$$

We have $F_0 = 3$, so $\lim_{z\to 0} f(z) = 3$. Thus, the desired generating function is given by the first solution.

2. Let t be a combinator having t_0 as a subterm. Then t either is equal to t_0 or is of the form of application $t = t_1\ t_2$, in which case either t_0 is a subterm of t_1 but not of t_2 or t_0 is a subterm of t_2 but not of t_1 or, finally, t_0 is a subterm of both t_1 and t_2. We get the following equation:

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

which can be simplified to

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

Solving this equation in f_{t_0} gives us two possible solutions:

$$\frac{-\sqrt{1-12z} + \sqrt{1-12z+4z^{n_0+1}}}{2z} \quad \text{and} \quad \frac{-\sqrt{1-12z} - \sqrt{1-12z+4z^{n_0+1}}}{2z}.$$

Since $n_0 \geq 1$, there is no term of size 0 having t_0 as a subterm. Thus, $\lim_{z\to 0} f_{t_0}(z) = 0$. The first function satisfies this condition, so this is the wanted generating function. □

The following theorem shows that the result similar to Theorem 47 is not valid in combinatory logic.

Theorem 54. *Let t_0 be a combinator. The density of combinators having t_0 as a subterm is* 1.

Proof. We prove the thesis applying Theorem 24. We start with normalizing functions f_{t_0} and f in such a way that the singularity located at $z = 1$ is the closest one to the origin. Hence, we define functions $\overline{f_{t_0}}(z) := z f_{t_0}(z/12)$ and $\overline{f}(z) := z f(z/12)$. We get

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

This representation reveals that the only singularity of $\overline{f_{t_0}}(z)$ and $\overline{f}(z)$ located in $|z| \leq 1$ is indeed at $z = 1$. We have to remember that the multiplication by z and the change of a caliber of the radius of convergence for functions f_{t_0} and f affect sequences represented by the new functions. Therefore, $\overline{f_{t_0}}$ and $\overline{f}$ enumerate sequences $(12)^{n-1}\,([z^{n-1}]\{f_{t_0}(z)\})$ and $(12)^{n-1}\,([z^{n-1}]\{f(z)\})$, respectively.

Now, let us consider functions $\widetilde{f}$ and $\widetilde{f_{t_0}}$ satisfying the following equations: $\widetilde{f}(\sqrt{1-z}) = \overline{f}(z)$ and $\widetilde{f_{t_0}}(\sqrt{1-z}) = \overline{f_{t_0}}(z)$. They are defined in the following way:

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

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

$$(\widetilde{f_{t_0}})'(z) = -\frac{1}{2} + \frac{\left(2\,z - 8(n_0+1)z\,\left(\frac{1-z^2}{12}\right)^{n_0}\right)}{4\,(1-z^2)\,\sqrt{z^2 + 4\,\left(\frac{1-z^2}{12}\right)^{n_0+1}}}, \qquad (\widetilde{f})'(z) = -\frac{1}{2}.$$

Finally, by computing the values of those derivatives at $z = 0$ we get $(\widetilde{f_{t_0}})'(0) = -\frac{1}{2}$ and $(\widetilde{f})'(0) = -\frac{1}{2}$.

To complete the proof we apply Theorem 24, obtaining:

$$\lim_{n\to\infty} \frac{[z^n]\{f_{t_0}(z)\}}{[z^n]\{f(z)\}} = \lim_{n\to\infty} \frac{(12)^{n-1}[z^{n-1}]\{f_{t_0}(z)\}}{(12)^{n-1}[z^{n-1}]\{f(z)\}} = \frac{(\widetilde{f_{t_0}})'(0)}{(\widetilde{f})'(0)} = 1. \qquad \square$$

Main Theorem 55. *The density of non-strongly normalizing combinators is* 1.

Proof. Let $\Omega = S\ I\ I\ (S\ I\ I)$. The combinator Ω reduces to itself and thus is not strongly normalizing. The thesis follows directly from Theorem 54, since the density of combinators containing Ω as a subterm is 1. $\square$

8 Discussion

The difference between Theorem 52 in the λ-calculus and Theorem 55 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, $\text{size}(T_1(t)) \leq k \cdot \text{size}(t)$. 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 but exponential. As far as we know, $\text{size}(T_2(t))$ is of order $3^{\text{size}(t)}$. The point is that T_2 has to code the variable binding in some way and this takes place.

8.1 Future work and open questions

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

1. Give the asymptotics of L_n or, at least, better upper and lower bounds.
2. Give the density of typable terms. Numerical experiments done by Jue Wang (see [19]) seem to show that this density is 0.
3. 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).

8.2 Possible applications

It is now popular to test programs written in functional languages using randomly generated set of inputs [2]. For higher-order functional programs where inputs are functions or algorithms for computing functions, this also means the ability to generate typical functions under certain known distributions.

For many typed languages such as OCaml or Haskell, testing functional programs usually means the possibility of supplying random typed lambda terms generated in compliance with theirs natural distribution, probably different for different type.

For untyped languages such as LISP, problem of testing programs is very close to capability of generating pure random λ-terms. In our case, those terms automatically enjoy important properties such as strong normalisation if they do not use recursive definition. However, it would be nice to have a distribution where terms with other computationally good properties have density one.

In light of our results, the distribution induced by the size of terms in combinatory logics is dramatically different because most term are not strongly normalizable while the distribution for pure λ-term with variables of size 0 enjoys the density 1 of strong normalisation. One could argue that the width at most 2 is a negative result for testing (tuple of functions are common and encoded with terms of arbitrary width).

Results presented in this paper are the starting point for the similar research based on the other notion of sizes necessary for applications. Possible direction of research is discussed in the next section.

8.3 Discussion on the various notion of sizes

The notion of sizes (which seems the most popular) where application, variables and abstraction have the same size is in fact not always the most natural, intuitively because variables require less information than application and abstrac-

tion. To illustrate this, here are two possible representations for terms: one in OCaml, the other in C.

The following definition in OCaml gives a size of term which count 3 words for application and abstraction and 2 words for each variable (because in OCaml you have a "tag" word for each constructor):

```
type var = int
type term =
  App of term * term
| Abs of var * term
| Var of var
let delta = Abs(0,App(Var 0, Var 0))
```

In C you may represent term using a pointer to a union type. We give here the type, and functions to construct lambda-terms:

```
typedef union term_union *term;
union term_union {
  struct { term first_arg ; term second_arg ; } app;
  struct { unsigned int var; term fun_body ; } abs;
};

term make_variable(unsigned var) {
  return (term) (4*var+1);
}
term make_application(term t1, term t2) {
  term t0 = (term) malloc(sizeof(union term_union));
  if (!t0) { fprintf(stderr, "not enough memory"); }
  t0->app.first_arg = t1; t0->app.second_arg = t2;
  return(t0);
}
term make_abstraction(unsigned var, term t) {
  term t0 = (term) malloc(sizeof(union term_union));
  if (!t0) { fprintf(stderr, "not enough memory"); }
  t0->abs.var = (4*var+3); t0->abs.fun_body = t;
  return(t0);
}
```

Here, we use the fact that pointers in C are always even. So we represent the variable numbered n by the integer $4n + 1$ using a type cast. To distinguish λ from application, the binding occurrence of the variable n is represented by $4n + 3$. In this representation, application and abstraction use two words and variable use no word at all. This model is equivalent (up to a constant factor) to the model for which we prove all our results.

Remark: If we reuse the same natural number for variable names whenever it is possible (for instance when the variable does not appear), then, the smallest term using the C representation, not representable in a 32 bits machine requires $2^{30} + 1$ abstractions followed by a sequence of applications with 2^{30} nodes to

make sure all the variables appear at least once. The total size is then $2^{31} + 1$. Therefore, we can consider that our density results are probably applicable in practice. To check this, we should check the speed of convergence of our density results.

If you consider that the notion of size needs to be correct for arbitrary large terms, then, the size of variable must be non constant (for instance, we could consider logarithmic size in base 2^{32} for variables) and we think that obtaining result in this case requires a major breakthrough in combinatorics.

References

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

[2] Koen Claessen and John Hughes *QuickCheck: A Lightweight Tool for Random Testing of Haskell Programs* Proc. Of International Conference on Functional Programming (ICFP), ACM SIGPLAN, 268-279, 2000

[3] H.B. Curry, R. Feys, *Combinatory Logic.* Vol. I. Amsterdam: North Holland, 1958.

[4] R. David, *Normalization without reducibility.* APAL n 107 (2001) p 121-130.

[5] R. David, *A short proof of the strong normalization of the simply typed lambda calculus.* `http://www.lama.univ-savoie.fr/~david/`

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

[7] 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.

[8] A. Genitrini, J. Kozik, *Quantitative comparison of Intuitionistic and Classical logics - full propositional system*, LFCS09, Lecture Notes in Computer Science 5407 (2009) pp. 280-294.

[9] A. Genitrini, J. Kozik, M. Zaionc, *Intuitionistic vs. Classical Tautologies, Quantitative Comparison*, Lecture Notes in Computer Science 4941, (2008) pp. 100-109.

[10] 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.

[11] Z. Kostrzycka and M. Zaionc. *Statistics of intuitionnistic versus classical logic*, Studia Logica, 76(3):307-328, 2004.

[12] 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.

[13] M. Moczurad, J. Tyszkiewicz and M. Zaionc. *Statistical properties of simple types,* Mathematical Structures in Computer Science, 10(5):575-594, 2000.

[14] M. Schönfinkel, *Über die Bausteine der mathematischen Logik*, Mathematische Annalen 92, pp. 305-316, 1924.

[15] R.M. Smullyan, *To Mock a Mockingbird and Other Logic Puzzles: Including an Amazing Adventure in Combinatory Logic*, Knopf, 1985.

[16] 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`.

[17] M. Zaionc, *On the asymptotic density of tautologies in logic of implication and negation*, Reports on Mathematical Logic, vol 39, pp 67-87, 2005.

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

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

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