Lambda terms of bounded unary height*

Olivier Bodini[†] Danièle Gardy[‡]
Bernhard Gittenberger[§]

Abstract

We aim at the asymptotic enumeration of lambda-terms of a given size where the order of nesting of abstractions is bounded whereas the size is tending to infinity. This is done by means of a generating function approach and singularity analysis. The generating functions appear to be composed of nested square roots which exhibit unexpected phenomena. We derive the asymptotic number of such lambda-terms and it turns out that the order depends on the bound of the height. Furthermore, we present some observations when generating such lambda randomly and explain why powerful tools for random generation, such as Boltzmann samplers, face serious difficulties in generating lambda-terms.

1 Introduction

Roughly speaking, a lambda-term is a formal expression built of variables and a quantifyer λ which in general occurs more than once and acts on one of the free variables of the subsequent subterm. λ-calculus is a set of rules for manipulating lambda-terms and was invented by Church and Kleene in the 30ies (see [24, 25, 9]) in order to investigate decision problems. It plays an important rôle in computability theory, for automatic theorem proving or as a basis for some programming languages, e.g. LISP. Due to its flexibility it can be used for a formal description of programming in general and is therefore an essential tool for analyzing programming languages.

Recently, there has been rising interest in random structures related to logic in general (see [30] [18], [19], and [11]) and in the properties of random lambda-terms in particular, see [10].

For analyzing the structure of random lambda-terms it is important to know the number of lambda-terms of a given size. It turns out that this is a very hard problem. For instance, translating the counting problem to generating functions, the resulting generating function has a radius of convergence equal to zero. Thus none of the classical methods of analytic combinatorics (see [16]) is applicable. Therefore we study in this paper a simpler structure, obtained by bounding the nested number of abstractions, i.e. the unary height (to be formally defined in the next section) of lambda-terms. Note that this simpler structure is indeed of practical relevance: the nested number of abstractions in lambda-terms which occur in computer programming is in general bounded. E.g., for implementing lambda-calculus we need to bound the height of the underlying stack, which is determined by the maximal allowed number of nested abstractions.

The plan of the paper is as follows: In Section 2, we formally define the objects of our interest and derive generating functions for the associated counting problems, which are expressed as a finite sequence of nested radicals. Section 3 is devoted to the the study of these nested radicals: we concentrate on the sequence of radii of convergence and on the type of their singularities. Then we are in position to determine in Section 4 the detailed asymptotic behaviour of the number of lambda-terms with fixed unary height. Finally, we investigate how our theoretical results fit with simulations and discover some challenging facts on the average behaviour of a random lambda-term in Section 5.

2 A combinatorial description for lambda-terms

A lambda-term is a formal expression which is described by the context-free grammar

$$T ::= a \ | \ (T * T) \ | \ \lambda a.T$$

where a is a variable. Concatenating terms, $(T * T)$, is called *application* and assumed to be non-commutative. Using the quantifyer is called *abstraction*. Iterated abstraction is also non-commutative, i.e., the terms $\lambda x.\lambda y.T$ and $\lambda y.\lambda x.T$ are considered to be different. Furthermore, each abstraction binds a variable and each variable can be bound by at most one abstraction. A variable which is not bound by an abstraction is called free. A lambda-term without free variables is called closed, otherwise open.

A lambda-term can be represented as an enriched tree, i.e., a graph built from a tree by adding certain directed edges (pointers): First we construct a Motzkin tree, i.e., a planar rooted tree where each node has outdegree 0, 1, or 2, if the edges are directed away from the

*Supported by ANR projects BOOLE NT09-432755 and Magnum, Austrian Science Foundation FWF, NFN-Project S9604, ÖAD, grant F03/2010, and PHC Amadeus 23091UD.

[†]LIP6, University Paris 6, and LIAFA, University Paris 7.

[‡]PRiSM, CNRS UMR 8144 and Université de Versailles Saint-Quentin, 78035 Versailles, France.

[§]Inst. of Discrete Mathematics and Geometry, T.U. Wien, Wiedner Hauptstr. 8-10/104, A-1040 Wien,Austria

root. We respectively denote by the terms leaves, unary nodes, and binary nodes, the nodes with out-degree 0, 1, and 2. In this tree each application corresponds to a binary node, each abstraction corresponds to a unary node, and each variable to a leaf. The fact that an abstraction λ binds a variable v is represented by adding a directed edge, from the unary node corresponding to the particular abstraction λ towards the leaf labelled by v. Therefore each unary node x of the Motzkin tree is carrying (zero, one or more) pointers to leaves taken from the subtree rooted at x; all leaves receiving a pointer from x (or, generally, from the same unary node) correspond to the same variable; and each leaf can receive at most one pointer.

For instance, the terms $(\lambda x.(x * x) * \lambda y.y)$ and $\lambda y.(\lambda x.x * \lambda x.y)$ correspond to the enriched trees T_0 and T_1 in Fig. 1, respectively. In particular, these terms are closed lambda-terms, since every variable is bound by an abstraction, i.e., every leaf receives exactly one pointer.

Figure 1: Two examples of lambda-terms: Each unary node corresponds to an abstraction λx binding all leaves below it which are labelled by x. Binary nodes correspond to applications merging their two subtrees t_1 and t_2 to the more complex structure $t_1 * t_2$.

The size of a lambda-term is the number of nodes in the corresponding enriched tree. It is defined recursively by

$$\begin{aligned} |x| &= 1, \\ |\lambda x.T| &= 1 + |T|, \\ |(S * T)| &= 1 + |S| + |T|. \end{aligned}$$

As mentioned in the introduction, we are interested in lambda-terms with bounded unary height. Other simplifications are possible, such as bounding the number of pointers for each unary nodes. Such terms are studied in [2] and are related to BCI and BCK logics as introduced in [22, 21, 23]. For their relations to lambda-calculus see for instance [20].

Definition 2.1. *Consider a lambda-term and its associated enriched tree T. The* unary height *of a vertex v of T, denoted by $l_u(v)$, is defined as number of unary nodes on the path connecting v with the root. The* unary height *of T, $l_u(T)$, is given by* $\max_{v \text{ vertex of } T} l_u(v)$.

Thus, an upper bound on the unary height means that we are dealing with lambda-terms where the number of nested abstractions is bounded.

In order to count lambda-terms of a given size we set up a formal equation which is then translated into generating functions. Since the class of enriched trees is isomorphic to the class of lambda-terms we do not distinguish between those classes in the sequel.

Let $\mathcal{L}$ denote the class of open lambda-terms and introduce the following atomic classes: the class of application nodes $\mathcal{N}$, the class of abstraction nodes $\mathcal{U}$, the class of free leaves $\mathcal{F}$, and the class of bound leaves $\mathcal{B}$. Then the class $\mathcal{L}$ can be described as follows:

$$\mathcal{L} = \mathcal{F} + \left(\mathcal{N} \times \mathcal{L}^2\right) + (\mathcal{U} \times \mathit{subs}(\mathcal{F} \to \mathcal{F} + \mathcal{B}, \mathcal{L}))$$

where the substitution operator $\mathit{subs}(\mathcal{F} \to \mathcal{F} + \mathcal{B}, \mathcal{L})$ corresponds to replacing some free leaves in $\mathcal{L}$ by bound leaves.

This specification gives rise to a functional equation for the bivariate generating function

$$L(z, f) = \sum_{t \text{ lambda-term}} z^{|t|} f^{\#\text{free leaves in } t}$$

which reads as follows:

$$L(z, f) = fz + zL(z, f)^2 + zL(z, f + 1).$$

In particular, the formal generating function for lambda-terms without free variables is:

$$\begin{aligned} L(z, 0) &= [f^0]L(z, f) \\ &= z^2 + 2z^3 + 4z^4 + 13z^5 + 42z^6 \\ &\quad +139z^7 + 506z^8 + 1915z^9 + 7558z^{10} + \cdots \end{aligned}$$

Note that these functional equations have to be considered in the framework of formal power series since the fast growth of the coefficients of the generating function implies that the radius of convergence of $L(z, 0)$ is zero (see Corollary 3.2 below).

Furthermore note, that the problem of counting closed or open lambda-terms is essentially the same. Indeed, the formal generating function for open lambda-terms can be derived from the formula $L(z, 1) = \frac{[f^0]L(z,f) - z[f^0]L(z,f)^2}{z}$. Consequently, the problems of enumerating lambda-terms with or without free variables are of the same difficulty and the solution for one of them yields the solution for the other one.

Now let us turn to terms of restricted unary height. Let $\mathcal{S}^{(k)}$ denote the class of closed lambda-terms with unary height less than or equal k; we want to set up an equation for the $\mathcal{S}^{(k)}$. Moreover, we set $r(T)$ to be the root of an enriched tree T and let $[r(T)..e]$ be the unique elementary path (i.e., pointers must not be used) connecting e and the root $r(T)$ of the tree.

Let us consider the combinatorial class $\mathcal{T}$ of rooted unary-binary trees (trees such that each internal node has one or two children) such that each leaf e of a tree t can be labelled with a label in $\{1, .., \kappa\}$ where $\kappa = \kappa(t, e)$ is the number of unary nodes in the unique path connecting e to the root of t. The size of t is the number of its nodes. Define the subclass $\mathcal{P}^{(i,k)}$ of $\mathcal{T}$ to be the class of unary-binary trees T such that for each leaf e we have $i + l_u([r(T)...e]) \leq k$ and every leaf e carries a label in $\{1, .., i + l_u([r(T)...e])\}$ (the definition of the unary height l_u given in Definition 2.1 for a tree extends readily to a path).

For every positive integer k, the class $\mathcal{P}^{(0,k)}$ is isomorphic to the class $\mathcal{S}^{(k)}$ and the class $\mathcal{P}^{(1,k)}$ is isomorphic to the class obtained from $\mathcal{S}^{(k)}$ by allowing free leaves. For general i, consider a path of i unary nodes to which we append a lambda-term T from $\mathcal{P}^{(i,k)}$; then the number of possible labellings for each leaf is equal to the number of possible labellings in T, to which we add i new labels for the i nodes of the path. Hence the classes $\mathcal{P}^{(i,k)}$ can be recursively specified by

$$\mathcal{P}^{(k,k)} = k\mathcal{Z} + \mathcal{Z}{\mathcal{P}^{(k,k)}}^2$$

and, for $i < k$, by

$$\mathcal{P}^{(i,k)} = i\mathcal{Z} + \mathcal{Z}\mathcal{P}^{(i+1,k)} + \mathcal{Z}{\mathcal{P}^{(i,k)}}^2. \tag{2.1}$$

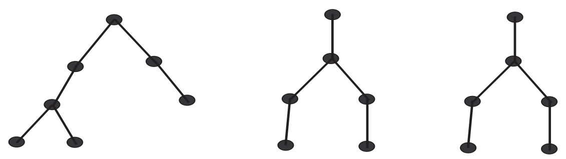

Figure 2: The trees P_0 and P_1 are elements of $\mathcal{P}^{(0,2)} \cong \mathcal{S}^{(2)}$ and correspond to T_0 and T_1 in Fig. 1. The tree P_2 is an element of $\mathcal{P}^{(i,2)}$ for $i \geq 2$: Both leaves have the same unary height as the tree itself, namely 2, and the label of the right (resp. left) leaf exceeds its unary height by 2 (resp. 1). Thus $P_2 \notin \mathcal{P}^{(i,2)}$ for $i < 2$.

Using the traditional correspondance between specifications and generating functions we obtain for $i < k$

$$P^{(i,k)}(z) = \frac{1 - \sqrt{1 - 4\,iz^2 - 4\,z^2 P^{(i+1,k)}(z)}}{2z}$$

and

$$P^{(k,k)}(z) = \frac{1 - \sqrt{1 - 4\,z^2 k}}{2z}.$$

Thus, this recursive specification gives directly the generating function $S^{(k)}(z)$ associated to $\mathcal{S}^{(k)}$. We get

$$S^{(k)}(z) = (1/2z)(1 - \sqrt{\ldots.})$$

where the part under the radical is

$$1 - 2z + 2z\sqrt{\cdots\sqrt{1 - 4\,(k-1)\,z^2 - 2z + 2z\sqrt{1 - 4kz^2}}}.$$

Note that for $n \leq k$ we have $[z^n]S^{(k)}(z) = [z^n]L(z, 1)$ and thus $S^{(k)}(z)$ converges to $L(z, 1)$ in the sense of formal convergence of power series (cf. [16, p. 731]).

In the next two sections we consider the singularities of this generating function and determine its dominant one together with its type. Then we use this information to obtain the asymptotic behaviour of its coefficients.

3 Toward an asymptotic analysis: jump over the radicands

Nested structures appear frequently in combinatorial objects; many are the structures that lead to generating functions in the form of continued fractions (see for example [13, 8]). Nested radicals are less frequent; they can appear when enumerating binary non plane trees [27, 16, 7], where there appears a "continued square-root" expansion. When bounding the number of nestings in such trees (which amounts to writing down a generating function whose coefficients are exact up to some size n_0, but differ from n_0 upwards), the innermost radicand is the one that determines the dominant singularity, hence the asymptotic behaviour. We know of no previous example where the determination of the significant radicand fluctuates according to the number of nestings allowed.

We now consider how to determine the dominant singularity of the function $S^{(k)}(z)$: it is built of nested radicands, thus its dominant singularity must be at a point where one of the radicands vanishes. Theorem 3.1 below gives the "dominant" radicand in $S^{(k)}(z)$, i.e., the radicand having a zero which is the dominant singularity of $S^{(k)}(z)$.

Definition 3.1. *We say that a function $f(z)$ has a singularity of type $\left(1 - \frac{z}{\rho}\right)^\alpha$ at $z = \rho$ if*

$$f(z) \sim c\left(1 - \frac{z}{\rho}\right)^\alpha,$$

as $z \to \rho$ inside the domain of analyticity and for some constant c.

The following sequence will turn out to be crucial for our counting problem:

DEFINITION 3.2. *Define a sequence $(u_k)_{k\geq 0}$ for integer k as*

$$\begin{aligned} u_0 &= 0; \\ u_k &= u_{k-1}^2 + k \qquad \text{for} \qquad k>0; \\ u_{k-1} &= \sqrt{u_k - k} \qquad \text{for} \qquad k<0. \end{aligned}$$

REMARK 3.1. *In [1] it has been shown that for any doubly exponential sequence $\mathbf{v} = (v_k)_{k\geq 0}$ the limit $\chi_{\mathbf{v}} := \lim_{k\to\infty} v_k^{1/2^k}$ exists and the sequence can be represented by $v_k = \lfloor \chi_{\mathbf{v}}^{2^k} \rfloor$. Since $(u_k)_{k\geq 0}$ is doubly exponential, we can apply this result and $\lim_{k\to\infty} u_k^{1/2^k}$ can be numerically approximated by $\chi \simeq 1.36660956...$*

THEOREM 3.1. *Let $N_k = u_k^2 - u_k + k$, with u_k as in Definition 3.2. Define i such that $k \in [N_i, N_{i+1})$. If $k \neq N_i$, then the dominant radicand of $S^{(k)}(z)$ is the i-th radicand, and the dominant singularity is algebraic of type $\left(1-\frac{z}{\rho}\right)^{1/2}$. Otherwise the i-th and the $(i+1)$-st radicand simultaneously vanish at the dominant singularity of $S^{(k)}(z)$, which is algebraic of type $\left(1-\frac{z}{\rho}\right)^{1/4}$. Here the first radicand is the innermost one, then we count outwards.*

The rest of this section is devoted to the proof of Theorem 3.1: we show that the i^{th} radicand, when restricted to the real part of its definition domain, is decreasing, and use this to prove that it has a single real positive root, which turns out to be the dominant singularity.

Let us begin with the case $k=1$: The generating function of $\mathcal{S}^{(1)}$ is

$$S^{(1)}(z) = \frac{1-\sqrt{1-2z+2z\sqrt{1-4z^2}}}{2z}.$$

Thus, $S^{(1)}(z)$ has a dominant singularity at $z=\frac{1}{2}$ of type $(1-2z)^{\frac{1}{4}}$, and a singularity at $z=-\frac{1}{2}$ of type $(1+2z)^{\frac{1}{2}}$ which turns out to give a negligible contribution to the asymptotics. Moreover, $S^{(1)}(z)$ is clearly analytic in a disk of radius $1/2+\varepsilon$ with two notches in $z=\pm 1/2$. Hence we can directly apply a classical transfer theorem [15] to obtain the asymptotic number of closed lambda-terms with unary height at most 1, as

$$[z^n]S^{(1)}(z) \sim \frac{1}{4}\frac{2^{\frac{1}{4}}2^n n^{-\frac{5}{4}}}{\Gamma(\frac{3}{4})}, \text{ as } n\to\infty.$$

Now let k grow: the dominant singularity no longer comes from the innermost radical. For instance, when $k=2$, the singularity from the second innermost radical becomes dominant; when $k=9$, it is the singularity from the third innermost radical which becomes dominant. This phenomenon requires some explaination.

Let us denote by $R_{i,k}(z)$ the i^{th} radicand ($1 \leq i \leq k+1$) of $S^{(k)}(z)$, according to the numbering from the innermost outwards as adopted in the assertion of Theorem 3.1, i.e., we have

$$P^{(i,k)}(z) = \frac{1-\sqrt{R_{k-i,k}(z)}}{2z}.$$

We can write the radicands recursively as follows :

$$R_{1,k}(z) := 1-4kz^2$$

and, for $i\geq 2$:

$$R_{i,k}(z) = 1-4\,(k-i+1)z^2 - 2z + 2z\sqrt{R_{i-1,k}(z)},$$

which gives

$$\begin{aligned} R_{i,k}(z) &= 1-4(k-i+1)z^2 - 2z \\ &+2z\sqrt{\cdots\sqrt{1-4\,(k-1)\,z^2-2z+2z\left(\sqrt{1-4kz^2}\right)}} \end{aligned}$$

What are the roots of such a radicand $R_{i,k}$? Recall that we obtained $R_{i,k}$ when solving the equation for the generating function of $\mathcal{P}^{(i,k)}$ obtained from (2.1):

$$R_{k-i,k}(z) = 1-4z^2\left(i+P^{(i+1,k)}(z)\right).$$

Since $P^{(i+1,k)}(z)$ is the generating function of certain trees (of which there exist some of any size), there exists a sequence $a_{n,i,k}$ of strictly positive numbers such that $R_{i,k}(z) = 1-\sum_{n\geq 2} a_{n,i,k}z^n$. Assume that $R_{i,k}(z)$ has a unique real positive root x_0 (which we shall prove later on); can there be other (imaginary) roots $z = x_0e^{I\theta}$ of same modulus? If so, then we would have

$$1 = \sum_{n\geq 2} a_{n,i,k}x_0^n = |\sum_{n\geq 2} a_{n,i,k}x_0^n e^{In\theta}|$$

which can only hold if $e^{In\theta}=1$ whenever $a_{n,i,k}\neq 0$. As the $a_{n,i,k}$ are basically the coefficients of some $P^{(j,k)}$, i.e., the numbers of lambda-terms in some suitable class, we can easily check that this is not so. Hence proving that $R^{(i,k)}(z)$ has a single real positive root will ensure that this root is the zero of smallest modulus of $R^{(i,k)}(z)$.

We thus turn to the behaviour of the generating functions on the positive real axis and determine the interval where the radicands are positive.

Lemma 3.1. *For every k and $1 \leq i \leq k$, the real function $R_{i,k}(x)$ is decreasing on the positive part of its real domain of definition.*

Proof. By induction on i: $R_{1,k}(x)$ is clearly decreasing. Now,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}R_{i,k}(x) &= -8\,(k-i+1)\,x - 2 \\ &\quad +2\sqrt{R_{i-1,k}(x)} + \frac{x\frac{\mathrm{d}}{\mathrm{d}x}R_{i-1,k}(x)}{\sqrt{R_{i-1,k}(x)}}.\end{aligned}$$

But $R_{i-1,k}(0) = 1$, hence $-2 + 2\sqrt{R_{i-1,k}(x)} \leq -2 + 2\sqrt{R_{i-1,k}(0)} = 0$; as by induction $\frac{\mathrm{d}}{\mathrm{d}x}R_{i-1,k}(x) \leq 0$, we obtain that $\frac{\mathrm{d}}{\mathrm{d}x}R_{i,k}(x) \leq 0$.

Corollary 3.1. *For every k and $1 \leq i \leq k$, the real function $R_{i,k}(x)$ has at most one real positive root.*

In order to proceed, consider the restriction of $P^{(i,k)}(z)$ to the real line and denote the domain where $P^{(i,k)}(z)$ is defined as a real function by $D(i,k)$. Now remember that we have a nested construction: the domains $D(i,k)$ are themselves nested.

Lemma 3.2. *For every $k \geq 1$ we have $\forall i < i'$: $D(i,k) \subseteq D(i',k)$. Moreover, if $D(i,k) = D(i+1,k)$ then we have $\forall j < i : D(j,k) = D(j+1,k)$.*

Proof. The first assertion is obvious, since when one of the inner radicands becomes negative the whole function is not defined any more.

In order to show the second assertion, consider the finite sequence $D(k,k), D(k-1,k), \ldots, D(0,k)$. We know that this sequence is a decreasing chain of intervals. Assume $D(i,k) \subseteq D(i+1,k)$ but $D(i,k) \neq D(i+1,k)$: the upper ends of the intervals are different. Let z_i denote the unique positive singularity of $\sqrt{R_{i,k}}$, i.e. $D_{i,k} = [0, z_i[$. Then our assumption implies that the outer radicand $R_{k-i+1,k}$ becomes null for a value of x strictly smaller than the value where the inner radicand $R_{k-i,k}$ is singular, i.e., $z_{k-i+1} < z_{k-i}$. In this case, the dominant singularity cannot be the singularity of the inner radicand $R_{k-i,k}$. Conversely, if $D(i+1,k) = D(i,k)$, then the least positive singularity of $P^{(i,k)}(z)$ cannot coincide with a zero of its outermost radicand and must therefore be the singularity of $P^{(i+1,k)}(z)$.

Consequently, the sequence of zeroes $(z_i)_{i=1,\ldots,k+1}$ of the radicands $R_{i,k}$ is decreasing and eventually becomes stationary. The same holds for the sequence of intervals. Thus the dominant singularity of $S^{(k)}$ is the smallest value of this (finite) sequence: the position $\hat{\imath}(k)$ of the dominant radicand $R_{\hat{\imath}(k),k}$ (i.e. the radicand such that $R_{\hat{\imath}(k),k} = 0$ at the singularity of $S^{(k)}$ is the dominant singularity of $S^{(k)}$) is the greatest i such that $D(i,k) = D(i+1,k)$.

For instance, for $k = 8$, the sequence of upper bounds for $D(8,8), ...D(0,8)$ is $[x_8 \simeq 0.1768, x_7 \simeq 0.168, x_6 = x_7, ...]$. So the second radicand $\hat{\imath}(8) = 2$ is dominant. For $k = 9$, the sequence of upper bounds for $D(9,9), ...D(0,9)$ is $[x_9 \simeq 0.1667, x_8 \simeq 0.15716, x_7 \simeq 0.15714, x_6 = x_7, ...]$. Here the third radicand $\hat{\imath}(9) = 3$ is dominant.

As is usual in classical analytic combinatorics, we need to know the dominant singularity ρ of the global generating function $S^{(k)}$, in order to evaluate its asymptotic behaviour. But this is not enough: we need to know *which* radicand leads to the singularity ρ. Recall that we call this radicand "dominant". Determining it corresponds precisely to finding the least integer value i such that $R_{i,k}(x) = 0$ and $R_{i+1,k}(x) \geq 0$. This is equivalent to solving the system

$$\begin{aligned} 1 - 4(k-i')x^2 - 2x &= 0 \\ R_{i'+1,k}(x) &= 0 \end{aligned}$$

in the variables $(x, i') \in R^+$, then taking $i = \lceil i' \rceil$. Indeed, it suffices to remark that the function $i \mapsto 1 - 4(k-i)z(i)^2 - 2z(i)$, with $z(i)$ the unique root of $R_{i,k}(x) = 0$, is an increasing function. This function is defined for integer i; however we can extend it to a function $\Phi(x)$ defined over the real positive numbers, with Φ a decreasing function with continuous derivative. We obtain $\Phi'(x) = 4z(x)^2 - (2(4k-4x))z(x)(\frac{\mathrm{d}z(x)}{\mathrm{d}x}) - 2\frac{\mathrm{d}z(x)}{\mathrm{d}x})$. As $\frac{\mathrm{d}z(x)}{\mathrm{d}x} < 0$, we have that $\Phi'(x) > 0$.

Let us give the first values of the localization of the dominant radicand in the following table, where the first column gives the generating function, the second and third ones the rank of the dominant radicand and value of the dominant singularity.

Function	Radicand	Singularity
$S^{(1)}$	{1,2}	0.5
$S^{(2)}$	2	0.3438
$S^{(3)}$	2	0.2760
...	...	...
$S^{(8)}$	{2,3}	0.1667
$S^{(9)}$	3	0.1571
...	...	...
$S^{(134)}$	3	0.0418
$S^{(135)}$	{3,4}	0.0417
$S^{(136)}$	4	0.0415
...	...	...

This table shows an unexpected phenomenon. For some critical values: $1, 8, 135, \ldots$, we have a "jump" from a radicand to its successor; this jump occurs precisely when two successive radicands cancel for the same value.

Our next goal is to describe explicitly this sequence of critical values.

Proposition 3.1. *We have*

$$R_{s,u_k^2-u_k+k}\left(\frac{1}{2u_k}\right)=\left(\frac{u_{k-s-1}}{u_k}\right)^2.$$

Proof. By induction on k.

Lemma 3.3. *Let ρ_k denote the dominant singularity of $S^{(N_k)}(z)$ where $N_k = u_k^2 - u_k + k$ with u_k as in Definition 3.2. Then we have $\rho_k = 1/2u_k$.*

Proof. For $s = k$ and $s = k+1$, $R_{s,N_k}(\frac{1}{2u_k}) = 0$.

The sequence (u_k) is doubly exponential: the localization of the dominant radicand is given by 1, 8, 135, 21760, 479982377, 230404115058374088, 53086056457022411574281640206019007, ...; the sequence of the dominant singularities is 1/2, 1/6, 1/24, 1/296, 1/43818, 1/960008574, 1/460808231076756752,...

Corollary 3.2. *The radius of convergence of the generating function $L(z,0)$ enumerating all lambda-terms is zero.*

Proof. The number of lambda-terms of size k being greater than the number of lambda-terms of the same size and unary height p for any p, the radius of convergence of the global generating function $L(0,z)$ is smaller than (or equal to) the radius of convergence ρ_k of the function $S^{(N_k)}$, for any k. But the sequence of these radii converges to 0.

4 Asymptotic analysis, and transition between different behaviours.

We are now in the position to give the asymptotic behaviour of the number of lambda-terms with bounded unary height.

Theorem 4.1. *Let $(N_k)_{k\geq 0}$ be as in Theorem 3.1 and $(u_k)_{k\geq 0}$ as in Definition 3.2. The following asymptotic relations hold:*

$$[z^n]\mathcal{S}^{(N_k)} \sim \frac{1}{\Gamma(3/4)} h_k n^{-5/4}(2u_k)^n, \ \textit{as } n\to\infty,$$

where ρ_k is the root of the dominant radicand of $S^{(N_k)}$ and

$$h_k = \left(-\frac{u_k}{2}\frac{d}{dz}R_{k,N_k}(\rho_k)\right)^{1/4}\prod_{i=k}^{N_k-1}\frac{1}{2u_{-i}}. \tag{4.2}$$

If m is in $]N_k, N_{k+1}[$, then there exists a suitable constant h_m such that

$$[z^n]S^{(m)} \sim \frac{1}{\Gamma(1/2)} h_m n^{-3/2}(\rho_k)^{-n}, \ \textit{as } n\to\infty.$$

Proof. 1. We first consider the case when the unary height is one of the N_k: then the dominant singularity is algebraic of order 1/4.

The generating function $S^{(N_k)}(z)$ has a dominant singularity in $\rho_k = \frac{1}{2u_k}$. We prove in the sequel that a singular expansion around 0 gives $S^{(N_k)}(z) \sim \tau_k - h_k(1-\frac{z}{\rho_k})^{\frac14}$. The expressions of τ_k and h_k need some attention.

Let us begin by defining the bivariate function

$$\check{S}^{(k)}(z,Y) = (1/2z)\left(1-\sqrt{1-2z+2z\sqrt{\cdots}}\right)$$

with the second radical being

$$\sqrt{1-4z^2-2z+2z\sqrt{\cdots\sqrt{1-4kz^2-2z+2zY}}}.$$

We have $\check{S}^{(k-i)}(z,\sqrt{R_{i,k}(z)}) = S^{(k)}(z)$. Now a first-order singular expansion of $S^{(k)}(z)$ around ρ (denoted by $\mathrm{DL}(S^{(k)}(z),\rho,1)$) is

$$\check{S}^{(k)}(\rho,0) + \left(\frac{\partial}{\partial Y}\check{S}^{(k)}\right)(\rho,0)\cdot \mathrm{DL}(\sqrt{R_{i,k}(z)},\rho,1).$$

A short calculation gives $\tau_k = \check{S}^{(N_k)}(\rho_k,0) = u_k - u_{k-N_k-1}$. Now, by inductive derivation we obtain

$$\frac{\partial}{\partial Y}\check{S}^{(N_k)}(\rho_k,0) = -u_k\prod_{i=k}^{N_k-1}\frac{1}{2u_{-i}}.$$

The final point is to determine the asymptotic expansion $\mathrm{DL}(\sqrt{R_{i,k}(z)},\rho,1)$. This can be done without difficulty and we obtain

$$\mathrm{DL}(\sqrt{R_{k,N_k}(z)},\rho_k,1) \sim \frac{1}{u_k^{3/4}}\left(-\frac12\frac{\mathrm{d}}{\mathrm{d}z}R_{k,N_k}(\rho_k)\right)^{1/4}\left(1-\frac{z}{\rho_k}\right)^{1/4}.$$

The theorem follows from the next lemma, which gives the value of the derivative of R_{k,N_k} at ρ_k.

Lemma 4.1. *$\frac{d}{dz}R_{k,N_k}(\rho_k) = w_{k-1,k}$ where $w_{k-1,k}$ is defined recursively by $w_{0,k} = -\frac{4N_k}{u_k}$, $w_{i,k} = -\frac{4(N_k-i)}{u_k} - 2 + 2\frac{u_{k-i}}{u_k} + \frac{w_{i-1,k}}{2u_{k-i}}$.*

2. If the unary height is not one of the N_k, the dominant singularity is again algebraic, but with order 1/2.

Numerical computations for the coefficients of asymptotic expansions when $k = 1, 8, 135$ give:

$$\begin{aligned}[z^n]S^{(1)}(z) &\sim \frac{2^{1/4}}{4\,\Gamma(\frac34)}\cdot\left(\frac1n\right)^{5/4}\cdot 2^n\\ &\sim 0.2426128012\cdot\left(\frac1n\right)^{5/4}\cdot 2^n;\end{aligned}$$

$$[z^n]S^{(8)}(z)| \sim \frac{6^{1/4}}{1152\,\Gamma(\frac{3}{4})}\cdot\frac{\alpha}{\beta}\cdot\left(\frac{1}{n}\right)^{5/4}\cdot 6^n$$
$$\sim 9.318885373\cdot 10^{-5}\left(\frac{1}{n}\right)^{5/4} 6^n$$

where we have exact expressions for α and β:

$$\alpha = \sqrt{6+\sqrt{5+\sqrt{4+\sqrt{3+\sqrt{3}}}}};$$
$$\beta = \sqrt{5+\sqrt{4+\sqrt{3+\sqrt{3}}}}\cdot\sqrt{4+\sqrt{3+\sqrt{3}}}\cdot\sqrt{3+\sqrt{3}}\cdot\left(2/3+1/9\sqrt{5+\sqrt{4+\sqrt{3+\sqrt{3}}}}\right)$$

and

$$[z^n]S^{(135)}(z) \sim 7.116999389\cdot 10^{-158}\left(\frac{1}{n}\right)^{5/4} 24^n.$$

The constant factor in the asymptotic expression turns out to decrease very quickly, which leads us to the following observation: when the maximal unary height k grows, the asymptotic regime is not the one we may observe on a "reasonable" (up to some tens of thousands) number of values. We have plotted in Figure 3 the ratio between the number of lambda-terms with unary height exactly k and size n, and the number of lambda-terms of size n (without restriction on the height). The figure suggests that, for any given size n, the unary height is close to a Gaussian distribution. In particular, this gives some experimental justification to the change of behaviour: the wave indicates the "good" estimate for the number of abstractions in a lambda-term; for instance, if we consider lambda-terms of size 198, then the vast majority of these terms has a unary height between 25 and 50.

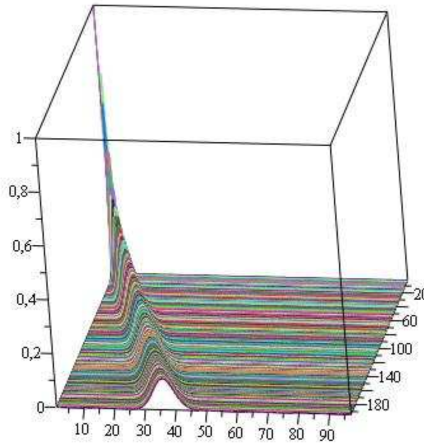

Figure 3: Distribution of lambda-terms of size $n \in [1,...,198]$ and unary height $k \in [1,...,98]$

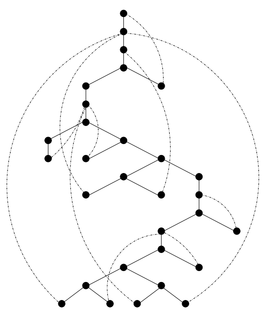

Figure 4: A random lambda-term of size 30, with the edges from unary nodes to leaves

5 Random generation and observations

5.1 Random generation of lambda-terms To get a feeling of the "average" behaviour of a combinatorial object, a method of choice is the random generation of terms of large size. We considered two methods to try to generate a random lambda-term of bounded unary height: the recursive method [17] and Boltzmann sampling. Boltzmann samplers are powerful tools to generate objects in specified combinatorial classes at random. They were introduced in [12] and extended furthermore by numerous authors [3, 4, 5, 6, 14, 28, 29]. Note that theoretically, a Boltzmann sampler can generate a tree of size close to n on average in linear time. We considered Boltzmann sampling of a closed term, with different success depending on the unary height: the efficiency decreases very quickly as the maximal unary height grows. When $k = 8$, we can generate terms of size 10000 in a few seconds on a standard personal computer. Figure 5.1 presents a term of size 6853 with unary height bounded by 8.[1] However,

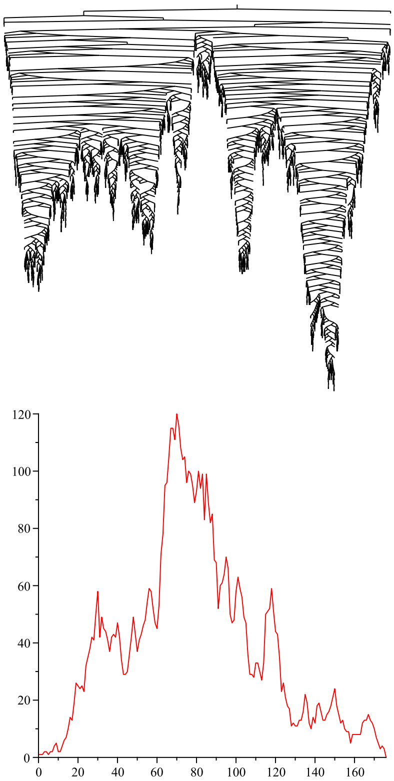

Figure 5: A random lambda-term of unary height ≤ 8 and its profile

if we consider trees with a maximal unary height of 135, a Boltzmann sampler is not able to produce objects of size larger than 200 in a "reasonable" time (less than one day). The explanation of the phenomenon is as follows: an "average" random lambda-term begins with a large number of unary nodes; cf. Figure 5.1 (see also [10] for a result on the same vein on a related model); drawing the sufficient number of unary nodes has very low probability in the Boltzmann process. Figure 6 gives the various probabilities of drawing a leaf, a unary node, or a binary node, plotted against the recursive depth (number of recursive calls to the generator). After a (long!) starting phase where the probability of stopping is more than 90%, the Boltzmann sampler becomes efficient. In other words, Boltzmann sampling is linear,

[1]For large sizes and for the sake of readability, we have not indicated the edges between a unary node and the leaf labels.

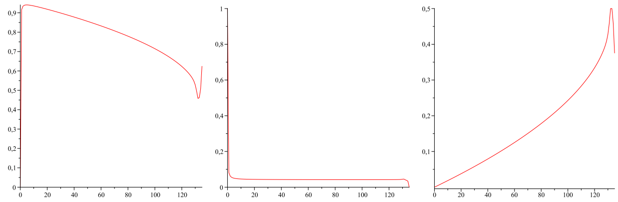

Figure 6: Left, the probability that the singular Boltzmann sampler $\Gamma\mathcal{P}^{(k,135)}$ of objects in $\mathcal{P}^{(k,135)}$ stops immediately,. Middle, the probability that the sampler $\Gamma\mathcal{P}^{(k,135)}$ calls $\Gamma\mathcal{P}^{(k-1,135)}$. Right, the probability that the sampler $\Gamma\mathcal{P}^{(k,135)}$ independently calls 2 generators $\Gamma\mathcal{P}^{(k,135)}$.

but with a constant depending on the maximum unary height which grows *very* quickly: the recursive form of the specification of lambda-terms and its varying behaviour are not conductive to random generation with a Boltzmann sampler.

We have thus turned to the recursive method; using the Maple package Combstruct, we have been able to generate quickly enough lambda-terms of size 200 and unary height bounded by 200–which means that there is de facto no restriction on the unary height of the lambda-term. Figure 7 shows what can be considered as a generic lambda-term for this size.

5.2 Profile and average behaviour of a lambda-term Being able to draw repeatedly random lambda-terms allows us to make tentative conjectures on their various parameters: profile, depth, etc. Figure 7 shows, together with a "generic" lambda-term, its profile (number of nodes at each level) and (far right) the profile averaged on 500 random lambda-terms, together with the average profile of a planar binary tree. From this we can make several empirical observations.
- The distribution of the profiles is poorly concentrated (this is also the case for planar binary trees).
- The levels containing the larger number of nodes are much farther from the root than in binary trees.
- A simulation of the distribution for the total – binary and unary – height (not presented here) also shows a clear difference with planar binary trees: the average depth seems to grow linearly, not in $\sqrt{n}$ as for binary trees. In the same vein, the width of a lambda-terms appears to grow as $\log n$.
- A random lambda-term usually begins with a large number of successive unary nodes interspersed with a few binary nodes; most binary nodes appear further down.

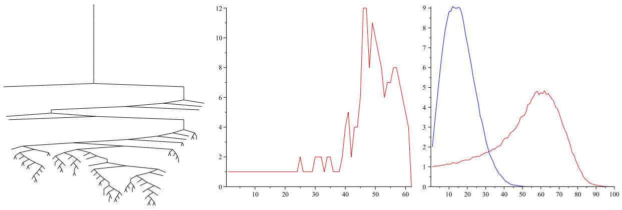

Figure 7: Left, a lambda-term of size 200. Middle, its profile. Right, the average profile (red) computed over 500 random lambda-terms, compared with the average profile for planar binary tree (blue, Airy function)

6 Conclusion and perspectives

We have shown in this paper that even a restricted class of lambda-terms exhibits an outstanding combinatorial complexity. Among others, we have discovered the unexpected behaviour of the position for the dominant radicand, which jumps according to some function behaving as $ln(ln(k))$, with k the maximum unary height of a lambda-term. Theorem 4.1 characterizes precisely these jumps and the asymptotic number of lambda-terms with bounded height.

A byproduct of our work concerns Boltzmann samplers: by trying to use them for the random generation of lambda-terms, we have pushed them to their limit. We feel that it might be possible to improve Boltzmann random generation, when we wish to apply it to combinatorial structures whose Boltzmann distribution is concentrated around the smallest sizes.

Finally, in terms of average properties and growth, lambda-terms widely differ from the usual models for trees such as simple families or increasing trees, for which we know the behaviour of classical parameters: number of trees of given size, profile, etc. Indeed they seem to behave, in some sense, like "ornamented" paths, i.e. long strings on which are grafted relatively small subterms.

Of course, such results need to be explained and quantified more rigorously. Let us also mention that the enumeration of (unrestricted) lambda-terms is still an open problem, which we shall probably need to solve if we are to study such parameters as the average unary height, or profile.

Acknowledgments. The authors thank M. Zaionc who introduced them to the problem of enumerating lambda-terms, P. Flajolet who gave the first insights on this problem, and P. Lescanne who provided input on lambda-calculus.

References

[1] Alfred V. Aho and Neil J. A. Sloane. Some Doubly Exponential Sequences. Fibonacci Quarterly, Vol. 11 (1970), pp. 429-437.

[2] Olivier Bodini, Danièle Gardy, and Alice Jacquot. Asymptotics and random sampling for formulae in intuitionist logical systems. 7^{th} GASCOM workshop, Montréal (Canada), September 2-4, 2010.

[3] Olivier Bodini, Danièle Gardy, and Olivier Roussel, Boys-and-girls birthdays and Hadamard products, 7^{th} international Conference on Lattice Paths and Applications, Siena (Italy), July 4-7, 2010.

[4] Olivier Bodini and Alice Jacquot, Boltzmann Samplers For Colored Combinatorial Objects, 6^{th} GASCOM workshop, Bibbiena (Italy), June 16-19, 2008.

[5] Olivier Bodini and Yann Ponty, Multi-dimensional Boltzmann sampling of languages, Conference on Analysis of Algorithms, AofA'10, Vienna (Austria), June 28-July 2, 2010.

[6] Manuel Bodirsky, Eric Fusy, Mihyun Kang, and Stefan Vigerske, An unbiased pointing operator for unlabeled structures, with applications to counting and sampling, SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms (Philadelphia, PA, USA), Society for Industrial and Applied Mathematics, 2007, pp. 356–365.

[7] Miklós Bóna and Philippe Flajolet, Isomorphism and Symmetries in Random Phylogenetic Trees, Journal of Applied Probability, vol 46 (2009), pp. 1005–1019.

[8] Jérémie Bouttier and Emmanuel Guitter, Planar maps and continued fractions, preprint, arXiv:1007:0419, 2010.

[9] Alonzo Church. An Unsolvable Problem of Elementary Number Theory. *Amer. J. Math.*, 58(2):345–363, 1936.

[10] René David, Katarzyna Grygiel, Jakub Kozik, Christophe Raffalli, Guillaume Theyssier, and Marek Zaionc. Asymptotically almost all λ-terms are strongly normalizing. Preprint, arXiv:math.LO/0903.5505, 2010.

[11] René David and Marek Zaionc. Counting proofs in propositional logic. *Arch. Math. Logic*, 48(2):185–199, 2009.

[12] Philippe Duchon, Philippe Flajolet, Guy Louchard, and Gilles Schaeffer, Boltzmann samplers for the random generation of combinatorial structures, Combinatorics, Probablity, and Computing **13** (2004), no. 4–5, 577–625, Special issue on Analysis of Algorithms.

[13] Philippe Flajolet, Combinatorial aspects of continued fractions, Discrete Mathematics 32 (1980), 125–161. Reprinted in the 35th Special Anniversary Issue of Discrete Mathematics, Volume 306, Issue 1011, Pages 992–1021 (2006).

[14] Philippe Flajolet, Eric Fusy, and Carine Pivoteau, Boltzmann sampling of unlabelled structures, Proceedings of ANALCO'07 (Analytic Combinatorics and Algorithms) Conference (SIAM Press, ed.), vol. 126, 2007, pp. 201–211.

[15] Philippe Flajolet and Andrew Odlyzko. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, 3(2):216-240, 1990.

[16] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.

[17] Philippe Flajolet, Paul Zimmerman, and Bernard Van Cutsem, A calculus for the random generation of labelled combinatorial structures, Theoretical Computer Science **132** (1994), no. 1-2, 1–35.

[18] Herve Fournier, Daniéle Gardy, Antoine Genitrini, and Bernhard Gittenberger. Complexity and limiting ratio of Boolean functions over implication. In *Proceedings of the 33rd International Symposium on Mathematical Foundations of Computer Science*, volume 5162/2008 of Lecture Notes in Comput. Sci., pages 347-262. Springer, 2008.

[19] Antoine Genitrini, Jakub Kozik, and Marek Zaionc. Intuitionistic vs. classical tautologies, quantitative comparison. In *Types for proofs and programs*, volume 4941 of *Lecture Notes in Comput. Sci.*, pages 100–109. Springer, Berlin, 2008.

[20] J. Roger Hindley. BCK and BCI logics, condensed detachment and the 2-property. *Notre Dame J. Formal Logic*, 34(2):231–250, 1993.

[21] Yasuyuki Imai and Kiyoshi Iséki. Corrections to: "On axiom systems of propositional calculi. I". *Proc. Japan Acad.*, 41:669, 1965.

[22] Yasuyuki Imai and Kiyoshi Iséki. On axiom systems of propositional calculi. I. *Proc. Japan Acad.*, 41:436–439, 1965.

[23] Kiyoshi Iséki and Shôtarô Tanaka. An introduction to the theory of BCK-algebras. *Math. Japon.*, 23(1):1–26, 1978/79.

[24] Stephen C. Kleene. A Theory of Positive Integers in Formal Logic. Part I. *Amer. J. Math.*, 57(1):153–173, 1935.

[25] Stephen C. Kleene. A Theory of Positive Integers in Formal Logic. Part II. *Amer. J. Math.*, 57(2):219–244, 1935.

[26] Albert Nijenhuis and Herbert S. Wilf, Combinatorial algorithms, Academic Press, 1978.

[27] Richard Otter, The number of trees, Ann. of Math. 49 (3), 583–599, 1948.

[28] Carine Pivoteau, Bruno Salvy, and Michèle Soria, Boltzmann oracle for combinatorial systems, Algorithms, Trees, Combinatorics and Probabilities, Discrete Mathematics and Theoretical Computer Science, 2008, Proceedings of the Fifth Colloquium on Mathematics and Computer Science. Blaubeuren, Germany. September 22-26, 2008, pp. 475–488.

[29] Olivier Roussel and Michèle Soria, Boltzmann sampling of ordered structures, Electronic Notes in Discrete Mathematics **35** (2009), 305–310.

[30] Marek Zaionc. On the asymptotic density of tautologies in logic of implication and negation. Reports on Mathematical Logic Vol. 39, 67-87, 2005.