# On the Number of Lambda Terms With Prescribed Size of Their De Bruijn Representation*

**Bernhard Gittenberger[1] and Zbigniew Gołębiewski[2]**

1 **Institute for Discrete Mathematics and Geometry, Technische Universität Wien, Wiedner Hauptstraße 8-10/104, 1040 Wien, Austria**
`gittenberger@dmg.tuwien.ac.at`

2 **Institute for Discrete Mathematics and Geometry, Technische Universität Wien, Wiedner Hauptstraße 8-10/104, 1040 Wien, Austria**
`zbigniew.golebiewski@pwr.edu.pl`

## Abstract

John Tromp introduced the so-called 'binary lambda calculus' as a way to encode lambda terms in terms of $0-1$-strings. Later, Grygiel and Lescanne conjectured that the number of binary lambda terms with $m$ free indices and of size $n$ (encoded as binary words of length $n$) is $o\left(n^{-3/2}\tau^{-n}\right)$ for $\tau \approx 1.963448\ldots$. We generalize the proposed notion of size and show that for several classes of lambda terms, including binary lambda terms with $m$ free indices, the number of terms of size $n$ is $\Theta\left(n^{-3/2}\rho^{-n}\right)$ with some class dependent constant $\rho$, which in particular disproves the above mentioned conjecture. A way to obtain lower and upper bounds for the constant near the leading term is presented and numerical results for a few previously introduced classes of lambda terms are given.

**1998 ACM Subject Classification** G.2.1 Combinatorics, D.1.6 Logic Programming

**Keywords and phrases** lambda calculus, terms enumeration, analytic combinatorics

**Digital Object Identifier** 10.4230/LIPIcs.STACS.2016.40

## 1 Introduction

The objects of our interest are lambda terms which are a basic object of lambda calculus. A lambda term is a formal expression which is described by the grammar $M ::= x \,|\, \lambda x.M \,|\, (M\, N)$ where $x$ is a variable, the operation $(M\, N)$ is called application, and using the quantifier $\lambda$ is called abstraction. In a term of the form $\lambda x.M$ each occurrence of $x$ in $M$ is called a bound variable. We say that a variable $x$ is free in a term $M$ if it is not in the scope of any abstraction. A term with no free variables is called closed, otherwise open. Two terms are considered equivalent if they are identical up to renaming of the variables, *i.e.*, more formally speaking, they can be transformed into each other by $\alpha$-conversion.

In this paper we are interested in counting lambda terms whose size corresponds to their De Bruijn representation (*i.e.* nameless expressions in the sense of [3]).

▶ **Definition 1.** A De Bruijn representation is a word described by the following specification:

$$M ::= n \,|\, \lambda M \,|\, M\, M$$

* This work was partially supported by FWF grant SFB F50-03.

where $n$ is a positive integer, called a De Bruijn index. Each occurrence of a De Bruijn index is called a variable and each $\lambda$ an abstraction. A variable $n$ of a De Bruijn representation $w$ is bound if the prefix of $w$ which has this variable as its last symbol contains at least $n$ times the symbol $\lambda$, otherwise it is free. The abstraction which binds a variable $n$ is the $n$th $\lambda$ before the variable when parsing the De Bruijn representation from that variable $n$ backwards to the first symbol.

For the purpose of the analysis we will use the notation consistent with the one used in [1]. This means that the variable $n$ will be represented as a sequence of $n$ symbols, namely as a string of $n-1$ so-called 'successors' $S$ and a so-called 'zero' 0 at the end. Obviously, there is a one to one correspondence between equivalence classes of lambda terms (as described in the first paragraph) and De Bruijn representations. For instance, the De Bruijn representation of the lambda-term $\lambda x.\lambda y.xy$ (which is e.g. equivalent to $\lambda a.\lambda b.ab$ or $\lambda y.\lambda x.yx$) is $\lambda\lambda 21$; using the notation with successors this becomes $\lambda\lambda((S0)0)$.

In this paper we are interested in counting lambda terms of given size where we use a general notion of size which covers several previously studied models from the literature. We count the building blocks of lambda terms, zeros, successors, abstractions and applications, with size $a, b, c$ and $d$, respectively. Formally, if $M$ and $N$ are lambda terms, then

$$|0| = a, \qquad |Sn| = |n| + b, \qquad |\lambda M| = |M| + c, \qquad |MN| = |M| + |N| + d.$$

Thus we have for the example given above $|\lambda\lambda((S0)0)| = 2a + b + 2c + d$. Assigning sizes for the symbols like above covers several previously introduced notions of size:

- so called 'natural counting' (introduced in [1]) where $a = b = c = d = 1$,
- so called 'less natural counting' (introduced in [1]) where $a = 0, b = c = 1, d = 2$.
- binary lambda calculus (introduced in [8]) where $b = 1, a = c = d = 2$,

▶ **Assumption 1.** *Throughout the paper we will make the following assumptions about the constants $a, b, c, d$:*

1. *$a, b, c, d$ are nonnegative integers,*
2. *$a + d \geq 1$,*
3. *$b, c \geq 1$,*
4. *$\gcd(b, c, a + d) = 1$.*

If the zeros and the applications both had size 0 (*i.e.* $a + d = 0$), then we would have infinitely many terms of the given size, because one could insert arbitrarily many applications and zeros into a term without increasing its size. If the successors or the abstractions had size 0 (*i.e.* $b$ or $c$ equals to 0), then we would again have infinitely many terms of given size, because one could insert arbitrarily long strings of successors or abstractions into a term without increasing its size. The last assumption is more technical in its nature. It ensures that the generating function associated with the sequence of the number of lambda-terms will have exactly one singularity on the circle of convergence, which is on the positive real line. The case of several singularities is not only technically more complicated, but it is for instance not even *a priori* clear which singularities are important and which are negligible. So we cannot expect that it differs from the single singularity case only by a multiplicative constant.

We mention that in [6] lambda terms with size function corresponding to $a = b = 0$ and $c = d = 1$ were considered, but another restriction was imposed on the terms.

**Notations.** We introduce some notations which will be frequently used throughout the paper: If $p$ is a polynomial, then $\mathrm{RootOf}\,\{p\}$ will denote the smallest positive root of $p$. Moreover, we will write $[z^n]f(z)$ for the $n$th coefficient of the power series expansion of $f(z)$ at $z = 0$ and $f(z) \prec g(z)$ (or $f(z) \preceq g(z)$) to denote that $[z^n]\, f(z) < [z^n]\, g(z)$ (or $[z^n]\, f(z) \le [z^n]\, g(z)$) for all integers $n$.

**Plan of the Paper.** The primary aim of this paper is the asymptotic enumeration of closed lambda terms of given size with the size tending to infinity. In the next section we define several classes of lambda terms as well as the generating function associated with them, present our main results and prove several auxiliary results which will be important in the sequel. We derive the asymptotic equivalent of the number of closed terms of given size up to a constant factor. This is established by construction of upper and lower bounds for the coefficients of the generating functions. These constructions are done in Sections 3 and 4. To get fairly accurate numerical bounds we present a method for improving the previously obtained bounds in Section 5. Finally, Section 6 is devoted to the derivation of very accurate results for classes of lambda terms which have been previously studied in the literature.

## 2 Main Results

In order to count lambda terms of a given size we set up a formal equation which is then translated into a functional equation for generating functions. For this we will utilise the symbolic method developed in [5].

Let us introduce the following atomic classes: the class of zeros $\mathcal{Z}$, the class of successors $\mathcal{S}$, the class of abstractions $\mathcal{U}$ and the class of applications $\mathcal{A}$. Then the class $\mathcal{L}_\infty$ of lambda terms can be described as follows:

$$\mathcal{L}_\infty = \textsc{Seq}(\mathcal{S}) \times \mathcal{Z} + \mathcal{U} \times \mathcal{L}_\infty + \mathcal{A} \times \mathcal{L}_\infty^2 \tag{1}$$

The number of lambda terms of size $n$, denoted by $L_{\infty,n}$, is $|\{t \in \mathcal{L}_\infty : |t| = n\}|$. Let $L_\infty(z) = \sum_{n\ge 0} L_{\infty,n} z^n$ be the generating function associated with $\mathcal{L}_\infty$. Then specification (1) gives rise to a functional equation for the generating function $L_\infty(z)$:

$$L_\infty(z) = z^a \sum_{j=0}^{\infty} z^{bj} + z^c L_\infty(z) + z^d L_\infty(z)^2. \tag{2}$$

Solving (2) we get

$$L_\infty(z) = \frac{1 - z^c - \sqrt{(1-z^c)^2 - \frac{4z^{a+d}}{1-z^b}}}{2z^d}, \tag{3}$$

which defines an analytic function in a neighbourhood of $z = 0$.

▶ **Proposition 2.** *Let* $\rho = \mathrm{RootOf}\left\{(1-z^b)(1-z^c)^2 - 4z^{a+d}\right\}$. *Then*

$$L_\infty(z) = a_\infty + b_\infty \left(1 - \frac{z}{\rho}\right)^{\frac{1}{2}} + O\left(\left|1 - \frac{z}{\rho}\right|\right), \tag{4}$$

*for some constants* $a_\infty > 0, b_\infty < 0$ *that depend on* $a, b, c, d$.

**Proof.** Let $f(z) = (1-z^b)(1-z^c)^2 - 4z^{a+d}$. Then $\rho$ is the smallest positive solution of $f(z) = 0$. If we compute derivative $f'(z) = -4(a+b)z^{a+b-1} - 2cz^{c-1}(1-z^b)(1-z^c) - bz^{b-1}(1-z^c)^2$

we can observe that all three terms are negative for $0 < z < 1$. Since $0 < \rho < 1$, the function $f(z)$ does not have a double root at $\rho$ and thus $L_\infty(z)$ has an algebraic singularity of type $\frac{1}{2}$ which means that its Newton-Puiseux expansion is of the form (4).

Since $L_\infty(z)$ is a power series with positive coefficients, we know that $a_\infty = L_\infty(\rho) > 0$ and $b_\infty < 0$. ◀

▶ **Corollary 3.** *The coefficients of $L_\infty(z)$ satisfy $[z^n]L_\infty(z) \sim C\rho^{-n}n^{-3/2}$, as $n \to \infty$, where $C = -b_\infty/(2\sqrt{\pi})$.*

Let us define the class of $m$-open lambda terms, denoted $\mathcal{L}_m$, as

$$\mathcal{L}_m = \{t \in \mathcal{L}_\infty : \text{a prefix of } m \text{ abstractions } \lambda \text{ makes } t \text{ a closed term}\}.$$

We remark that any $m$-open lambda term is obviously $m+1$-open as well. The number of $m$-open lambda terms of size $n$ is denoted by $L_{m,n}$ and the generating function associated with the class by $L_m(z) = \sum_{n\geq 0} L_{m,n}z^n$. Similarly to $\mathcal{L}_\infty$, the class $\mathcal{L}_m$ can be specified, and this specification yields the functional equation

$$L_m(z) = z^a \sum_{j=0}^{m-1} z^{bj} + z^c L_{m+1}(z) + z^d L_m(z)^2 \tag{5}$$

Note that $L_0(z)$ is the generating function of the set $\mathcal{L}_0$ of closed lambda terms.

Let $\mathcal{K}_m = \mathcal{L}_\infty \setminus \mathcal{L}_m$ and $K_m(z) = L_\infty(z) - L_m(z)$. Then using (2) and (5) we obtain

$$K_m(z) = z^a \sum_{j=m}^{\infty} z^{bj} + z^c K_{m+1}(z) + z^d K_m(z)L_\infty(z) + z^d K_m(z)L_m(z). \tag{6}$$

which implies

$$K_m(z) = \frac{z^{a+bm}}{(1-z^b)(1-z^d(L_\infty(z)+L_m(z)))} + \frac{z^c}{1-z^d(L_\infty(z)+L_m(z))}K_{m+1}(z). \tag{7}$$

Note that $K_m(z)$ as well as $L_m(z)$ define analytic functions in a neighbourhood of $z = 0$.

Let us state the main theorem of the paper:

▶ **Theorem 4.** *Let $\rho = \operatorname{RootOf}\left\{(1-z^b)(1-z^c)^2 - 4z^{a+d}\right\}$. Then there exist positive constants $\underline{C}$ and $\overline{C}$ (depending on $a, b, c, d$ and $m$) such that the number of $m$-open lambda terms of size $n$ satisfies*

$$\liminf_{n\to\infty} \frac{[z^n]\, L_m(z)}{\underline{C} n^{-\frac{3}{2}}\rho^{-n}} \geq 1 \quad \text{and} \quad \limsup_{n\to\infty} \frac{[z^n]\, L_m(z)}{\overline{C} n^{-\frac{3}{2}}\rho^{-n}} \leq 1, \tag{8}$$

▶ Remark. In case of given $a, b, c, d$ and $m$ we can compute numerically such constants $\underline{C}$ and $\overline{C}$. This will be done for some of the models mentioned in the introduction.

Before proving this theorem we will present the key ideas needed for our proof. We introduce the class $\mathcal{L}_m^{(h)}$ of lambda terms in $\mathcal{L}_m$ where the length of each string of successors is bounded by a constant integer $h$. As before, set $L_{m,n}^{(h)} = \left|\left\{t \in \mathcal{L}_m^{(h)} : |t| = n\right\}\right|$ and $L_m^{(h)}(z) = \sum_{n\geq 0} L_{m,n}^{(h)} z^n$. Then $L_m^{(h)}(z)$ satisfies the functional equation

$$L_m^{(h)}(z) = \begin{cases} z^a \sum_{j=0}^{m-1} z^{bj} + z^c L_{m+1}^{(h)}(z) + z^d L_m^{(h)}(z)^2 & \text{if } m < h, \\ z^a \sum_{j=0}^{h-1} z^{bj} + z^c L_h^{(h)}(z) + z^d L_h^{(h)}(z)^2 & \text{if } m \geq h. \end{cases} \tag{9}$$

Notice that for $m \geq h$ we have a quadratic equation for $L_m^{(h)}(z) = L_h^{(h)}(z)$ that has the solution

$$L_h^{(h)}(z) = \frac{1 - z^c - \sqrt{(1-z^c)^2 - 4z^{a+d}\frac{1-z^{bh}}{1-z^b}}}{2z^d}.$$

For $m < h$ we have a relation between $L_m^{(h)}(z)$ and $L_{m+1}^{(h)}(z)$ which gives rise to a representation of $L_m^{(h)}(z)$ in terms of nested radicands (*cf.* [2]) after all. Indeed, for $m < h$ we have

$$L_m^{(h)}(z) = \frac{1 - \sqrt{r_m(z) + 2z^c\sqrt{r_{m+1}(z) + 2z^c\sqrt{\cdots\sqrt{r_{h-1}(z) + 2z^c\sqrt{r_h(z)}}}}}}{2z^d} \tag{10}$$

where

$$r_j(z) = \begin{cases} 1 - 4z^{a+d}\frac{1-z^{jb}}{1-z^b} - 2z^c & \text{if } m \leq j < h-1, \\ 1 - 4z^{a+d}\frac{1-z^{(h-1)b}}{1-z^b} - 2z^c + 2z^{2c} & \text{if } j = h-1, \\ (1-z^c)^2 - 4z^{a+d}\frac{1-z^{bh}}{1-z^b} & \text{if } j = h. \end{cases}$$

▶ **Lemma 5.** *For all $m \geq 0$ the dominant singularity $\rho^{(h)} := \rho_m^{(h)}$ of $L_m^{(h)}(z)$ is independent of $m$. Moreover, we have $\lim_{h\to\infty} \rho^{(h)} = \rho$.*

**Proof.** One can check that the dominant singularity of $L_m^{(h)}(z)$ comes from smallest positive root of $r_h(z)$ and that it is of type $\frac{1}{2}$ (*i.e.* $L_m^{(h)}(z) = a_m^{(h)} + b_m^{(h)}\left(1 - \frac{z}{\rho^{(h)}}\right)^{\frac{1}{2}} + O\left(\left|1 - \frac{z}{\rho^{(h)}}\right|\right)$ as $z \to \rho_m^{(h)}$ for some constants $a_m^{(h)}, b_m^{(h)}$ depending on $m$ and $h$). Consequently, $\rho_m^{(h)}$ is independent of $m$. Notice that in the unit interval $r_h(z)$ converges uniformly to $r(z)$, the radicand in (3). Thus $\lim_{h\to\infty} \rho^{(h)} = \rho$ since $\rho$ is the smallest positive root of $r(z)$. ◀

Let us begin with computing the radii of convergence of the functions $K_m(z)$ and $L_m(z)$. For the case of binary lambda calculus Lemmas 7 and 8 were already proven in [7]. To extend those results to our more general setting, we will use different techniques.

▶ **Lemma 6.** *For all $m \geq 0$ the radius of convergence of $K_m(z)$ equals $\rho$ (the radius of convergence of $L_\infty(z)$).*

**Proof.** Inspecting (7) reveals that the key part is $\frac{1}{1-z^d(L_\infty(z)+L_m(z))}$. This is the generating function of a sequence of combinatorial structures associated with the generating function $z^d(L_\infty(z) + L_m(z))$. One can check that we are not in the supercritical sequence schema case (*i.e.* a singularity of considered fraction does not come from the root of its denominator, see [5, pp. 293]) because $1 - \rho^d\left(L_\infty(\rho) + L_m(\rho)\right) > 0$. This follows from

$$\rho^d\left(L_\infty(\rho) + L_m(\rho)\right) \leq 2\rho^d L_\infty(\rho) = 1 - \rho^c < 1.$$

The first inequality holds because $L_\infty(\rho) \geq L_m(\rho)$ for all $m \geq 0$ and the second one because $\rho > 0$. Moreover, the radius of convergence of $L_m(z)$ is larger than or equal to the radius of convergence of $L_\infty(z)$ because $\mathcal{L}_m \subseteq \mathcal{L}_\infty$. Therefore, for all $m \geq 0$ the radius of convergence of $K_m(z)$ equals $\rho$, the radius of convergence of $L_\infty(z)$. ◀

▶ **Lemma 7.** *All the functions $L_m(z)$, $m \geq 0$, have the same radius of convergence.*

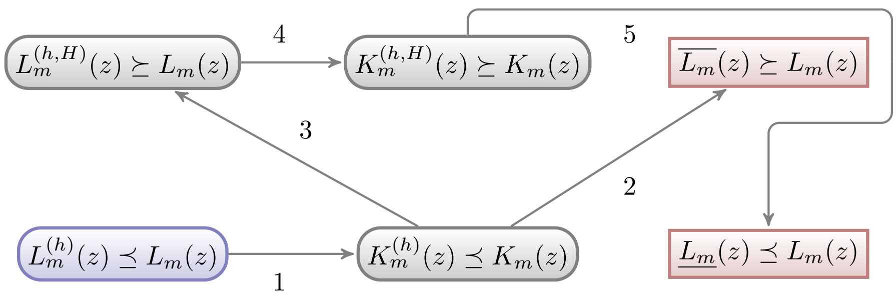


**Figure 1** The diagram illustrates the idea how we obtain the upper and the lower bound (in terms of the coefficients) for the function $L_m(z)$. Starting point is denoted by a blue node, the finish nodes are red.

**Proof.** Let $\rho_m$ denote the radius of convergence of the function $L_m(z)$. From the definition of the function $L_m(z)$ it is known that for all $m \geq 0$ and for all $n$ we have $[z^n]\, L_m(z) \leq [z^n]\, L_{m+1}(z)$ and therefore $\rho_m \geq \rho_{m+1}$. Moreover, from (5) we know

$$L_{m+1}(z) = -z^{a-c} \sum_{j=0}^{m-1} z^{bj} + z^{-c} L_m(z) - z^{d-c} L_m(z)^2.$$

Notice that $\rho_m \leq 1$ because $\rho_m \leq \rho^{(h)} < 1$ for $h \geq m$. Then due to the fact that $z^{-c} L_m(z) - z^{d-c} L_m(z)^2$ has radius of convergence bigger or equal $\rho_m$, we have $\rho_m \leq \rho_{m+1}$. ◀

▶ **Lemma 8.** *For all $m \geq 0$ the radius of convergence of $L_m(z)$ equals $\rho$.*

**Proof.** Take $L_m^{(m)}$, defined in (9). Recall that $\rho^{(m)}, \rho_m$ and $\rho$ denote the radii of convergence of $L_m^{(m)}(z), L_m(z)$ and $L_\infty(z)$, respectively. Notice that for all $m, n \geq 0$ we have $[z^n]\, L_m^{(m)}(z) \leq [z^n]\, L_m(z) \leq [z^n]\, L_\infty(z)$ and thus $\rho^{(m)} \geq \rho_m \geq \rho$. Now, the assertion follows from Lemmas 5 and 7. ◀

In the next sections we will present how to obtain an upper and a lower bound for $[z^n]\, L_m(z)$. The idea is to construct auxiliary functions satisfying certain inequalities and to use them to construct further ones until we have the desired bound. The procedure follows the flowchart depicted in Fig. 1.

## 3 Upper Bound for $[z^n]\, L_m(z)$

Notice that for all integers $h$ and $m$ we have $\mathcal{L}_m^{(h)} \subset \mathcal{L}_m$. Moreover, for all $m, h \geq 0$ there exists $n_m^{(h)}$ such that $[z^n]\, L_m^{(h)}(z) = [z^n]\, L_m(z)$ if $n < n_m^{(h)}$ and $[z^n]\, L_m^{(h)}(z) \leq [z^n]\, L_m(z)$ else. We will use those properties of $L_m^{(h)}(z)$ in order to derive a lower bound for the asymptotics of $[z^n]\, K_m(z)$.

Note that (6) corresponds to an equation of the form $\mathcal{K}_m = \mathcal{F}(\mathcal{K}_m, \mathcal{K}_{m+1}, \mathcal{L}_\infty, \mathcal{L}_m)$ where $\mathcal{F}$ is obtained from (6) by replacing addition and multiplication by their combinatorial counterparts and generating functions by their corresponding sets. Now, define the new set $\mathcal{K}_m^{(h)} := \mathcal{F}\left(\mathcal{K}_m^{(h)}, \mathcal{K}_{m+1}^{(h)}, \mathcal{L}_\infty, \mathcal{L}_m^{(h)}\right)$. From the construction of $\mathcal{F}$ and the properties of $\mathcal{L}_m^{(h)}$ we know that $\mathcal{K}_m^{(h)} \subseteq \mathcal{K}_m$. Let $K_{m,n}^{(h)} = \left|\left\{t \in \mathcal{K}_m^{(h)} : |t| = n\right\}\right|$ and $K_m^{(h)}(z) = \sum_{n \geq 0} K_{m,n}^{(h)} z^n$.

Then $K_m^{(h)}(z)$ satisfies the functional equation

$$K_m^{(h)}(z) = z^a \sum_{j=m}^{\infty} z^{bj} + z^c K_{m+1}^{(h)}(z) + z^d K_m^{(h)}(z) L_\infty(z) + z^d K_m^{(h)}(z) L_m^{(h)}(z). \tag{11}$$

In fact, what we did is that we replaced in the application operation every $m$-open lambda term (corresponding to the subterm $z^d K_m(z) L_m(z)$ of (6)) by an $m$-open lambda term where each string of successors has bounded length (corresponding to $z^d K_m(z) L_m^{(h)}(z)$). Solving (11) we get after some computations

$$K_m^{(h)}(z) = \frac{z^{a-cm}}{1-z^b} \sum_{j=m}^{\infty} z^{j(b+c)} \prod_{i=m}^{j} \frac{1}{1 - z^d \left(L_\infty(z) + L_i^{(h)}(z)\right)} =: S_{m,\infty}(z). \tag{12}$$

▶ **Lemma 9.** *Let $\rho, a_\infty, b_\infty$ be as in Proposition 2 and $\tilde{c}_i = 1/(1-\rho^d(a_\infty + L_i^{(h)}(\rho)))$ and $\tilde{d}_i = b_\infty \rho^d/(1-\rho^d(a_\infty + L_i^{(h)}(\rho)))^2$. Then $K_m^{(h)}(z)$ admits the expansion*

$$K_m^{(h)}(z) = c_m^{(h)} + d_m^{(h)} \left(1 - \frac{z}{\rho}\right)^{\frac{1}{2}} + O\left(\left|1 - \frac{z}{\rho}\right|\right), \quad \textit{as } z \to \rho, \tag{13}$$

*where*

$$c_m^{(h)} = \begin{cases} S_{m,h-1}(\rho) + R_{c_m^{(h)}} & \textit{if } m < h, \\ \frac{\rho^{a+bm}}{(1-\rho^b)\left(1-\rho^{b+c}-\rho^d\left(a_\infty + L_h^{(h)}(\rho)\right)\right)} & \textit{else}, \end{cases}$$

$$d_m^{(h)} = \begin{cases} \frac{\rho^{a-cm}}{1-\rho^b} \sum_{j=m}^{h-1} \rho^{j(b+c)} \sum_{i=m}^{j} \frac{\tilde{d}_i}{\tilde{c}_i} \prod_{k=m}^{j} \tilde{c}_k + R_{d_m^{(h)}} & \textit{if } m < h, \\ \frac{b_\infty \rho^{a+bm+d}}{(1-\rho^b)\left(1-\rho^{b+c}-\rho^d\left(a_\infty + L_h^{(h)}(\rho)\right)\right)^2} & \textit{else}, \end{cases}$$

*with*

$$R_{c_m^{(h)}} = \frac{\rho^{a+bh+c(h-m)}}{(1-\rho^b)\left(1-\rho^{b+c}-\rho^d\left(a_\infty + L_h^{(h)}(\rho)\right)\right)} \prod_{i=m}^{h-1} \tilde{c}_i,$$

$$R_{d_m^{(h)}} = \frac{b_\infty \rho^{a+bh+c(h-m)+d}}{(1-\rho^b)\left(1-\rho^{b+c}-\rho^d\left(a_\infty + L_h^{(h)}(\rho)\right)\right)} \left(\prod_{i=m}^{h-1} \tilde{c}_i\right) \times \left(\sum_{i=m}^{h-1} \tilde{c}_i + \frac{1}{1-\rho^{b+c}-\rho^d\left(a_\infty + L_h^{(h)}(\rho)\right)}\right).$$

**Proof.** Let us recall that for all $m \geq h$ we have $L_m^{(h)}(z) = L_h^{(h)}(z)$. Therefore we can split the infinite sum $S_{m,\infty}(z)$ in (12) into the finite one $S_{m,h-1}(z)$ and the rest $S_{h,\infty}(z)$.

**Case I:** $m < h$. First, consider the finite sum $S_{m,h-1}(z)$. As in the proof of Lemma 6 we identify the key term, show that we are not in the supercritical case, and expand by means of Proposition 2. Eventually, this yields

$$\prod_{i=m}^{j} \frac{1}{1 - z^d(L_\infty(z) + L_i^{(h)}(z))} = \tilde{c}_{m,j} + \tilde{d}_{m,j} \left(1 - \frac{z}{\rho}\right)^{\frac{1}{2}} + O\left(\left|1 - \frac{z}{\rho}\right|\right)$$

where $\tilde{c}_{m,j} = \prod_{i=m}^{j} \tilde{c}_i$ and $\tilde{d}_{m,j} = \sum_{i=m}^{j} \frac{\tilde{d}_i}{\tilde{c}_i} \prod_{k=m}^{j} \tilde{c}_k$. Since $\left(1-\frac{z}{\rho}\right)^{\frac{1}{2}}$ does neither depend on $m$ nor on $j$ and $\frac{z^{a-cm}}{1-z^b}$ has poles only on the unit circle, for all $m \geq 0$ we have

$$S_{m,h-1}(z) = \tilde{\tilde{c}}_m + \tilde{\tilde{d}}_m \left(1-\frac{z}{\rho}\right)^{\frac{1}{2}} + O\left(\left|1-\frac{z}{\rho}\right|\right)$$

where $\tilde{\tilde{c}}_m = \frac{\rho^{a-cm}}{1-\rho^b} \sum_{j=m}^{h-1} \rho^{j(b+c)} \tilde{c}_{m,i}$ and $\tilde{\tilde{d}}_m = \frac{\rho^{a-cm}}{1-\rho^b} \sum_{j=m}^{h-1} \rho^{j(b+c)} \tilde{d}_{m,i}$.

Now, let us look on the infinite part of the sum in (12). We will refer to its contributions to the first and second coefficient of the Newton-Puiseux expansion (13) as the remainders $R_{c_m^{(h)}}$ and $R_{d_m^{(h)}}$, respectively. Since for all $m \geq h$ we have $L_m^{(h)}(z) = L_h^{(h)}(z)$, the sum can be rewritten as

$$S_{h,\infty}(z) = \left(\prod_{i=m}^{h-1} \frac{1}{1-z^d\left(L_\infty(z)+L_i^{(h)}(z)\right)}\right) \cdot \frac{z^{a+bh+c(h-m)}}{(1-z^b)\left(1-z^{b+c}-z^d\left(L_\infty(z)+L_h^{(h)}(z)\right)\right)}.$$

We already know how to handle the product part of this expression, so let us consider the fraction $\frac{z^{a+bh+c(h-m)}}{(1-z^b)\left(1-z^{b+c}-z^d\left(L_\infty(z)+L_h^{(h)}(z)\right)\right)}$. Similarly to before, we have to check that the singularity of this function does not come from the root of the denominator but from $L_\infty(z)$ (it cannot come from $L_h^{(h)}$ because it has a bigger radius of convergence than $L_\infty(z)$). So, we have to show the inequality $1-\rho^{b+c}-\rho^d\left(L_\infty(\rho)+L_h^{(h)}(\rho)\right) > 0$. But from $L_\infty(\rho) \geq L_h^{(h)}(\rho)$ and $0 < \rho^{b+c} < \rho^c < 1$ we obtain $\rho^d(L_\infty(\rho)+L_h^{(h)}(\rho)) \leq 2\rho^d L_\infty(\rho) = 1-\rho^c < 1-\rho^{b+c}$ and hence the desired inequality indeed holds. Now, similarly to the previous case we use the Newton-Puiseux expansion of $L_\infty(z)$ at $\rho$ to derive an expansion of the infinite part of the sum in (12) and get the asserted result.

**Case II:** $m \geq h$. This case is easier, because the finite part of the sum in (12) does not exist and the other part can be evaluated to a closed form which can be treated as above. ◀

Using the transfer lemmas of [4] (applied to $K_m^{(h)}(z)$) and $[z^n]\, K_m^{(h)}(z) \leq [z^n]\, K_m(z)$, we get $\liminf_{n\to\infty} ([z^n]\, K_m(z)) \cdot \Gamma(-1/2) n^{3/2} \rho^n / d_m^{(h)} \geq 1$.

▶ **Corollary 10.** *The number of $m$-open lambda terms of size $n$ satisfies*

$$\limsup_{n\to\infty} \frac{[z^n]\, L_m(z)}{\overline{C} n^{-\frac{3}{2}} \rho^{-n}} \leq 1 \textit{ where } \overline{C} = \frac{b_\infty - d_m^{(h)}}{\Gamma\left(-\frac{1}{2}\right)}.$$

## 4 Lower Bound for $[z^n]\, L_m(z)$

The idea behind obtaining a lower bound for $[z^n]\, L_m(z)$ is similar to the one used for the upper bound. First we will find an upper bound for $[z^n]\, K_m(z)$ using the function

$$L_m^{(h,H)}(z) = \sum_{n\geq 0} L_{m,n}^{(h,H)} z^n = \begin{cases} L_\infty(z) - K_m^{(h)}(z) & \text{if } m < H, \\ L_\infty(z) & \text{else.} \end{cases} \tag{14}$$

Notice that for all $m, h, H, n \geq 0$ we have $[z^n]\, L_m(z) \leq [z^n]\, L_m^{(h,H)}(z)$. Let $\mathcal{L}_m^{(h,H)}$ denote the class of combinatorial structures associated with $L_m^{(h,H)}(z)$ and define the new set $\mathcal{K}_m^{(h,H)} := \mathcal{F}\left(\mathcal{K}_m^{(h,H)}, \mathcal{K}_{m+1}^{(h,H)}, \mathcal{L}_\infty, \mathcal{L}_m^{(h,H)}\right)$. From the construction of $\mathcal{F}$ and the above properties of $\mathcal{L}_m^{(h,H)}$ we know that $\mathcal{K}_m \subseteq \mathcal{K}_m^{(h,H)}$. Let $K_{m,n}^{(h,H)} = \left|\left\{t \in \mathcal{K}_m^{(h,H)} : |t| = n\right\}\right|$ and $K_m^{(h,H)}(z) = \sum_{n\geq 0} K_{m,n}^{(h,H)} z^n$. Then $K_m^{(h,H)}(z)$ is given by

$$K_m^{(h,H)}(z) = z^a \sum_{j=m}^{\infty} z^{bj} + z^c K_{m+1}^{(h,H)}(z) + z^d K_m^{(h,H)}(z) L_\infty(z) + z^d K_m^{(h,H)}(z) L_m^{(h,H)}(z).$$

Solving this equation and using (14) we get

$$\begin{aligned}
K_m^{(h,H)}(z) &= \frac{z^{a-cm}}{1-z^b} \sum_{j=m}^{H-1} z^{j(b+c)} \prod_{i=m}^{j} \frac{1}{1-z^d\left(L_\infty(z)+L_m^{(h,H)}(z)\right)} \\
&= \frac{z^{a-cm}}{1-z^b} \sum_{j=m}^{H-1} z^{j(b+c)} \prod_{i=m}^{j} \frac{1}{1-z^d\left(2L_\infty(z)-K_i^{(h)}(z)\right)} + \\
&\quad \frac{z^{a+bH+c(H-m)}}{(1-z^d)\left(1-z^{b+c}-2z^d L_\infty(z)\right)} \left(\prod_{i=m}^{H-1} \frac{1}{1-z^d\left(2L_\infty(z)-K_i^{(h)}(z)\right)}\right). \qquad (15)
\end{aligned}$$

▶ **Lemma 11.** *Let $\rho$ be the radius of convergence of the function $L_\infty(z)$. Then the generating function $K_m^{(h,H)}(z)$ admits the following expansion*

$$K_m^{(h,H)}(z) = c_m^{(h,H)} + d_m^{(h,H)} \left(1-\frac{z}{\rho}\right)^{\frac{1}{2}} + O\left(\left|1-\frac{z}{\rho}\right|\right), \qquad (16)$$

*where*

$$c_m^{(h,H)} = \frac{\rho^{a-cm}}{1-\rho^b} \sum_{j=m}^{H-1} \rho^{j(b+c)} \prod_{i=m}^{j} \frac{1}{1-\rho^d\left(2a_\infty - c_i^{(h)}\right)} + R_{c_m^{(h,H)}},$$

$$d_m^{(h,H)} = \frac{\rho^{a-cm}}{1-\rho^b} \sum_{j=m}^{H-1} \rho^{j(b+c)} \sum_{i=m}^{j} \frac{\rho^d\left(2b_\infty - d_i^{(h)}\right)}{1-\rho^d\left(2a_\infty - c_i^{(h)}\right)} \prod_{i=m}^{j} \frac{1}{1-\rho^d\left(2a_\infty - c_i^{(h)}\right)} + R_{d_m^{(h,H)}},$$

*$a_\infty, b_\infty$ and $c_i^{(h)}, d_i^{(h)}$ come from the expansion of $L_\infty(z)$ and $K_i^{(h)}(z)$, respectively, at $\rho$ (see Proposition 2 and the proof of Lemma 9) and*

$$R_{c_m^{(h,H)}} = \frac{\rho^{a+bH+c(H-m)}}{(1-\rho^d)\left(1-\rho^{b+c}-2\rho^d a_\infty\right)} \left(\prod_{i=m}^{H-1} \frac{1}{1-\rho^d\left(2a_\infty - c_i^{(h)}\right)}\right),$$

$$\begin{aligned}
R_{d_m^{(h,H)}} &= \frac{\rho^{a+bH+c(H-m)+d}}{(1-\rho^d)\left(1-\rho^{b+c}-2\rho^d a_\infty\right)} \left(\prod_{i=m}^{H-1} \frac{1}{1-\rho^d\left(2a_\infty - c_i^{(h)}\right)}\right) \\
&\quad \times \left(\frac{2b_\infty}{1-\rho^{b+c}-2\rho^d a_\infty} + \sum_{i=m}^{H-1} \frac{\left(2b_\infty - d_i^{(h)}\right)}{1-\rho^d\left(2a_\infty - c_i^{(h)}\right)}\right).
\end{aligned}$$

As in the previous section, we apply the transfer lemmas of [4] to $K_m^{(h,H)}(z)$ and use $[z^n]\, K_m^{(h,H)}(z) \geq [z^n]\, K_m(z)$ to arrive at the following result:

▶ **Corollary 12.** *The number of $m$-open lambda terms of size $n$ satisfies*

$$\liminf_{n\to\infty} \frac{[z^n]\, L_m(z)}{\underline{C} n^{-\frac{3}{2}} \rho^{-n}} \geq 1 \text{ where } \underline{C} = \frac{b_\infty - d_m^{(h,H)}}{\Gamma\left(-\frac{1}{2}\right)}.$$

# 5 Improvement of the Bounds

The bounding functions $\underline{L_m}(z) = L_\infty(z) - K_m^{(h,H)}(z)$ and $\overline{L_m}(z) = L_\infty(z) - K_m^{(h)}(z)$ derived in the previous sections can be used in a straightforward way to compute numerical values for $\underline{C}$ and $\overline{C}$, given concrete values for $a, b, c, d$. If we choose $h$ and $H$ big enough, then this will give us proper bounds. But in practice they still leave a large gap, at least for values $h$ and $H$ that allow us to perform the computation within a few hours on a standard PC. For instance, in case of the natural counting for $h = H = 15$ we get $\underline{C}^{(nat)} \approx 0.00404525\ldots$ and $\overline{C}^{(nat)} \approx 0.18086721\ldots$.

We show a simple way to improve $\underline{C}$ and $\overline{C}$. Let us introduce the functions $L_{m,M}^{(h)}(z)$ and $L_{m,M}^{(h,H)}(z)$, defined by the following equations:

$$L_{m,M}^{(h)}(z) = \begin{cases} z^a \sum_{j=0}^{m-1} z^{bj} + z^c L_{m+1,M}^{(h)}(z) + z^d L_{m,M}^{(h)}(z)^2 & \text{if } m < M, \\ L_\infty(z) - K_M^{(h)}(z) & \text{if } m = M, \end{cases}$$

$$L_{m,M}^{(h,H)}(z) = \begin{cases} z^a \sum_{j=0}^{m-1} z^{bj} + z^c L_{m+1,M}^{(h,H)}(z) + z^d L_{m,M}^{(h,H)}(z)^2 & \text{if } m < M, \\ L_\infty(z) - K_M^{(h,H)}(z) & \text{if } m = M. \end{cases}$$

These two functions admit a representation in terms of nested radicals which is similar to (10):

$$L_{m,M}^{(h)}(z) = \frac{1}{2z^d} \cdot \left( 1 - \sqrt{1 - 4z^{a+d}\frac{1-z^{mb}}{1-z^b} - 2z^c + 2z^c \sqrt{\cdots \sqrt{1 - 4z^{a+d}\frac{1-z^{(M-1)b}}{1-z^b} - 2z^c + 2z^c \sqrt{L_\infty(z) - K_M^{(h)}(z)}}}} \right),$$

$$L_{m,M}^{(h,H)}(z) = \frac{1}{2z^d} \cdot \left( 1 - \sqrt{1 - 4z^{a+d}\frac{1-z^{mb}}{1-z^b} - 2z^c + 2z^c \sqrt{\cdots \sqrt{1 - 4z^{a+d}\frac{1-z^{(M-1)b}}{1-z^b} - 2z^c + 2z^c \sqrt{L_\infty(z) - K_M^{(h,H)}(z)}}}} \right).$$

So, in $L_{m,M}^{(h)}(z)$ and $L_{m,M}^{(h,H)}(z)$ we are using exact expressions for $L_m(z)$ up to some constant $M$ and then replace $L_M(z)$ by a function that is its upper and lower bound, respectively. Numerical results for some previously studied notions of size (see Sections 6.1 and 6.2) reveal a significant improvement in closing the gap between the constants $\overline{C}, \underline{C}$ obtained by utilising the functions $L_{m,M}^{(h)}(z), L_{m,M}^{(h,H)}(z)$.

# 6 Results for Some Previously Introduced Notions of Size

## 6.1 Natural Counting

▶ **Lemma 13.** *The following bounds hold*

$$\liminf_{n\to\infty} \frac{[z^n]\, L_0^{(nat)}(z)}{\underline{C}^{(nat)} n^{-\frac{3}{2}} \rho^{-n}} \geq 1 \quad \textit{and} \quad \limsup_{n\to\infty} \frac{[z^n]\, L_0^{(nat)}(z)}{\overline{C}^{(nat)} n^{-\frac{3}{2}} \rho^{-n}} \leq 1 \tag{17}$$

*where* $\rho = \mathrm{RootOf}\{-1 + 3x + x^2 + x^3\} \approx 0.295598\ldots$ *and* $\underline{C}^{(nat)}, \overline{C}^{(nat)}$ *are computable constants with numerical values* $\underline{C}^{(nat)} \approx 0.00404525\ldots$ *and* $\overline{C}^{(nat)} \approx 0.18086721\ldots$.

**Table 1** Numbers rounded up to 7 digits.

| $h, H$ | $c_0^{(h)}$ | $d_0^{(h)}$ | $c_0^{(h,H)}$ | $d_0^{(h,H)}$ |
|---|---|---|---|---|
| 1 | 0.855448 | −1.153959 | 1.086200 | −3.803686 |
| 2 | 0.898032 | −1.313246 | 0.979519 | −2.581823 |
| 3 | 0.917305 | −1.397536 | 0.958215 | −2.324953 |
| 4 | 0.927248 | −1.444672 | 0.950295 | −2.236290 |
| 5 | 0.932849 | −1.472308 | 0.946185 | −2.192353 |
| 6 | 0.936128 | −1.488826 | 0.943824 | −2.167379 |
| 7 | 0.938055 | −1.498647 | 0.942443 | −2.152790 |
| 8 | 0.939174 | −1.504385 | 0.941643 | −2.144335 |
| 9 | 0.939813 | −1.507673 | 0.941187 | −2.139511 |
| 10 | 0.940172 | −1.509525 | 0.940931 | −2.136799 |
| 11 | 0.940372 | −1.510556 | 0.940788 | −2.135291 |
| 12 | 0.940482 | −1.511125 | 0.940710 | −2.134460 |
| 13 | 0.940543 | −1.511438 | 0.940667 | −2.134004 |
| 14 | 0.940576 | −1.511608 | 0.940643 | −2.133755 |
| 15 | 0.940594 | −1.511701 | 0.940630 | −2.133619 |

**Proof.** For 'natural counting' we have $a = b = c = d = 1$. From Theorem 4 we know that the radius of convergence of $L_0^{(nat)}(z)$ equals $\rho = \text{RootOf}\left\{-1+3x+x^2+x^3\right\} \approx 0.295598\ldots$. We can also easily get $L_\infty(z) \sim a_\infty + b_\infty\sqrt{1-z/\rho}$ with

$$a_\infty = \frac{1-\rho}{2\rho} \approx 1.19149\ldots \text{ and } b_\infty = \frac{1}{\rho-1}\sqrt{\frac{1+\rho+\rho^2-\rho^3}{2\rho}} \approx 2.15093\ldots,$$

and from the transfer lemmas of [4] we obtain $[z^n]\,L_\infty(z) \sim b_\infty n^{-3/2}\rho^{-n}/\Gamma(-1/2) \simeq (0.606767\ldots)\cdot n^{-\frac{3}{2}}(3.38298\ldots)^n$, as $n \to \infty$.

Since we are most interested in the enumeration of closed lambda terms, we examine the multiplicative constants in the leading term of the asymptotical lower and upper bound for $[z^n]\,L_0(z)$. From the formulas in Lemmas 9 and 11 we have computed the values for $c_0^{(h)}, d_0^{(h)}$ and $c_0^{(h,H)}, d_0^{(h,H)}$ for different constants $h$ and $H$ (see Table 1).

As expected, the bigger $h$ and $H$ are, the more accurate is the bound we get (in this case for $[z^n]\,K_0(z)$). Taking the values of $d_0^{(h,H)}$ and $d_0^{(h)}$ for $h = H = 15$, Corollaries 12 and 10 yield $\underline{C}^{(nat)} \approx 0.00404525\ldots$ and $\overline{C}^{(nat)} \approx 0.18086721\ldots$. Notice that using the values of $d_0^{(h,H)}$ to compute $\underline{C}^{(nat)}$ gives non-trivial values only for $h > 7$ (for $1 \le h \le 7$ we get negative numbers because in this case we have $\left|d_0^{(h,H)}\right| > |b_\infty|$). ◀

Figure 2 illustrates the bounds we obtained and the exact values of the coefficients $[z^n]\,L_0^{(nat)}(z)$ for $10 \le n \le 150$. Applying the approach discussed in Section 5 for $h = H = M = 13$ we get the following improvement.

▶ **Lemma 14.** *The following bounds hold*

$$\liminf_{n\to\infty} \frac{[z^n]\,L_0^{(nat)}(z)}{\underline{\underline{C}}^{(nat)} n^{-\frac{3}{2}}\rho^{-n}} \ge 1 \quad \text{and} \quad \limsup_{n\to\infty} \frac{[z^n]\,L_0^{(nat)}(z)}{\overline{\overline{C}}^{(nat)} n^{-\frac{3}{2}}\rho^{-n}} \le 1 \tag{18}$$

*where* $\rho = \text{RootOf}\{-1+3x+x^2+x^3\} \approx 0.295598\ldots$ *and* $\underline{\underline{C}}^{(nat)}, \overline{\overline{C}}^{(nat)}$ *are computable constants with numerical values* $\underline{\underline{C}}^{(nat)} \approx 0.07790995266\ldots$ *and* $\overline{\overline{C}}^{(nat)} \approx 0.07790998229\ldots$.

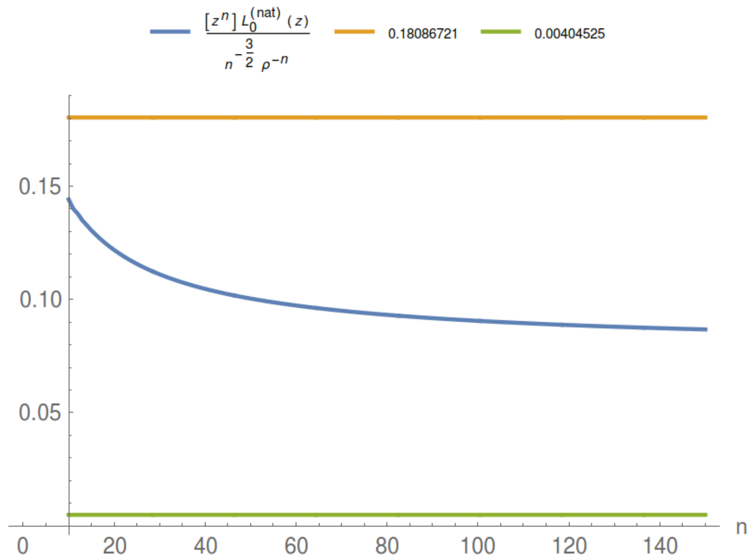


**Figure 2** Plot of the exact value of $\frac{[z^n]L_0^{(nat)}(z)}{n^{-\frac{3}{2}}\rho^{-n}}$ and computed lower and upper bound for $h = H = 15$ and $10 \leq n \leq 150$.

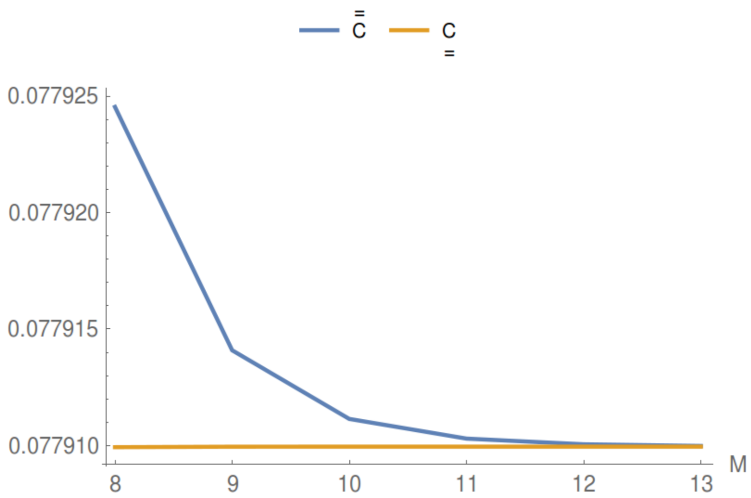


**Figure 3** Plot of the numerical values for the constants $\underline{\underline{C}}^{(nat)}, \overline{\overline{C}}^{(nat)}$ for $8 \leq h = H = M \leq 13$.

Figure 3 illustrates how the improvement discussed in Section 5 allows to reduce the gap between the constants $\underline{\underline{C}}^{(nat)}, \overline{\overline{C}}^{(nat)}$ for the lower and the upper bound.

## 6.2 Binary Lambda Calculus

▶ **Lemma 15.** *The following bounds hold*

$$\liminf_{n\to\infty} \frac{[z^n]\, L_0^{(bin)}(z)}{\underline{\underline{C}}^{(bin)} n^{-\frac{3}{2}} \rho^{-n}} \geq 1 \quad \text{and} \quad \limsup_{n\to\infty} \frac{[z^n]\, L_0^{(bin)}(z)}{\overline{\overline{C}}^{(bin)} n^{-\frac{3}{2}} \rho^{-n}} \leq 1 \tag{19}$$

*where* $\rho = \text{RootOf}\{-1 + x + 2x^2 - 2x^3 + 3x^4 + x^5\} \approx 0.509308\ldots$ *and* $\underline{\underline{C}}^{(bin)}, \overline{\overline{C}}^{(bin)}$ *are computable constants with numerical values* $\underline{\underline{C}}^{(bin)} \approx 0.01252417\ldots$ *and* $\overline{\overline{C}}^{(bin)} \approx 0.01254593\ldots$.

In order to prove this lemma it is enough to recall that in case of binary lambda calculus the size defining constants are $b = 1$ and $a = c = d = 2$. Then we used the functions $L_{0,13}^{13,13}(z), L_{0,13}^{13,13}(z)$ to obtain the numerical constants stated in Lemma 15.

## References


1 Maciej Bendkowski, Katarzyna Grygiel, Pierre Lescanne, and Marek Zaionc. A natural counting of lambda terms. *CoRR*, abs/1506.02367, 2015. URL: `http://arxiv.org/abs/1506.02367`.

2 Olivier Bodini, Danièle Gardy, and Bernhard Gittenberger. Lambda-terms of bounded unary height. In Philippe Flajolet and Daniel Panario, editors, *Proceedings of the Eighth Workshop on Analytic Algorithmics and Combinatorics, ANALCO 2011, San Francisco, California, USA, January 22, 2011*, pages 23–32. SIAM, 2011. `doi:10.1137/1.9781611973013.3`.

3 N. G. de Bruijn. Lambda calculus notation with nameless dummies, a tool for automatic formula manipulation, with application to the Church-Rosser theorem. *Nederl. Akad. Wetensch. Proc. Ser. A* **75**=*Indag. Math.*, 34:381–392, 1972.

4 Philippe Flajolet and Andrew M. Odlyzko. Singularity analysis of generating functions. *SIAM J. Discrete Math.*, 3(2):216–240, 1990. `doi:10.1137/0403019`.

5 Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, New York, NY, USA, 1 edition, 2009.

6 Katarzyna Grygiel and Pierre Lescanne. Counting and generating lambda terms. *Journal of Functional Programming*, 23:594–628, 2013. `doi:10.1017/S0956796813000178`.

7 Katarzyna Grygiel and Pierre Lescanne. Counting terms in the binary lambda calculus. In *DMTCS. 25th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms*. Discrete Mathematics & Theoretical Computer Science, Jun 2014.

8 John Tromp. Binary lambda calculus and combinatory logic. In Marcus Hutter, Wolfgang Merkle, and Paul M.B. Vitanyi, editors, *Kolmogorov Complexity and Applications*, number 06051 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2006. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany. URL: `http://drops.dagstuhl.de/opus/volltexte/2006/628`.