
ASYMPTOTICALLY ALMOST ALL λ -TERMS ARE STRONGLY NORMALIZING

RENÉ DAVID^a, KATARZYNA GRYGIEL^b, JAKUB KOZIK^c, CHRISTOPHE RAFFALLI^d,
GUILLAUME THEYSSIER^e, AND MAREK ZAIONC^f

^{a,d,e} LAMA, CNRS, Université de Savoie, 73376 Le Bourget-du-Lac, France
e-mail address: {rene.david, christophe.raffalli, guillaume.theyssier}@univ-savoie.fr

^{b,c,f} Theoretical Computer Science, Jagiellonian University, Łojasiewicza 6, Kraków, Poland
e-mail address: {Katarzyna.Grygiel, Jakub.Kozik, zaionc}@tcs.uj.edu.pl

ABSTRACT. We present a quantitative analysis of various (syntactic and behavioral) properties of random λ -terms. Our main results show that asymptotically, almost all 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.

1. INTRODUCTION

Since the pioneering work of Church, Turing *et al.*, more than 70 years ago, a wide range of computational models has been introduced. It has been shown that the feasible models are all equivalent in the sense of computational power. However, this equivalence says nothing about what *typical* programs or machines of each of these models do.

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 language of this kind (see [11, 16, 2] for similar work on other models of

2012 ACM CCS: [Mathematics of Computing]: Discrete Mathematics—Combinatorics—Combinatoric problems.

Key words and phrases: lambda-calculus, combinatorics, normalisation, combinatory logic.

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 .

computation). To our knowledge, the only work on this subject is some experiments carried out by Jue Wang (see [19]). Most interesting properties of λ -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 upper and lower bounds for this super-exponential number (see Section 5). 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 random λ -term. In particular, we show that almost every closed λ -term begins with “many” lambdas (the precise meaning is given in Theorem 6.3). Moreover, each of them binds “many” occurrences of variables (Theorems 6.5, 6.7 and 6.9). Finally, given any fixed closed λ -term, almost no λ -term has this term as a subterm (Theorem 6.13).

We also give results on the behavior of terms, which is our original motivation. We show that a random term is strongly normalizing (SN for short) with asymptotic probability 1. Let us recall that, in general, knowing whether a term is 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 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 a subterm. This implies that almost every term is not SN . The difference of results concerning strong normalization between λ -calculus and combinatory logic is not contradictory since the coding of bound variables in combinatory logic induces a large increase of size. This is discussed in Section 8.

Our interest in statistical properties of computational objects, like λ -terms or combinators, is a natural extension of similar work 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, of random formulas). For the purely implicational logic with one variable (and simple type systems), the exact value of the density of true formulas has been computed in [14, 21]. Quantitative relationship between intuitionistic and classical logics (based on the same language) has also been analyzed. The exact value describing how large the intuitionistic fragment of the classical logic with one variable is has been determined in [12]. For results with more than one variable, or with other logical connectives, consult [8, 10, 9].

The organization of the paper is as follows. In Section 2 we recall basic definitions and facts about λ -calculus and combinatory logic. Section 3 gives combinatorial notations which we will need in our proofs. It introduces generating functions and basic techniques to compute asymptotics. The notion of density and its basic properties is introduced in Section 4. The lower and upper bounds for the number of λ -terms of size n are given in Section 5. In Section 6 we prove theorems about random λ -terms using coding which is an injective and size-preserving function on terms. Our main result establishing that the set of strongly normalizable terms has density 1 appears at the end of this section in Theorem 6.18. Section 7 contains results in combinatory logic, namely the fact that every fixed term appears in almost every term. The main result of this section, in Theorem 7.3, states that

the density of non-strongly normalizing combinators is 1. Finally Section 8 discusses future work, open questions and possible applications of results.

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. We do this mainly to make our notations and conventions precise. It should be enough for defining the notion of size, but for substitution and reduction and normalization we recommend [1].

Definition 2.1. 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 $t_1 t_2 \dots t_n$ without parentheses for $(\dots (t_1 t_2) \dots t_n)$.

As usual, λ -terms are considered modulo 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 2.2. By a λ -tree we mean a rooted tree of the following form there are two kinds of inner nodes – labeled with $@$ and with λ . Nodes labeled with $@$ have two successors: left and right. Nodes labeled with λ have only one successor. Each Leaf of a tree is labeled either with a variable or with a pointer to one of the λ nodes above it.

For every λ -term t we define the λ -tree $G(t)$ in the following way:

- If t is a variable x , then $G(t)$ is a single node labeled with x .
- If $t = t_1 t_2$, then $G(t_1 t_2)$ is a tree with the root labeled with $@$ and two subtrees $G(t_1)$ (left) and $G(t_2)$ (right).
- If $t = \lambda x.u$, then $G(t)$ is obtained from $G(u)$ in four steps:
 - add a new root labeled with λ ;
 - connect the new root with $G(u)$;
 - connect all leaves of $G(u)$ labeled with x with the new root;
 - remove all labels x .

Observation 1. If T is a λ -tree then $T = G(t)$ for some λ -term t . Terms t and u are α -convertible iff $G(t)$ and $G(u)$ are the same tree.

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 2.3. 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 this occurrence of λx is *binding* in t if x has a free occurrence in a .

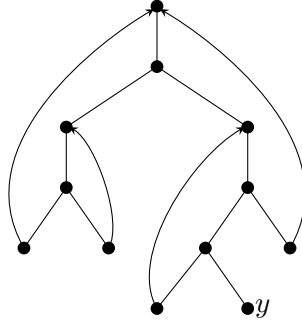


Figure 1: The λ -tree representing the term $\lambda z.(\lambda u.zu)(\lambda u.uyz)$ (labels of inner nodes are not shown in the figure and can be recovered from their degrees)

- (3) The *unary height* of t is the maximum number of lambdas on a branch in the λ -tree of t .
- (4) Two lambdas in t are called *incomparable* if there is no branch in the λ -tree containing both of them. The λ -*width* of t (or simply *width* of t when there is no ambiguity) is the maximum number of pairwise incomparable binding lambdas. *Remark:* a closed λ -term has width at least 1.
- (5) We say that t has k *head lambdas* if its λ -tree starts with at least k unary nodes.

Definition 2.4.

- When t and u are terms, $t[x := u]$ denotes the *capture avoiding substitution* of u for the free occurrences of the variable x in t . Bound variables of t may have to be renamed to avoid capture of free variables in u .
- A term of the form $(\lambda x.t)u$ is called a β -*redex*. A λ -term is in *normal form* if it does not contain β -redex subterms. The least relation \triangleright on terms satisfying $(\lambda x.t)u \triangleright t[x := u]$ and closed under contexts is called β -*reduction*.
- A term t is (*weakly*) *normalizing* if there is a finite reduction sequence starting from t and ending in a normal form.
- A term t is *strongly normalizing* (SN) if all reduction sequences starting from t are finite. If t is SN, we denote by $\eta(t)$ the length of its longest reduction. The fact that such a longest reduction exists follows from König's lemma. If t is not SN, $\eta(t) = +\infty$.

In the λ -tree representation, a redex is a subtree of the λ -tree. Therefore β -reduction can be seen as an operation on λ -trees (see Fig. 2).

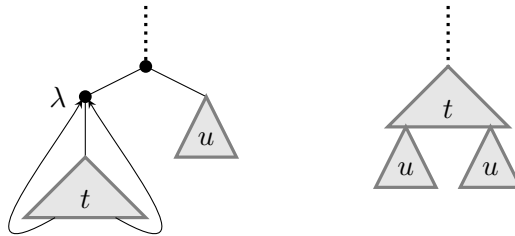


Figure 2: β -reduction scheme

Definition 2.5. The *size* of a term (denoted by $\text{size}(\cdot)$) is defined recursively as follows:

- (i) $\text{size}(x) = 0$ if x is a variable,
- (ii) $\text{size}(\lambda x.t) = 1 + \text{size}(t)$,
- (iii) $\text{size}(t u) = 1 + \text{size}(t) + \text{size}(u)$.

As we can see, $\text{size}(t)$ is the number of inner nodes in the λ -tree $G(t)$.

Notation 2.6. 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 $(L_n)_{n \in \mathbb{N}}$ has been done. Moreover, typical combinatorial techniques do not seem to apply easily for this task.

2.2. Innocuous and safe λ -terms. This section introduces the notion of *safe λ -terms* which is a sufficient condition for being SN (Proposition 2.16).

Definition 2.7.

- (1) Let t be a term of width 1. We say that t is *innocuous* if there is no binding λ on the leftmost branch of t (this includes the root of t).
- (2) We say that t is *safe* if either it has width at most 1 or if it has width 2 and for $(u v)$ being the smallest subterm of t of width 2, at least one of the terms u and v is innocuous.

Definition 2.8.

- 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 σ be a substitution. By $t[\sigma]$ we denote the term obtained from t by simultaneous replacement of all free occurrences of variables x from the domain of σ by $\sigma(x)$.
- A *context* is a λ -term with a unique hole denoted by \square . Traditionally, contexts are defined by a BNF grammar:

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

- When E is a context and t is a term, $E[t]$ denotes the result of replacing the hole in E by t allowing captures (i.e. the lambdas in E can bind variables in t).
- For a context E , we define $\eta(E)$ as $\eta(E[x])$ and $\text{size}(E)$ as $\text{size}(E[x])$ where x is an arbitrary variable not captured by E .
- In a few cases, we need contexts with multiple holes. When E is a context with exactly n holes, $E[t_1, \dots, t_n]$ denotes the term where the holes of E are substituted from the leftmost to the rightmost by terms t_1, \dots, t_n (in this order).

In some proofs in this section we use the following basic fact concerning strong normalization of λ -terms:

Fact 2.9. Let t be a λ -term.

- If $t = (x t_1 \dots t_n)$, for some variable x , with $n \geq 0$, then $\eta(t) = \eta(t_1) + \dots + \eta(t_n)$. Moreover t is SN if and only if t_1, \dots, t_n are SN.
- If $t = \lambda x.u$, then $\eta(t) = \eta(u)$ and t is SN if and only if u is SN.
- If $t = ((\lambda x.u) v t_1 \dots t_n)$ with $n \geq 0$ and t is SN, then $\eta(t) > \eta(u[x := v] t_1 \dots t_n)$ and $\eta(t) > \eta(v) + \eta(t_1) + \dots + \eta(t_n)$. Moreover t is SN if and only if v and $(u[x := v] t_1 \dots t_n)$ are SN.

These three cases cover all possible forms of t . Moreover, if x is a variable, then t is *SN* if and only if $(t x)$ is *SN*.

Proof. These facts are “folklore”, but they are not trivial to prove directly from the definition of β -reduction and the proof is not found in the usual literature. Here, we give a proof sketch using the fact that Barendregt’s [1] *perpetual norm* (length of the *perpetual reduction strategy*) is in fact the length of the longest reduction. This is proved in [15].

The perpetual strategy is the strategy that reduces the left-most redex first, except when this redex is a K-redex $((\lambda x.u)t$ when x is not free in u). In this case, the redex is reduced only when t and u are normal. For a formal definition see [1] or [15].

The equality about $\eta(t)$ in the first two items are immediate from this, by induction on the length of the reduction.

Using the perpetual norm, we have

$$\eta((\lambda x.u) v t_1 \dots t_n) = 1 + \max(\eta(u[x := v] t_1 \dots t_n), \eta(v) + \eta(u t_1 \dots t_n)).$$

The two terms in the max correspond respectively to the case where x occurs free in u and the case where the redex is a K-redex.

For the equivalence, one direction comes from the fact that subterms and reducts of an *SN* term are *SN*. For the other direction we have to prove that if v and $(u[x := v] t_1 \dots t_n)$ are *SN* then so is $t = ((\lambda x.u) v t_1 \dots t_n)$. This is done by induction on $\eta(u) + \eta(v) + \eta(t_1) + \dots + \eta(t_n)$ looking at the different possible reductions of t .

The fact that if t is *SN* then so is $(t x)$ is proved using the perpetual norm to establish that $\eta(t x) \leq \eta(t) + 1$ (in fact $\eta(t x) = \eta(t) + 1$ if t reduces to a term starting with λ and $\eta(t x) = \eta(t)$ otherwise). \square

Lemma 2.10. *The set of terms of width at most 1 is closed under β -reduction.*

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

Let t be a term of width 1. First, let us remark that all binding lambdas 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 has no free occurrences in u and $t' = E[u]$ or it has some free occurrence in u and v must have width 0, which means that every variable of v is either free in t or bound by some lambda occurring in the context E . It is clear that t' is still of width 1 because the binding lambdas remain on one branch. \square

Lemma 2.11. *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 number 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 2.10. To prove this, we consider a β -reduction: $t = E[(\lambda x.u) v] \triangleright E[u[x := v]] = t'$ and distinguish two cases:

- If x does not occur in u , then $N_1(t)$ is non-increasing. Moreover, it is decreasing if v contains some binding lambdas or if E binds some variables that occur only in v . Therefore, if $N_1(t)$ is constant, then $N_0(t)$ is decreasing: we erase at least one non-binding λ and do not transform binding ones into non-binding ones.
- If x occurs 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 lambdas. Therefore, $N_1(t)$ is decreasing. \square

Lemma 2.12. *If u has width 0 and t_1, \dots, t_n are SN terms, then the term $t = (u t_1 \dots t_n)$ is SN.*

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

- If $u = x$, the result is trivial by Fact 2.9.
- If $u = (u' v)$, v has width 0 and is SN because of Lemma 2.11. We conclude by induction on u' .
- For $u = \lambda x.u'$ we consider two cases: if $n = 0$, the result follows from Lemma 2.11; otherwise, by Fact 2.9, it is enough to show that the head reduct of t is SN. But, since u has width 0, this reduct is $(u' t_2 \dots t_n)$ and the result follows from the induction hypothesis. \square

Lemma 2.13. *Let $t \in SN$ be a term and σ be a substitution such that, for each x , there is k such that $\sigma(x) = (u v_1 \dots v_k)$ where u has width 0 and $v_1 \dots v_k$ are SN. Then $t[\sigma] \in SN$.*

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

- If $t = \lambda x.t_1$ or if $t = (x t_1 \dots t_n)$ with x not in the domain of σ , it is enough to prove that for all i , $t_i[\sigma]$ is SN. This follows from the induction hypothesis because $\eta(t_i) \leq \eta(t)$ and $\text{size}(t_i) < \text{size}(t)$.
- If $t = ((\lambda x.u) v t_1 \dots t_n)$ we show that $v[\sigma]$ and $(u[x := v] t_1 \dots t_n)[\sigma]$ are SN and apply Fact 2.9. This follows from the induction hypothesis because $\eta(v) < \eta(t)$ for the first point and because $\eta(u[x := v] t_1 \dots t_n) < \eta(t)$ for the second.
- If $t = (x t_1 \dots t_n)$ where x is in the domain of σ . Then we have $t[\sigma] = (\sigma(x) t_1[\sigma] \dots t_n[\sigma])$ which is SN by Lemma 2.12 because $t_1[\sigma], \dots, t_n[\sigma]$ are SN by the induction hypothesis and $\sigma(x) = (u v_1 \dots v_k)$ where u has width 0 and $v_1 \dots v_k$ are SN. \square

Definition 2.14. We define the set of contexts of width at most 1 by the following BNF grammar (where $\overline{\Lambda}_0$ denotes the set of λ -terms of width 0):

$$E := [] \mid \lambda x.E \mid (E \overline{\Lambda}_0) \mid (\overline{\Lambda}_0 E).$$

This definition means that all the binding lambdas are on the path from the root to the hole of the context.

Lemma 2.15. *Let E be a context of width 1 and $u \in SN$ be a term. Then $E[u] \in SN$.*

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

If $E = (E_1 v)$, where $v \in \overline{\Lambda}_0$, then $E[u] = (E_1[u] x)[x := v]$ where x is a fresh variable. $E_1[u]$ is SN by induction hypothesis because $\text{size}(E_1) < \text{size}(E)$. Therefore $(E_1[u] x)$ is SN by Fact 2.9 and finally $(E_1[u] x)[x := v]$ is SN by Lemma 2.13.

If $E = (v E_1)$, then $E[u] = (x E_1[u])[x := v]$ where x is a fresh variable and $E_1[u]$ is SN by induction hypothesis because $\text{size}(E_1) < \text{size}(E)$. Therefore $(x E_1[u])$ is SN and finally $(x E_1[u])[x := v]$ is SN by Lemma 2.13. \square

Proposition 2.16. *All safe terms are SN.*

Proof. If t has width at most one, the result follows directly from Lemma 2.11. 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 as $E[(t_1 t_2)]$ where E is a context of width at most 1 and t_1 and t_2 are both of width 1. By Lemma 2.15, it is therefore enough to show that $(t_1 t_2)$ is SN.

We know that t is safe. This means that at least t_1 or t_2 is innocuous. If t_i is innocuous, it can be written $F[(u v)]$ where u has width 0, v has width 1 and F belongs to the family of contexts defined by the following BNF grammar:

$$F := [] \mid \lambda_{-}F \mid (F \bar{\Lambda}_0)$$

where λ_{-} denotes non-binding lambdas and $\bar{\Lambda}_0$ denotes terms of width 0.

The context F is defined precisely to denote the beginning of the leftmost branch until we reach an application node whose argument is of width 1. The definition of innocuous terms together with the definition of width 1 ensures the existence of such an application node on the leftmost branch.

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

In both cases, $(t_1 t_2) = t'[x := (u v)]$. We can conclude by Lemma 2.13 that $(t_1 t_2)$ is *SN* since u has width 0 and t' and v are *SN* (by Lemma 2.11, since they have width 1). \square

2.3. Combinatory logic. Combinatory logic is a theoretical model of computation introduced by Moses Schönfinkel in [17] and many years later rediscovered and deeply studied by Haskell Curry in [4]. For the main reference to the subject we refer to [1]. A very intelligible approach towards this subject can be found in [18]. It is a well known fact that both models, the lambda calculus and the combinatory logic, are equivalent in the sense of expressive power. It turns out, however, that these two models differ radically as regards the behavior of random terms.

Definition 2.17. Combinatory logic

(1) The set \mathcal{F} of combinatory 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 λ -terms i.e. we write $t_1 t_2 \dots t_n$ without parentheses for $(\dots (t_1 t_2) \dots t_n)$.

(2) The *reduction* on combinators is the least compatible relation \triangleright satisfying the following rules:

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

Combinatory terms can be considered as rooted binary trees whose leaves are labeled with combinators K , S and I and inner nodes are labeled with an application operation. Accordingly, every reduction rule can be seen as a transformation of combinatory trees.

Definition 2.18. A combinatory term is in *normal form* if no reduction can be performed. A term M is *normalizing* if there is a reduction sequence starting from M and ending in a normal form N . A term M is *strongly normalizing* if all reduction sequences are finite.

Definition 2.19. 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:

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

As we can see $\text{size}(t)$ is the number of inner nodes of the combinatory tree of t .

Notation 2.20. For an integer n , we denote by \mathcal{F}_n the set of combinatory terms of size n . The set \mathcal{F}_n is finite and we denote its cardinality by F_n .

3. COMBINATORIAL RESULTS

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

Definition 3.1. Let $f, g: \mathbb{N} \rightarrow \mathbb{R}$.

- (i) Functions f and g are said to be *asymptotically equal* iff $\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 1$. We denote it by $f \sim g$.
- (ii) The *asymptotic inequality* $f \gtrsim g$ holds iff there exists a function $h: \mathbb{N} \rightarrow \mathbb{R}$ such that $h \sim g$ and $f(n) \geq h(n)$ for all n .
- (iii) A function f is said to be of the *smaller order* than g iff $\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 0$. We denote it by $f \in o(g)$.
- (iv) A function f is said to be *subexponential* in n iff there exists $h: \mathbb{N} \rightarrow \mathbb{R}$ 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 $n \mapsto n \ln(n)$) 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 their generating functions (see [20] for introductory reference). This is the approach we take to determine the asymptotic fraction of certain combinatory logic trees of a given size.

The following theorem is a well-known result in the theory of generating functions. Its derivation from the Szegő Lemma (see [13]) can be found, e.g., in [22] (Theorem 22). We denote by $[z^n]\{v(z)\}$ the coefficient of z^n in the expansion of v .

Theorem 3.2. 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 $z = 1$:

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

where $w_1 \neq 0$.

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

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

3.2. Catalan numbers. We denote by $C(n)$ the n -th Catalan numbers, i.e., the number of binary trees with n inner nodes. We use the following classical result (see, for example, [7, Ch. IV.1]).

Proposition 3.3.

- $C(n+1) = \sum_{i=0}^n C(i)C(n-i)$ for $n > 0$ and $C(0) = 1$. From this we have $C(n+1) \geq \sum_{i=0}^n C(i)$.
- $C(n) = \frac{1}{n+1} \binom{2n}{n} = \prod_{i=2}^n \frac{n+i}{i}$. From this we have $\frac{C(n)}{C(n-1)} = \frac{2(2n-1)}{n+1}$
- $C(n) \sim \frac{4^n}{n^{3/2}\sqrt{\pi}}$ 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 variables 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 3.4. • $M(n, k) = C(k-1) \binom{n+k-1}{n-k+1}$.

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

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 $2k-1$ 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 a 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 method (see, e.g., [7, Ch.VII.4] for exact computations). \square

4. DENSITIES

4.1. Main notations. For any finite set A we denote by $\#A$ its cardinality. To attribute a precise meaning to sentences like “asymptotically almost all λ -terms have property P ” we use the following definition of asymptotic density.

Definition 4.1. Let $B \subset \Lambda$, assume that B contains closed terms of every large enough size. For $A \subseteq B$, if the limit

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

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

Remarks and notations

- The asymptotic density $d_B(A)$ can also be interpreted as an asymptotic probability of finding a λ -term from the class A among all λ -terms from B .
- d_B is finitely additive: if A_1 and A_2 are disjoint classes of λ -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 , meeting the condition of definition 4.1, 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

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

is not true for all classes of pairwise disjoint sets $\{A_i\}_{i \in \mathbb{N}}$. A counterexample for the equation is to take $B = \Lambda$ and A_i the singleton containing the i -th lambda term from our language under any natural enumeration 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 closed λ -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 analogy to research 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 closed λ -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 which we provide is rather imprecise but sufficient for our purpose.

Theorem 5.1. *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 closed λ -terms of size n with k head lambdas 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, $L_n \geq LB(n, k)$ for all $k = 1, \dots, n$. Let $k = \left\lceil \frac{n}{\ln(n)} \right\rceil$. Then we get:

$$\begin{aligned} L_n &\geq C \left(n - \left\lceil \frac{n}{\ln(n)} \right\rceil \right) \left(\left\lceil \frac{n}{\ln(n)} \right\rceil \right)^{n - \left\lceil \frac{n}{\ln(n)} \right\rceil + 1} \\ &\sim \frac{4^{n - \left\lceil \frac{n}{\ln(n)} \right\rceil}}{\left(n - \left\lceil \frac{n}{\ln(n)} \right\rceil \right)^{3/2} \sqrt{\pi}} \left(\left\lceil \frac{n}{\ln(n)} \right\rceil \right)^{n - \left\lceil \frac{n}{\ln(n)} \right\rceil + 1} && \text{by Proposition 3.3} \\ &\gtrsim \left(\frac{4n}{\ln(n)} \right)^{n - \frac{n}{\ln(n)}} \frac{1}{p(n)} && \text{for some positive polynomial } p \\ &\gtrsim \left(\frac{(4 - \varepsilon)n}{\ln(n)} \right)^{n - \frac{n}{\ln(n)}} && \text{since } \left(\frac{4}{4 - \varepsilon} \right)^{n - \frac{n}{\ln(n)}} \gtrsim p(n). \end{aligned}$$

□

5.2. Number of lambdas in a term. In this part we focus on the number of unary and binary nodes in random λ -terms. We need the following lemma:

Lemma 5.2. *For all sufficiently large n , the function $f(p) = p^{n-p+1}$ is*

- (i) *decreasing on $[\frac{3n}{\ln(n)}, +\infty)$,*
- (ii) *increasing on $(0, \frac{n}{3\ln(n)}]$.*

Proof. Let us start by computing the derivative of the function f on $(0, +\infty)$:

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

- (i) We want to show that $f'(p) < 0$ for any $p \in \left[\frac{3n}{\ln(n)}, +\infty \right)$. 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. The last inequality is obvious for sufficiently large n .

- (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

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

which finishes the proof. The last inequality is obvious for sufficiently large n . \square

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

Notation 5.3. Let \mathcal{A} denote the class of closed terms $t \in \mathcal{A}$ that satisfies 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 5.4. *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 the number of considered terms: $p^{n-p+1}M(n, n - p + 1)$.

Now, we show that each of properties (i)–(iii) characterizing the class \mathcal{A} is valid for random terms. Obviously, property (iii) implies property (ii), but our proof of (iii) uses (ii) as intermediate result so we make it explicit.

- (i) Let P_n denote the number of closed terms of size n containing more than $\frac{3n}{\ln(n)}$ lambdas.

$$\text{We have } P_n \leq \sum_{p \geq \frac{3n}{\ln(n)}} p^{n-p+1} M(n, n - p + 1).$$

By Lemma 5.2 the function $p \mapsto p^{n-p+1}$ is decreasing in the interval $\left[\frac{3n}{\ln(n)}, n\right]$. Thus,

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

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 3.4, $M(n) \sim \left(\frac{1}{3-2\sqrt{2}}\right)^n \frac{1}{\sqrt{\pi n}^{\frac{3}{2}}}$. Using this equivalence, we deduce that there is some positive constant γ such that we have:

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

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, which finishes the proof.

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

By Lemma 5.2 the function $p \mapsto p^{n-p+1}$ is increasing in the interval $\left[0, \frac{n}{3\ln(n)}\right]$. Thus,

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

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

$$\begin{aligned} \frac{R_n}{L_n} &\lesssim \frac{M(n) \left(\frac{n}{3\ln(n)} \right)^{n+1-\frac{n}{3\ln(n)}}}{\left(\frac{(4-\varepsilon)n}{\ln(n)} \right)^{n-\frac{n}{\ln(n)}}} \\ &\lesssim \gamma \frac{\left(\frac{1}{3-2\sqrt{2}} \right)^n \left(\frac{n}{3\ln(n)} \right)^{n+1-\frac{n}{3\ln(n)}}}{n^{\frac{3}{2}} \left(\frac{(4-\varepsilon)n}{\ln(n)} \right)^{n-\frac{n}{\ln(n)}}} && \text{for some } \gamma > 0 \\ &\lesssim \frac{\left(\frac{1}{3-2\sqrt{2}} \right)^n \left(\frac{n}{3\ln(n)} \right)^{n-\frac{n}{3\ln(n)}}}{\left(\frac{(4-\varepsilon)n}{\ln(n)} \right)^{n-\frac{n}{\ln(n)}}} && \text{since } \frac{\gamma n}{3\ln(n)} \lesssim n^{\frac{3}{2}} \\ &= \left(\frac{1}{3(4-\varepsilon)(3-2\sqrt{2})} \right)^n \left(\frac{3(4-\varepsilon)^3 n^2}{(\ln(n))^2} \right)^{\frac{n}{3\ln(n)}} \\ &= \left(\frac{e^{2/3}}{3(4-\varepsilon)(3-2\sqrt{2})} \right)^n \left(\frac{3(4-\varepsilon)^3}{(\ln(n))^2} \right)^{\frac{n}{3\ln(n)}} && \text{since } n^{\frac{2n}{3\ln(n)}} = e^{\frac{2}{3}n}. \end{aligned}$$

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

- (iii) Let S_n be the number of closed terms of size n with more than $\frac{n}{3\ln(n)}$ lambdas and with the unary height less than $\frac{n}{3\ln(n)}$. Such a term has at most $n - \frac{n}{3\ln(n)} + 1$ leaves and each of them can be bound by one of 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. \square

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

Lemma 5.5. *Let $\alpha(n)$ be either $n \mapsto \lceil \frac{3n}{\ln(n)} \rceil$ or $n \mapsto \lfloor \frac{3n}{\ln(n)} \rfloor$. Then the function $n \mapsto \binom{3n}{\alpha(n)}$ is subexponential.*

Proof. Using the Stirling formula

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

we obtain, for some polynomial function $\gamma(n)$, the asymptotic majoration:

$$\begin{aligned} \binom{3n}{\alpha(n)} &\lesssim \gamma(n) \frac{(3n)^{3n}}{\left(3n - \left\lfloor \frac{3n}{\ln(n)} \right\rfloor\right)^{3n - \left\lfloor \frac{3n}{\ln(n)} \right\rfloor} \left(\left\lceil \frac{3n}{\ln(n)} \right\rceil\right)^{\left\lceil \frac{3n}{\ln(n)} \right\rceil}} \\ &\lesssim \gamma(n) E(n) \end{aligned}$$

where $E(n)$ can be written

$$E(n) = \frac{3^{3n}}{\left(3 - \frac{\left\lfloor \frac{3n}{\ln(n)} \right\rfloor}{n}\right)^{3n - \left\lfloor \frac{3n}{\ln(n)} \right\rfloor} \left(\frac{\left\lceil \frac{3n}{\ln(n)} \right\rceil}{n}\right)^{\left\lceil \frac{3n}{\ln(n)} \right\rceil}}$$

Let us compute the logarithm of $E(n)$:

$$\begin{aligned} \ln(E(n)) &= 3n \ln(3) - \left(3n - \left\lfloor \frac{3n}{\ln(n)} \right\rfloor\right) \ln\left(3 - \frac{\left\lfloor \frac{3n}{\ln(n)} \right\rfloor}{n}\right) - \left\lceil \frac{3n}{\ln(n)} \right\rceil \ln\left(\frac{\left\lceil \frac{3n}{\ln(n)} \right\rceil}{n}\right) \\ &\leq 3n \ln(3) - \left(3n - \frac{3n}{\ln(n)} - 1\right) \ln\left(3 - \frac{3}{\ln(n)} - \frac{1}{n}\right) - \left(\frac{3n}{\ln(n)} + 1\right) \ln\left(\frac{3}{\ln(n)}\right) \end{aligned}$$

After some simplifications we obtain that $\ln(E(n)) \lesssim 3n \frac{\ln \ln(n)}{\ln(n)} + o\left(n \frac{\ln(\ln(n))}{\ln(n)}\right)$. Since the polynomial function $\gamma(n)$ belongs to $o\left(e^{\alpha n \frac{\ln(\ln(n))}{\ln(n)}}\right)$ for any positive α , we finally deduce that:

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

Theorem 5.6. *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 $\frac{n}{3 \ln(n)}$ lambdas. According to Theorem 5.4 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 greater by one. We compute the upper bound for T_n in the following way:

- first, we consider binary trees built on at most $n - \lfloor \frac{n}{3 \ln(n)} \rfloor$ binary nodes — their number does not exceed Catalan number $C\left(n - \lfloor \frac{n}{3 \ln(n)} \rfloor + 1\right)$ (the +1 in the argument is obtained through Proposition 3.3 because we sum $C(i)$ over all possible i up to $n - \lfloor \frac{n}{3 \ln(n)} \rfloor$),
- then, we insert in such trees at most $\frac{3n}{\ln(n)}$ (the maximum number of lambdas) unary nodes — this can be done in less than $\binom{\lceil \frac{3n}{\ln(n)} \rceil}{\lfloor \frac{3n}{\ln(n)} \rfloor}$ ways ($3n - \lfloor \frac{3n}{\ln(n)} \rfloor$ is an upper bound for the number of possible places for insertions into 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 be bound by at most $\frac{3n}{\ln(n)}$ lambdas — thus the number of possible ways of binding is not greater than $\left(\frac{3n}{\ln(n)}\right)^{n+1 - \frac{n}{3 \ln(n)}}$.

Thus, we get

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

Using the asymptotic expansion of Catalan numbers (Proposition 3.3), we obtain

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

for any $\varepsilon > 0$. The last line follows from the fact that $\binom{\lceil \frac{3n}{\ln(n)} \rceil}{\lfloor \frac{3n}{\ln(n)} \rfloor}$ is subexponential (by Lemma 5.5). \square

Remark. The ratio between the upper and lower bounds obtained for L_n is exponential, but L_n is super-exponential itself.

6. PROOFS USING CODING

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

We consider successive sets of terms X_1, \dots, X_k with $X_{i+1} \subseteq X_i$ and we prove:

- (1) X_1 has density 1 (Theorem 5.4);
- (2) X_{i+1} has density 1 because $X_i \setminus X_{i+1}$ has density 0 (successive theorems of this section).

By choice of X_k , we finally get that SN terms have density 1. Below, these sets X_1, X_2, \dots are denoted $\mathcal{A}, \mathcal{B}, \dots$ and depend on some parameters (integers or functions).

Some proofs need the following lemma:

Lemma 6.1. *Let A_n be a sequence of non empty finite sets of terms and B_n be subsets of A_n . Let $(A_{n,i})_{i \in I_n}$ be a partition of A_n and let $B_{n,i} = A_{n,i} \cap B_n$. 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:*

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

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:

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

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

6.1. The number of lambdas in head position. We start with showing that a random term starts with a long chain of lambdas. In the next theorem and until the end of the paper, we denote by g a lower bound on the length of this chain (as a function of the size of the term). Theorem 6.3 below shows that any $g \in o(\sqrt{n/\ln(n)})$ is an admissible lower bound. However, the reader can think of g as the function $n \mapsto \ln(n)^2 + 3$ since the main theorem (Theorem 6.18) and all intermediate results can be proved using this particular choice of g (see Proposition 6.17).

Notation 6.2. Let $g: \mathbb{N} \rightarrow \mathbb{N}$ We define \mathcal{B}^g as the class of terms t such that

1. $t \in \mathcal{A}$ (see Notation 5.3),
2. t has at least $g(\text{size}(t))$ 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 $\overline{\mathcal{B}_n^g}$ the set of terms from $\overline{\mathcal{B}^g}$ of size n .

Theorem 6.3. *Let $g: \mathbb{N} \rightarrow \mathbb{N}$ be a function such that $g \in o(\sqrt{n/\ln(n)})$. The density of \mathcal{B}^g in Λ is 1.*

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

Let $n_0 > 1$ be such that $g(n) < \frac{n}{3\ln(n)}$ for all $n \geq n_0$. Such n_0 exists because $g \in o(\sqrt{n/\ln(n)})$. In the rest of the proof we always assume that $n \geq n_0$.

We define a partition of $\overline{\mathcal{B}_n^g}$ as follows (see Figure 3). Let \vec{t} be a non-empty sequence of (not necessarily closed) terms such that each of the elements of \vec{t} starts with a λ . Let $\ell \geq 1$ be an integer such that $0 \leq n - \ell - \text{size}(\vec{t}) \leq g(n)$, where $\text{size}(\vec{t})$ denotes the sum of sizes of its components. We define $\overline{\mathcal{B}_n^g}(\vec{t}, \ell)$ as the set of terms of the form:

$$\lambda x_1 \dots \lambda x_{p \cdot v} [t_1, \dots, t_k]$$

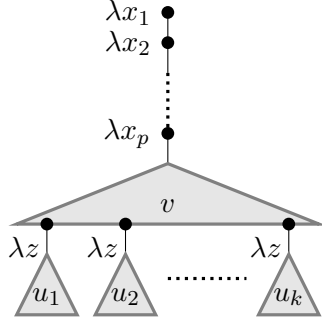


Figure 3: A term from $\overline{\mathcal{B}}_n^g(\vec{t}, \ell)$ where $\vec{t} = (\lambda z.u_1, \dots, \lambda z.u_k)$

where v is a purely applicative context with k holes, $\vec{t} = (t_1, \dots, t_k)$ and $p = n - \ell - \text{size}(\vec{t})$. Therefore, ℓ is the size of the applicative context v (where the hole are counted with size 0 like variables).

First, it is clear that nonempty sets $\overline{\mathcal{B}}_n^g(\vec{t}, \ell)$ form a partition of $\overline{\mathcal{B}}_n^g$: they are pairwise disjoint by definition and every $u \in \overline{\mathcal{B}}_n^g$ belongs to \mathcal{A} so it contains some λ not in the chain of head lambdas (because $p \leq g(n) < \frac{n}{3 \ln(n)}$), therefore it belongs to some $\overline{\mathcal{B}}_n^g(\vec{t}, \ell)$ for some non-empty \vec{t} and some $\ell \geq 1$.

Terms from $\overline{\mathcal{B}}_n^g(\vec{t}, \ell)$ differ only by applicative contexts, so the cardinality of $\overline{\mathcal{B}}_n^g(\vec{t}, \ell)$ 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 for all $n \geq n_0$:

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

Let $t \in \overline{\mathcal{B}}_n^g(\vec{t}, \ell)$ and $\vec{t} = (t_1, \dots, t_k)$ for some $k \geq 1$ and v be the purely applicative context in the decomposition of t . We can write $t_i = \lambda z.u_i$. Consider the term

$$t' = \lambda z \lambda x_1 \dots \lambda x_p.(u_1 (u_2 (\dots (u_{k-1} u_k) \dots)))$$

which is of size

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

We rename bound variables, so that a variable distinct from z in t is renamed to x_k where k is number of lambdas from the root to the lambda binding that variable (inclusive). Let V_n be the set of variables $\{x_1, \dots, x_{\lfloor \frac{n}{3 \ln(n)} \rfloor}\}$. Let $\lambda y.s$ denote the term rooted at the leftmost deepest λ of term t' .

Since the unary height of t is the same as that of t' , and since $t \in \mathcal{A}$, all the variables in V_n are bound on the path from the root to $\lambda y.s$ (in the worst case, y is $x_{\lfloor \frac{n}{3 \ln(n)} \rfloor}$ and must also be counted on the path).

Let $U_{n,\ell}$ be the set of purely applicative (therefore not closed) terms of size $\ell - 1$ whose variables are chosen from V_n . There are at least

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

elements in $U_{n,l}$.

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

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

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

From this, for n large enough, we get $P(n, \ell) < Q(n, \ell)$ (uniform convergence of ψ is needed only later) and there exists an injective function $h_{n,\ell}$ which assigns an element from $U_{n,l}$ to any purely applicative context using variables in $\{x_1, \dots, x_p\}$ (i.e. applicative context v used in the decomposition of a term in $\overline{\mathcal{B}}_n^g(\vec{t}, \ell)$).

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

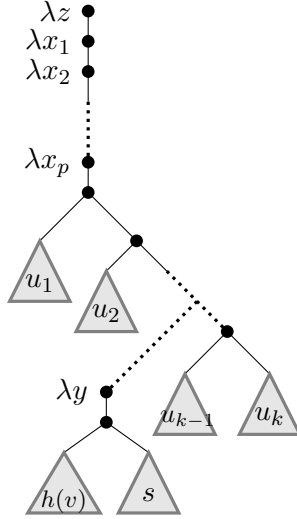


Figure 4: The term $\varphi_n^{\mathcal{B}}(t)$ from Theorem 6.3

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

Let $\varphi_n^{\mathcal{B}} = \bigcup_{\ell, \vec{t}} \varphi_{n,\vec{t},\ell}$. The function $\varphi_n^{\mathcal{B}}$ is an injection because codomains of the $\varphi_{n,\vec{t},\ell}$ are all disjoint by construction. Since the sets $\overline{\mathcal{B}}_n^g(\vec{t}, \ell)$ form a partition of $\overline{\mathcal{B}}_n^g$, by means of Lemma 6.1, it is enough to show that $\frac{P(n, \ell)}{Q(n, \ell)}$ tends uniformly in l to 0 as n tends to infinity, which was done above. \square

6.2. Head lambdas bind “many” occurrences. Now we are ready to present some theorems showing that in a random term head lambdas are used, i.e. they really bind some variables. The first result shows that in a random term many of head lambdas are binding.

Notation 6.4. Let $g: \mathbb{N} \rightarrow \mathbb{N}$ be a function such that $g \in o(\sqrt{n/\ln(n)})$. 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 $n \mapsto g(n) + 1$,
2. each of first $g(\text{size}(t))$ head lambdas in t is binding.

Additionally, we denote by $\overline{\mathcal{D}}_n^g = \mathcal{B}^{g+1} \setminus \mathcal{D}^g$ the complement 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 6.5. Let $g: \mathbb{N} \rightarrow \mathbb{N}$ be a function such that $g \in o(\sqrt{n/\ln(n)})$. The density of \mathcal{D}^g in Λ is 1.

Proof. Let us fix $g \in o(\sqrt{n/\ln(n)})$. We construct a family of codings $\varphi_n^{\mathcal{D}}: \overline{\mathcal{D}}_n^g \rightarrow \Lambda_n$ such that their images are negligible in Λ_n , i.e. the fraction $\#\varphi_n^{\mathcal{D}}(\overline{\mathcal{D}}_n^g)/L_n$ 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 lambda in t does not bind any variable. Take

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

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 $g(n) + 1$ head lambdas. By Theorem 6.3, the density of such terms in Λ is zero. Since the function $\varphi_n^{\mathcal{D}}$ is injective, the density of $\overline{\mathcal{D}}^g$ is also zero. \square

Notation 6.6. Let $g, h: \mathbb{N} \rightarrow \mathbb{N}$ be functions such that $g \in o(\sqrt{n/\ln(n)})$, $g(n) \geq 3$ 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 the first three lambdas in t is greater than $h(\text{size}(t))$.

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

Theorem 6.7. Let $g, h: \mathbb{N} \rightarrow \mathbb{N}$ be functions such that $g \in o(\sqrt{n/\ln(n)})$, $g(n) \geq 3$ 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 $\varphi_n^{\mathcal{E}}: \overline{\mathcal{E}}_n^{g,h} \rightarrow \Lambda_n$ 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 lambdas 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 $\overline{\mathcal{E}}_n^{g,h}$. There are at most $3^{h(n)}$ elements in the class $[t]$.

Let $\psi(t) = \lambda xy.u[x_1 := y, x_2 := y, x_3 := y]$. The size of $\psi(t)$ is $n - 1$. Let $\lambda a.v$ be the subterm of $\psi(t)$ such that λa is the leftmost deepest λ in $\psi(t)$. Denote by $V(t)$ the set of variables introduced by lambdas occurring in $\psi(t)$ on the path from λa to λy . Note that the variable x occurs neither in $\psi(t)$ nor in $V(t)$.

By Theorem 5.4(iii), there are at least $\frac{n}{3 \ln(n)} - 2$ such lambdas. As $h \in o\left(\log_3\left(\frac{n}{\ln(n)}\right)\right)$, we have

$$\lim_{n \rightarrow \infty} \frac{3^{h(n)}}{\left(\frac{n}{3 \ln(n)} - 2\right)} = 0.$$

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

We define $\varphi_n^{\mathcal{E}}(t)$ as the term obtained from $\psi(t)$ by replacing the subterm $\lambda a.v$ with $\lambda a.(w v)$, where $w = f_{[t]}(t)$.

All terms from the image $\varphi_n^{\mathcal{E}}(\overline{\mathcal{E}_n^{g,h}})$ start with a λ that binds no variable. By Theorem 6.5 we know that such terms are negligible in Λ_n . Since $\varphi_n^{\mathcal{E}}$ is injective, the density of $\overline{\mathcal{E}_n^{g,h}}$ is zero, as well. \square

Notation 6.8. Let k and ℓ be natural numbers. Let $g: \mathbb{N} \rightarrow \mathbb{N}$ be functions such that $g \in o(\sqrt{n/\ln(n)})$, $g(n) \geq 3$ for all n , $\lim_{n \rightarrow \infty} g(n) = \infty$, 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 complement 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 6.9. *Let k and ℓ be integers. Let $g: \mathbb{N} \rightarrow \mathbb{N}$ be a function such that $g \in o(\sqrt{n/\ln(n)})$, $\lim_{n \rightarrow \infty} g(n) = \infty$, 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 6.7, 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 $\mathcal{E}_n^{g,h}(m, k)$ the set of terms from $\mathcal{E}_n^{g,h}$ with exactly m leaves bound by the 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 definition, terms from $\overline{\mathcal{G}_n^{g,k,\ell}}(m)$ have exactly m leaves bound by the first k lambdas and at least one of these lambdas binds at most ℓ variables.

Consider the equivalence relation \sim_n on $\mathcal{E}_n^{g,h}(m, k)$ defined analogously to the relation with the same notation within the proof of Theorem 6.7, but with respect to the first k (instead of three) head lambdas. Denote by $[t]$ the equivalence class of t for this relation.

Let $t \in \mathcal{E}_n^{g,h}(m, k)$. By hypothesis on g and for large enough n , each of the first k head lambdas of t are binding. Of the m leaves bound by these lambdas, give the k leftmost leaves distinct labels. 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 lambdas which binds ℓ' leaves with $1 \leq \ell' \leq \ell$ (we can choose them out of m ones) and the other leaves (their number is equal to $m - \ell' \leq m - 1$) can be bound by $k - 1$ lambdas. Thus, we obtain the upper bound equal to $\sum_{1 \leq \ell' \leq \ell} k \binom{m}{\ell'} (k - 1)^{m - \ell'} \leq km^\ell (k - 1)^{m - 1}$. This holds because $\sum_{1 \leq \ell' \leq \ell} \binom{m}{\ell'} \leq m^\ell$ which can be proved by induction over ℓ when $m \geq 2$ (here $m \geq k \geq 3$).

Therefore, the quotient of the two cardinalities is less than

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

As n tends to infinity, the above quotient tends to 0 uniformly in m . To establish this, we define $f(x) = x^\ell R^{\frac{x}{2}}$ with $R = \frac{k - 1}{k} < 1$. Thus we have

$$k^k m^\ell \left(\frac{k - 1}{k} \right)^{m - 1} \leq k^k f(m) R^{\frac{m}{2} - 1}$$

Then, $f'(x) = x^{\ell - 1} R^{\frac{x}{2}} (l + x \frac{\ln(R)}{2})$ and we see that $f(x)$ reaches a maximum on \mathbb{R}_+ for $x = A = -\frac{2\ell}{\ln(R)}$ (which is a positive constant because $R = \frac{k - 1}{k} < 1$), which gives:

$$k^k m^\ell \left(\frac{k - 1}{k} \right)^{m - 1} \leq k^k f(A) R^{\frac{m}{2} - 1} \leq k^k f(A) R^{\frac{h(n)}{2} - 1}$$

For $t \in \Lambda_n$ and $m \geq h(n)$ the sets $[t] \cap \overline{\mathcal{G}_n^{g,k,\ell}}(m)$ form a partition of $\overline{\mathcal{G}_n^{g,k,\ell}}$. Now Lemma 6.1 finishes the proof. \square

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

Notation 6.10. Let k and ℓ be positive integers. Let $g: \mathbb{N} \rightarrow \mathbb{N}$ be a function such that $g \in o(\sqrt{n/\ln(n)})$, $\lim_{n \rightarrow \infty} g(n) = \infty$, 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 $\overline{\mathcal{H}^{g,k,\ell}} = \mathcal{G}^{g,k,\ell} \setminus \mathcal{H}^{g,k,\ell}$ the complement of the class $\mathcal{H}^{g,k,\ell}$ in $\mathcal{G}^{g,k,\ell}$ and by $\overline{\mathcal{H}_n^{g,k,\ell}}$ the set of terms from $\overline{\mathcal{H}^{g,k,\ell}}$ of size n .

Lemma 6.11. *Let k and ℓ be positive integers. Let $g: \mathbb{N} \rightarrow \mathbb{N}$ be a function such that $g \in o(\sqrt{n/\ln(n)})$, $\lim_{n \rightarrow \infty} g(n) = \infty$, 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 $\varphi_n^{\mathcal{H}}$ from $\overline{\mathcal{H}_n^{g,k,\ell}}$ into the set of terms whose leading λ binds only one variable occurrence.

Let t be a term from $\overline{\mathcal{H}_n^{g,k,\ell}}$. Let t_1 be a subterm rooted at a highest leftmost 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 $\varphi_n^{\mathcal{H}}(t) = \lambda z. t'$. The result follows from Theorem 6.9. \square

6.3. A random term avoids any fixed closed term.

Notation 6.12. 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 $g: \mathbb{N} \rightarrow \mathbb{N}$ be a function such that $g \in o(\sqrt{n/\ln(n)})$, $g(n) \geq 3$ for all n and $\lim_{n \rightarrow \infty} g(n) = \infty$. 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 $\bigcup_{i \leq j} \Lambda_i$ 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 complement of the class $\mathcal{I}^{g,j}$ in $\mathcal{H}^{g,k(j),k(j)}$ and by $\overline{\mathcal{I}_n^{g,j}}$ the set of terms from $\overline{\mathcal{I}^{g,j}}$ of size n .

Theorem 6.13. *Let j be a positive integer and let $g: \mathbb{N} \rightarrow \mathbb{N}$ be a function such that $g \in o(\sqrt{n/\ln(n)})$, $g(n) \geq 3$ for all n and $\lim_{n \rightarrow \infty} g(n) = \infty$. 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 $\varphi_n^{\mathcal{I}}: \overline{\mathcal{I}_n^{g,j}} \rightarrow \Lambda_n$ 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 from $\bigcup_{i \leq j} \Lambda_i$ to $\{1, \dots, k(j)\}$.

Let n be an integer satisfying $g(n) > k(j)$ and $n > k(j) + j$. Let $t \in \overline{\mathcal{I}_n^{g,j}}$. By hypothesis the term t belongs to \mathcal{B}^{g+1} , so it has more than $k(j)$ head lambdas since $k(j) < g(n)$ (see Figure 5).

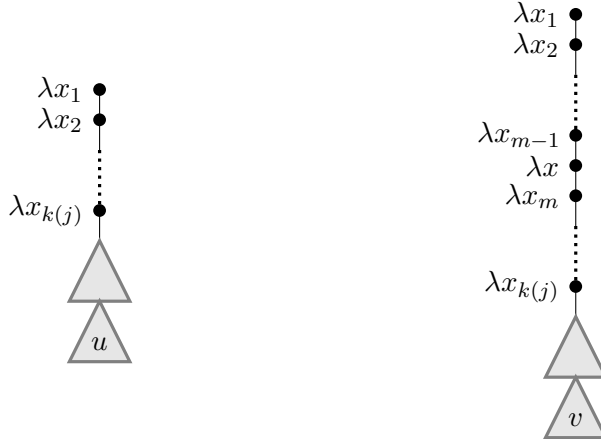


Figure 5: Terms $t \in \overline{\mathcal{I}_n^{g,j}}$ and $\varphi_n^{\mathcal{I}}(t)$ from Theorem 6.13

In term t , consider the smallest m such that $f(u) = m$ for some closed u occurring in t (there is at least one such m because $t \in \overline{\mathcal{I}_n^{g,j}}$). 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 $\varphi_n^{\mathcal{I}}(t)$ obtained by replacing the leftmost deepest occurrence of subterm u in s by the term $v = (x(x(\dots(x)x)\dots))$ of size $i - 1$ where i is the size of u (see Figure 5). Thus, the size of $\varphi_n^{\mathcal{I}}(t)$ is equal to n .

By Theorem 6.9, each of the first $k(j)$ head lambdas in a term from $\mathcal{H}^{g,k(j),k(j)}$ of size n binds more than $k(j)$ variables. Therefore, among the first $k(j)$ head lambdas of $\varphi_n^{\mathcal{I}}(t)$,

only the m -th λ binds less than $k(j)$ variables (recall that u is closed which means that the number of variables bound by λx_i for $1 \leq i \leq k(j)$ is the same in t and $\varphi_n^{\mathcal{I}}(t)$). Since $f(u) = m$ and f is injective, the function $\varphi_n^{\mathcal{I}}$ is injective. Terms from the image $\varphi_n^{\mathcal{I}}(\overline{\mathcal{I}_n^{g,j}})$ are not in $\mathcal{H}^{g,k(j),k(j)}$ since the m -th λ binds only $i \leq j \leq k(j)$ variables. Thus, those terms are negligible among all terms of size n . \square

6.4. The λ -width of a term. Let us recall that λ -width of a term is the maximum number of incomparable binding lambdas in the term. In the following proposition we show that λ -width of typical λ -terms is small.

Notation 6.14. Let $g: \mathbb{N} \rightarrow \mathbb{N}$ be a function such that $g \in o(\sqrt{n/\ln(n)})$, $g(n) \geq 3$ for all n and $\lim_{n \rightarrow \infty} g(n) = \infty$. By \mathcal{J}^g we denote the class of closed terms such that $t \in \mathcal{J}^g$ iff

1. $t \in \mathcal{G}^{g,1,4}$
2. λ -width of t is at most 2.

Additionally, we denote by $\overline{\mathcal{J}^g} = \mathcal{G}^{g,1,4} \setminus \mathcal{J}^g$ the complement of the class \mathcal{J}^g in $\mathcal{G}^{g,1,4}$ and by $\overline{\mathcal{J}_n^g}$ the set of terms from $\overline{\mathcal{J}^g}$ of size n .

Theorem 6.15. Let $g: \mathbb{N} \rightarrow \mathbb{N}$ be a function such that $g \in o(\sqrt{n/\ln(n)})$, $g(n) \geq 3$ for all n and $\lim_{n \rightarrow \infty} g(n) = \infty$. The density of \mathcal{J}^g in Λ is 1.

Proof. Let us fix a function g as in the assumptions of the theorem. We construct a family of codings $\varphi_n^{\mathcal{J}}: \overline{\mathcal{J}_n^g} \rightarrow \Lambda_n$ such that their images are negligible in Λ_n . Let t be an element of $\overline{\mathcal{J}_n^g}$, therefore the λ -width of t is at least 3. Let us denote by λx , λy and λz the first three highest leftmost pairwise incomparable binding lambdas (appearing in this order from left to right in t).

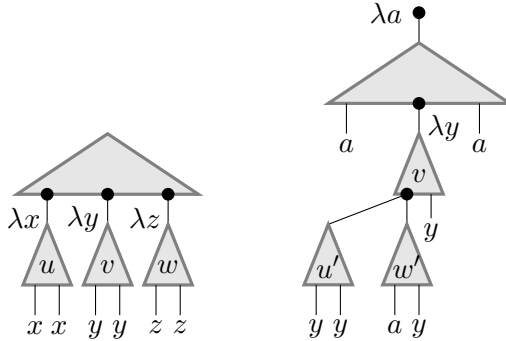


Figure 6: The terms t and $\varphi_n^{\mathcal{J}}(t)$ from Theorem 6.15

Let $\lambda x.u$, $\lambda y.v$ and $\lambda z.w$ be subterms rooted at those lambdas (see Figure 6). Let $u' = u[x := y]$, let a be a new variable, and 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 $\text{size}(\varphi_n^{\mathcal{J}}(t)) = \text{size}(t)$. Also note that since we choose the highest three incomparable binding lambdas 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 ν_l (resp. ν_r) be the deepest node above the two leftmost (resp. rightmost) occurrences of a . Remark that since there are exactly 3 occurrences of a , one of these two nodes is above the other. Let ν be the deepest one. λy is the first binding λ on the path from the node ν 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 6.9, the density of $\varphi_n^{\mathcal{J}}(\overline{\mathcal{J}}_n^g)$ is equal to zero. The injectivity of $\varphi_n^{\mathcal{J}}$ finishes the proof. \square

6.5. The density of strongly normalizable terms. From Theorem 6.15 (using $g(n) = \ln(n)^2 + 3$ for instance) we know that almost all terms are of width at most 2. In Section 2 we introduced the notion of 'safe' terms of width 2 which implies strong normalization (Proposition 2.16).

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

Notation 6.16. Let $g: \mathbb{N} \rightarrow \mathbb{N}$ be a function such that $g \in o(\sqrt{n/\ln(n)})$, $g(n) \geq 3$ for all n and $\lim_{n \rightarrow \infty} g(n) = \infty$. By \mathcal{K}^g we denote the class of closed terms such that $t \in \mathcal{K}^g$ iff

1. $t \in \mathcal{J}^g$,
2. t is safe.

Additionally, we denote by $\overline{\mathcal{K}}^g = \mathcal{J}^g \setminus \mathcal{K}^g$ the complement of the class \mathcal{K}^g in \mathcal{J}^g and by $\overline{\mathcal{K}}_n^g$ the set of terms from $\overline{\mathcal{K}}^g$ of size n . Note that terms from $\overline{\mathcal{K}}^g$ are of λ -width at most 2 and are unsafe, therefore they are of width exactly 2 (because terms of width 1 are safe by definition).

Proposition 6.17. *Let $g: \mathbb{N} \rightarrow \mathbb{N}$ be the function defined by $g(n) = \ln(n)^2 + 3$. The density of \mathcal{K}^g 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 $\overline{\mathcal{K}}^g$ in Λ is 0. Let us divide the set $\overline{\mathcal{K}}_n^g$ into two disjoint subsets:

$\overline{\mathcal{K}}_n^{g,1}$: the set of terms of size n such that neither of the lengths of paths from the branching node to the two highest incomparable binding lambdas is greater than $\ln(n)$,

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

We can construct a family of codings from the set $\overline{\mathcal{K}}_n^{g,1}$ into Λ_n in the following two steps:

- (1) Remove the 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.

- (2) Insert one non-binding lambda among the head lambdas of the term. By choice of g and by definition of \mathcal{K}^g , terms from $\overline{\mathcal{K}_n^{g,1}}$ have more than $\ln(n)^2$ head lambdas. Therefore we can encode the lengths of the paths from the branching node to the two highest binding lambdas as the position of this new lambda. By Theorem 6.7 the image of such a transformation has density 0.

For the set $\overline{\mathcal{K}_n^{g,2}}$ we do not construct an injection, but a relation that associates to terms in $\overline{\mathcal{K}_n^{g,2}}$ disjoint set of terms of cardinals greater than $\ln(n)/2$. This is enough to show that $\overline{\mathcal{K}_n^{g,2}}$ has density 0. Precisely, we proceed as follows:

- (1) Choose the leftmost path among the one or two paths longer than $\ln(n)$ (without loss of generality we can assume it is the left path, the case of the right one is analogous). Consider the binding lambda at the end of this path and let t_0 be the subterm rooted at this lambda. Let t_1, \dots, t_k be the right subtrees rooted at the binary nodes on the path between the branching node and t_0 (the path goes always to the left since the term is unsafe). By Lemma 6.11 at least half of the nodes on this path are binary (since there are no two consecutive non-binding lambdas in the tree). This means that $k \geq \ln(n)/2$. Moreover, the terms t_1, \dots, t_k contain no binding lambda otherwise, the lambda width of the term would be greater than 2.
- (2) Choose some leaf x belonging to some subtree among t_1, \dots, t_k and exchange it with the subterm t_0 . Independently of the choice of the leaf, the encoding can be reversed since:
- (a) the position of t_0 in the encoded term is uniquely identifiable as the highest binding lambda of the innocuous subtree below the branching node (the innocuousness identifies the modified branch);
 - (b) the position of the variable x in the encoded term is identifiable as the leftmost leaf of the subtree rooted at the branching node of the resulting term which is still of width 2 (in the case of the right branch, it is the leftmost leaf of the right sub-term of the branching node).

The encoding preserves size and the number of possibilities for the choice of a leaf x is the number of leafs of t_1, \dots, t_k , which is greater than $\ln(n)/2$. Therefore, terms from $\overline{\mathcal{K}_n^{g,2}}$ are negligible in Λ_n as n tends to infinity. \square

Main Theorem 6.18. *The set of strongly normalizable terms has density 1.*

Proof. Proposition 6.17 shows the existence of a set of safe terms that has density 1. Proposition 2.16 shows that they are all strongly normalizable. \square

7. COMBINATORY LOGIC

In this section we show that our main result about strong normalization of random λ -terms does not hold in the world of random combinatory terms. On the contrary, a random combinatory 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 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 the n -th Catalan number (see Proposition 3.3).

Proposition 7.1.

- (1) *The generating function f enumerating cardinality of the set of combinatorics (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 cardinality of the set of all combinatorics 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 combinatorics of size n . Since there are three combinatorics of size 0, we have $F_0 = 3$. Combinatorics of size $n \geq 1$ are built of two combinatorics of sizes i and $n - i - 1$ ($i = 0, \dots, n - 1$), respectively, 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 $\lim_{z \rightarrow 0} f(z) = 3$. Thus, the desired generating function is given by the first solution.

- (2) Let t be a combinator having t_0 as a subterm. Then either t is equal to t_0 or t 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:

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

which can be simplified to

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

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, $\lim_{z \rightarrow 0} f_{t_0}(z) = 0$. The first function satisfies this condition, so this is the wanted generating function. \square

The following theorem shows that the result similar to Theorem 6.13 is not valid in combinatorial logic.

Theorem 7.2. *Let t_0 be a combinator. The density of combinatorics having t_0 as a subterm is 1.*

Proof. We prove this result applying Theorem 3.2. We start by normalizing the functions f_{t_0} and f in such a way that the closest singularity to the origin is located at $z = 1$. 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

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

Since $\frac{\sqrt{1-z+4\left(\frac{z}{12}\right)^{n_0+1}}}{2}$ is analytic for $|z| \leq 1$, the representation above reveals that the only singularity of $\overline{f_{t_0}}(z)$ and $\overline{f}(z)$ located in $|z| \leq 1$ is indeed at $z = 1$ and both functions $\overline{f_{t_0}}$ and \overline{f} have expansions in the vicinity of $z = 1$ of forms $\sum_{p \geq 0} v_p (1-z)^{p/2}$ and $\sum_{p \geq 0} w_p (1-z)^{p/2}$, respectively, with $w_1 = -1/2 \neq 0$. We have to remember that the multiplication by z and the change 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 $(12)^{1-n} ([z^{n-1}] \{f_{t_0}(z)\})$ and $(12)^{1-n} ([z^{n-1}] \{f(z)\})$, 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:

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

By analyticity of functions $(\widetilde{f_{t_0}})'$ and $(\widetilde{f})'$ for $|z| < 1$, their derivatives in this circle exist and are as follows:

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

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 3.2, obtaining:

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

Main Theorem 7.3. *The density of non-strongly normalizing combinatorics 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 7.2, since the density of combinatorics containing Ω as a subterm is 1. \square

8. DISCUSSION

The difference between Theorem 6.18 in the λ -calculus and Theorem 7.3 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: there is a constant k such that, for all term t , $\text{size}(T_1(t)) \leq k \text{size}(t)$. 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 requires the use of many combinators.

8.1. Future work and open questions. We present here some questions left open.

- (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 for simple types.
- (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. For what notions of size do we get a density 1 as in Theorem 6.18 or a density 1 as in Theorem 7.3? Are there sizes for which the density is neither 0 nor 1?

8.2. Possible applications. It is now popular to test programs written in functional languages using randomly generated inputs [3]. For higher-order functional programs where inputs are functions, this also means the ability to generate typical functions under certain known distributions.

For many typed languages such as OCaml or Haskell, functional programs can be tested by supplying random typed λ -terms generated in compliance with their natural distribution (probably different for different types of programs).

For untyped languages such as LISP, the problem of testing programs is very close to the capability of generating pure random λ -terms. In our case, those terms automatically enjoy important properties such as strong normalization, if they do not use recursive definitions. However, it would be nice to have a distribution where terms with other computationally good properties have density 1.

One could argue that width at most 2 is a negative result since it shows that random terms do not contain any tuple of more than 2 functions, whereas 'natural' programs do contain such kind of subterms.

Anyway, results and methods presented in this paper can be used as the starting point for further research based on other notions of size which are meaningful for applications.

ACKNOWLEDGMENTS

We would like to thank the anonymous referees for their numerous, precise and fruitful remarks.

REFERENCES

- [1] H. Barendregt, *The Lambda Calculus: Its Syntax and Semantics*. Studies in Logic and The Foundations of Mathematics, vol. 103, North-Holland, 1984.
- [2] L. Boyer, G. Theyssier, *On Local Symmetries and Universality in Cellular Automata*. 26th International Symposium on Theoretical Aspects of Computer Science (STACS), Dagstuhl Seminar Proceedings, 2009, <http://stacs2009.informatik.uni-freiburg.de/proceedings.php>
- [3] K. Claessen, J. Hughes, *QuickCheck: A Lightweight Tool for Random Testing of Haskell Programs*. Proc. of International Conference on Functional Programming (ICFP), ACM SIGPLAN, pp. 268-279, 2000.
- [4] H.B. Curry, R. Feys, *Combinatory Logic*. Vol. I. Amsterdam: North Holland, 1958.
- [5] R. David, *Normalization without reducibility*. APAL 107 (2001), pp. 121-130.
- [6] R. David, *A short proof of the strong normalization of the simply typed lambda calculus*. <http://www.lama.univ-savoie.fr/~david/>
- [7] P. Flajolet, R. Sedgewick, *Analytic combinatorics*. Cambridge University Press, 2008.
- [8] 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.
- [9] A. Genitrini, J. Kozik, *Quantitative comparison of Intuitionistic and Classical logics – full propositional system*. LFCS09, Lecture Notes in Computer Science 5407, pp. 280-294, 2009.
- [10] A. Genitrini, J. Kozik, M. Zaionc, *Intuitionistic vs. Classical Tautologies, Quantitative Comparison*. Lecture Notes in Computer Science 4941, pp. 100-109, 2008.
- [11] J.D. Hamkins and A. Miasnikov, *The halting problem is decidable on a set of asymptotic probability one*. Notre Dame J. Formal Logic 47(4), pp. 515-524, 2006.
- [12] Z. Kostrzycka, M. Zaionc, *Statistics of intuitionistic versus classical logic*. Studia Logica, 76(3), pp. 307-328, 2004.
- [13] G. Szegő, *Orthogonal polynomials*. American Mathematical Society Colloquium Series Publication, 1967.
- [14] M. Moczurad, J. Tyszkiewicz, M. Zaionc, *Statistical properties of simple types*. Mathematical Structures in Computer Science, 10(5), pp. 575-594, 2000.
- [15] L. Regnier, *Une équivalence sur les lambda-termes*. Theoretical Computer Science, Volume 126(2), pp. 281-292, 1994.
- [16] A. Rybalov, *On the strongly generic undecidability of the Halting Problem*. Theoretical Computer Science, Volume 377, Issues 1-3, pp. 268-270, 31 May 2007.
- [17] M. Schönfinkel, *Über die Bausteine der mathematischen Logik*. Mathematische Annalen 92, pp. 305-316, 1924.
- [18] R.M. Smullyan, *To Mock a Mockingbird and Other Logic Puzzles: Including an Amazing Adventure in Combinatory Logic*. Knopf, 1985.
- [19] J. 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.
- [21] M. Zaionc, *On the asymptotic density of tautologies in logic of implication and negation*. Reports on Mathematical Logic, vol 39, pp. 67-87, 2005.
- [22] M. Zaionc, *Probability distribution for simple tautologies*. Theoretical Computer Science, 355(2), pp. 243-260, 2006.