Counting terms
in the binary lambda calculus

Katarzyna Grygiel
Theoretical Computer Science Department
Jagiellonian University in Kraków

Pierre Lescanne
Laboratoire de l’Informatique du Parallélisme
École Normale Supérieure de Lyon

Analysis of Algorithms
Paris, June 2014
1. Open problem: enumeration of lambda calculus terms

2. Attempts to solve the enumeration problem

3. Motivation

4. Binary lambda calculus

5. Enumerating binary lambda terms
Terms in untyped lambda calculus

Lambda terms:

\[ t ::= x \mid \lambda x.t \mid (t\ t) \]
Terms in untyped lambda calculus

Lambda terms:

\[ t ::= x \mid \lambda x.t \mid (t \ t) \]

Two alternative notions of size:

\[
\begin{align*}
|x| &= 0 \\
|\lambda x.t| &= 1 + |t| \\
|t \ s| &= 1 + |t| + |s|
\end{align*}
\]

\[
\begin{align*}
|x| &= 1 \\
|\lambda x.t| &= 1 + |t| \\
|t \ s| &= 1 + |t| + |s|
\end{align*}
\]
It’s nicer to look at trees

Combinatorial interpretation by lambda trees:

The lambda tree corresponding to the lambda term
\[ \lambda x. (\lambda z. z) x (\lambda y. y (\lambda w. x)) \]
Open problem

How many closed lambda terms of a given size are there?
Attempts to solve the enumeration problem

- Studying lambda polynomials and their coefficients
- Estimating the number of simpler structures
- Applying the theory of generating functions
- Analyzing lambda terms of bounded unary height
- Enumerating BCI/BCK terms
Let $T_{n,m}$ denote the number of terms of size $n$ with occurrences of at most $m$ distinct free variables:

$$T_{0,m} = m$$

$$T_{n+1,m} = T_{n,m+1} + \sum_{i=0}^{n} T_{i,m} T_{n-i,m}$$
Let \( T_{n,m} \) denote the number of terms of size \( n \) with occurrences of at most \( m \) distinct free variables:

\[
T_{0,m} = m \\
T_{n+1,m} = T_{n,m+1} + \sum_{i=0}^{n} T_{i,m} T_{n-i,m}
\]

The sequence \( (T_{n,0})_{n \geq 0} \) (A220894 in Sloane’s OEIS) enumerates closed terms. Its first values are:

\[
0, 1, 3, 14, 82, 579, 4741, 43977, 454283, 5159441, 63782411.
\]
Studying lambda polynomials and their coefficients (|x| = 0)

For every $n \geq 0$ we associate with $T_{n,m}$ a polynomial $P_n(m)$ in $m$.

\[
P_0(m) = m
\]

\[
P_{n+1}(m) = P_n(m + 1) + \sum_{i=0}^{n} P_i(m)P_{n-i}(m)
\]

For every $n \geq 0$ we have $P_n(0) = T_{n,0}$. 
Studying lambda polynomials and their coefficients (|x| = 0)

For every $n \geq 0$ we associate with $T_{n,m}$ a polynomial $P_n(m)$ in $m$.

$$P_0(m) = m$$

$$P_{n+1}(m) = P_n(m + 1) + \sum_{i=0}^{n} P_i(m)P_{n-i}(m)$$

For every $n \geq 0$ we have $P_n(0) = T_{n,0}$.

For $k > 0$ and $n \geq 0$, let $p_n^{[k]}$ denote the $k$-th leading coefficient of the polynomial $P_n$.

$$p_n^{[k]} = \frac{1}{2^{k-1}(k-1)!\sqrt{\pi}} \frac{4^n n^{(2k-5)/2}}{8n} \left( 1 + \frac{(2k-3)(2k-5)}{8n} + O\left(\frac{1}{n^4}\right) \right).$$
Estimating the number of simpler structures \(|x| = 0\)

David, KG, Kozik, Raffalli, Theyssier, Zaionc; 2011

The number of closed lambda terms of size \(n\) is bounded:

- from below by \((\frac{(4 - \varepsilon)n}{\ln n})^{n - \frac{n}{\ln n}}\),
- from above by \((\frac{(12 + \varepsilon)n}{\ln n})^{n - \frac{n}{3\ln n}}\).
The number of closed lambda terms of size $n$ is bounded:

- from below by
  \[
  \left( \frac{(4 - \varepsilon)n}{\ln n} \right)^{n - \frac{n}{\ln n}},
  \]
- from above by
  \[
  \left( \frac{(12 + \varepsilon)n}{\ln n} \right)^{n - \frac{n}{3\ln n}}.
  \]

Therefore

\[
T_{n,0} \sim n^n(\ln n)^{-n} O(e^n).
\]
Applying generating functions \((|x| = 1)\)

Let \(L_{n,k}\) denote the number of lambda terms of size \(n\) and with at most \(k\) distinct free variables and let \(\ell(z, f) = \sum_{n,k \geq 0} z^n f^k\). Hence, \(\ell(z, 0)\) is the generating function for the sequence enumerating closed lambda terms.
Applying generating functions $(|x| = 1)$

Let $L_{n,k}$ denote the number of lambda terms of size $n$ and with at most $k$ distinct free variables and let $\ell(z, f) = \sum_{n,k \geq 0} z^n f^k$. Hence, $\ell(z, 0)$ is the generating function for the sequence enumerating closed lambda terms.

Bodini, Gardy, Gittenberger; 2011

The function $\ell(z, f)$ satisfies the functional equation

$$\ell(z, f) = fz + z\ell^2(z, f) + z\ell(z, f + 1).$$

This leads to

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

Therefore,

$$\ell(z, 0) = \frac{1 - \sqrt{1 - 2z + 2z \sqrt{1 - 4z^2 - 2z + 2z \sqrt{1 - 8z^2 - 2z + 2z \sqrt{1 - \ldots}}}}}{2z}.$$
Analyzing terms of bounded unary height \(|x| = 1\)

Bodini, Gardy, Gittenberger; 2011

Consider closed terms with a limited (given) number of unary nodes on each branch.
Consider closed terms with a limited (given) number of unary nodes on each branch.

The generating function for the sequence enumerating such terms is the truncated version of $\ell(z, 0)$. 

Bodini, Gardy, Gittenberger; 2011

One observes the intriguing behavior of dominant singularities of such generating functions. Some singularities cancel one radical, some cancel two.

This leads to two different asymptotic behaviors, dependent on the value of unary height.
Analyzing terms of bounded unary height \((|x| = 1)\)

Bodini, Gardy, Gittenberger; 2011

Consider closed terms with a limited (given) number of unary nodes on each branch.

The generating function for the sequence enumerating such terms is the truncated version of \(\ell(z, 0)\).

One observes the intriguing behavior of dominant singularities of such generating functions.
Some singularities cancel one radicand, some cancel two.
This leads to two different asymptotic behaviors, dependent on the value of unary height.
Enumerating BCI/BCK terms \(|x| = 1\)

Special subclasses of lambda terms: \(BCI/BCK\) and their generalized versions: \(BCI(p)/BCK(p)\)
Attempts

Enumerating BCI/BCK terms \(|x| = 1\)

Special subclasses of lambda terms: \(BCI/BCK\) and their generalized versions: \(BCI(p)/BCK(p)\)

Bodini, Gardy, Gittenberger, Jacquot; 2013

The number of \(BCI(p)\)-terms of size \((2p + 1)n - 1\) is asymptotically

\[
\binom{2p - 2}{p - 1} \left(\frac{1.0844 \ldots}{p}\right)^{\frac{2p + 1}{2}} \left(\frac{(4p + 2)^p}{p!}\right)^{n - 1} \frac{p^2 - 2p}{n^{2p + 1}} (n - 1)!^p, \quad p \geq 2.
\]

K. Grygiel & P. Lescanne

Counting binary lambda terms

AofA 2014
Enumerating BCI/BCK terms ($|x| = 1$)

Special subclasses of lambda terms: $BCI/BCK$ and their generalized versions: $BCI(p)/BCK(p)$

Bodini, Gardy, Gittenberger, Jacquot; 2013

The number of $BCI(p)$-terms of size $(2p + 1)n - 1$ is asymptotically

$$\left(\frac{2p - 2}{p - 1}\right) \left(\frac{1.0844 \ldots}{2}ight) \frac{2p+1}{2} \left(\frac{(4p + 2)^p}{p!}\right)^{n-1} \frac{p^2 - 2p}{n^{2p+1}} (n - 1)!^p, \quad p \geq 2.$$  

Bodini, Gittenberger; 2014

The number of $BCK(2)$-terms of size $n$ is asymptotically

$$An^{2n/5}2^{n/5}e^{-2n/5} \exp\left(2^{-8/5}n^{4/5} + \frac{7 \cdot 2^{4/5}}{15}n^{3/5} - \frac{17 \cdot 2^{1/5}}{75}n^{2/5} - \frac{41 \cdot 2^{3/5}}{500}n^{1/5}\right)n^{-3/5},$$

where $A$ is some positive constant.
Motivation

- Lambda terms as programs
  → research on random programs and their behavior
Motivation

- Lambda terms as programs
  → research on random programs and their behavior

- Typed lambda terms as proofs of logical formulae
  → research on logical entities and their properties
Motivation

- Lambda terms as programs
  \(\rightarrow\) research on random programs and their behavior

- Typed lambda terms as proofs of logical formulae
  \(\rightarrow\) research on logical entities and their properties

- Lambda calculus as the basis for functional programming
  \(\rightarrow\) testing compilers by generating random lambda terms
Let’s eliminate names of variables from the notation of a \( \lambda \)-term:

\[
\begin{align*}
\text{variables } x, y, z, \ldots & \mapsto \text{de Bruijn indices } 1, 2, 3, \ldots \\
\lambda x. M & \mapsto \lambda M, \text{ where } M \text{ is the result of substituting each } x \text{ by } n, \\
\text{where } n & \text{ is the number of } \lambda \text{'s above the given occurrence of } x \\
MN & \mapsto MN
\end{align*}
\]
Lambda terms and de Bruijn indices

Regular lambda term

The lambda term \( \lambda x. (\lambda z. z) x (\lambda y. y (\lambda w. x)) \)
Lambda terms and de Bruijn indices

Lambda term in the de Bruijn version

The de Bruijn term $\lambda(\lambda_1)\lambda_1(\lambda_1(\lambda_3))$
Binary representation of lambda terms

Following John Tromp, we define the binary representation of de Bruijn indices in the following way:

\[
\begin{align*}
\hat{\lambda M} & = 00\hat{M}, \\
\hat{M N} & = 01\hat{M}\hat{N}, \\
\hat{i} & = 1'0.
\end{align*}
\]
Binary representation of lambda terms

Following John Tromp, we define the binary representation of de Bruijn indices in the following way:

\[
\begin{align*}
\hat{\lambda M} &= 00\hat{M}, \\
\hat{M N} &= 01\hat{M}\hat{N}, \\
\hat{i} &= 1^i0.
\end{align*}
\]

Given a \(\lambda\)-term, we define its size as the length of the corresponding binary sequence, i.e.,

\[
\begin{align*}
|i| &= i + 1, \\
|\lambda M| &= |M| + 2, \\
|M N| &= |M| + |N| + 2.
\end{align*}
\]
How many...?

Let $S_{m,n}$ denote the number of binary lambda terms of size $n$ with at most $m$ distinct free indices.
How many...?

Let $S_{m,n}$ denote the number of binary lambda terms of size $n$ with at most $m$ distinct free indices.

\[
S_{m,0} = S_{m,1} = 0,
\]

\[
S_{m,n+2} = [m \geq n + 1] + S_{m+1,n} + \sum_{k=0}^{n} S_{m,k}S_{m,n-k}.
\]

The sequence $(S_{0,n})_{n \geq 0}$ (A114852 in Sloane’s OEIS) enumerates closed binary lambda terms of size $n$. Its first 20 values are as follows:

0, 0, 0, 0, 1, 0, 1, 1, 2, 1, 6, 5, 13, 14, 37, 44, 101, 134, 298, 431.
Generating functions again in action

Now let us define the family of generating functions for sequences $(S_{m,n})_{n \geq 0}$:

$$s_m(z) = \sum_{n=0}^{\infty} S_{m,n} z^n.$$ 

Most of all, we are interested in the generating function for the number of closed terms, i.e.,

$$s_0(z) = \sum_{n=0}^{\infty} S_{0,n} z^n.$$
Recurring problem

Lemma

\[ s_m(z) = \frac{z^2(1 - z^m)}{1 - z} + z^2 s_{m+1}(z) + z^2 s_m(z)^2. \]
Enumerating binary terms

Recurring problem

Lemma

\[ s_m(z) = \frac{z^2 (1 - z^m)}{1 - z} + z^2 s_{m+1}(z) + z^2 s_m(z)^2. \]

This gives

\[ s_m(z) = \frac{1 - \sqrt{1 - 4z^4 \left( \frac{1 - z^m}{1 - z} + s_{m+1}(z) \right)}}{2z^2}. \]

Again we are stuck with infinitely nested radicals...
Let’s count all binary terms

Let $S_{\infty,n}$ denote the number of all (not necessarily closed) binary terms of size $n$.
We obtain the following recurrence relation:

\[
S_{\infty,0} = S_{\infty,1} = 0,
\]

\[
S_{\infty,n+2} = 1 + S_{\infty,n} + \sum_{k=0}^{n} S_{\infty,k}S_{\infty,n-k}.
\]

The sequence $(S_{\infty,n})_{n \geq 0}$ has the entry number A114851 in Sloane’s OEIS.
Its first 20 values are:

0, 0, 1, 1, 2, 2, 4, 5, 10, 14, 27, 41, 78, 126, 237, 399, 745, 1292, 2404, 4259.
No problem this time

Let $s_\infty(z)$ denote the generating function for the sequence $(S_{\infty,n})_{n \geq 0}$. 
No problem this time

Let $s_{\infty}(z)$ denote the generating function for the sequence $(S_{\infty,n})_{n \geq 0}$.

**Theorem**

The number of all binary lambda terms of size $n$ satisfies

$$S_{\infty,n} \sim \left(\frac{1}{\rho}\right)^n \cdot \frac{C}{n^{3/2}},$$

where $\rho = 0.509308127 \ldots$ and $C = 1.021874073 \ldots$. 
No problem this time

Let \( s_\infty(z) \) denote the generating function for the sequence \( (S_{\infty,n})_{n \geq 0} \).

**Theorem**

The number of all binary lambda terms of size \( n \) satisfies

\[
S_{\infty,n} \sim \frac{1}{\rho^n} \cdot \frac{C}{n^{3/2}},
\]

where \( \rho = 0.509308127 \ldots \) and \( C = 1.021874073 \ldots \).

For \( m \geq n - 1 \) we have \( S_{m,n} = S_{\infty,n} \). Therefore

\[
s_\infty(z) = \sum_{n=1}^{\infty} S_{n,n}z^n,
\]

which yields that \( [z^n]s_n = [z^n]s_\infty \).

Furthermore, \( s_\infty(z) = \lim_{m \to \infty} s_m(z) \).
Singularities of $s_m(z)$

**Lemma**

Let $\rho_m$ denote the dominant singularity of $s_m(z)$. Then for every $m$,

$$\rho_m = \rho_0.$$
Singularities of $s_m(z)$

Lemma

Let $\rho_m$ denote the dominant singularity of $s_m(z)$. Then for every $m$,

$$\rho_m = \rho_0.$$ 

Lemma

The dominant singularity of $s_0(z)$ is equal to the dominant singularity of $s_\infty(z)$, i.e.,

$$\rho_0 = \rho = 0.509308127\ldots$$
Sketch of the proof

Let us define functionals

\[
\Phi_m(F) = \frac{1 - \sqrt{1 - 4z^4 \left( \frac{1-z^m}{1-z} + F \right)}}{2z^2},
\]

\[
\Phi_\infty(F) = \frac{1 - \sqrt{1 - 4z^4 \left( \frac{1}{1-z} + F \right)}}{2z^2}.
\]
Sketch of the proof

Let us define functionals

\[ \Phi_m(F) = \frac{1 - \sqrt{1 - 4z^4 \left( \frac{1-z^m}{1-z} + F \right)}}{2z^2}, \]

\[ \Phi_\infty(F) = \frac{1 - \sqrt{1 - 4z^4 \left( \frac{1}{1-z} + F \right)}}{2z^2}. \]

Then

\[ s_m(z) = \Phi_m(s_{m+1}(z)). \]
Sketch of the proof

Let us define functionals

\[ \Phi_m(F) = \frac{1 - \sqrt{1 - 4z^4\left(\frac{1-z^m}{1-z} + F\right)}}{2z^2}, \]
\[ \Phi_\infty(F) = \frac{1 - \sqrt{1 - 4z^4\left(\frac{1}{1-z} + F\right)}}{2z^2}. \]

Then

\[ s_m(z) = \Phi_m(s_{m+1}(z)). \]

For \( z \in [0, \rho] \) we have

\[ \Phi_m(s_m(z)) \leq s_m(z) \leq s_{m+1}(z) \leq s_\infty(z). \]
Sketch of the proof II

Let $\tilde{s}_m(z)$ denote the fixed point of $\Phi_m$, i.e., $\tilde{s}_m(z)$ is defined as the solution of the equation

$$\tilde{s}_m(z) = \Phi_m(\tilde{s}_m(z)).$$
Let $\tilde{s}_m(z)$ denote the fixed point of $\Phi_m$, i.e., $\tilde{s}_m(z)$ is defined as the solution of the equation

$$\tilde{s}_m(z) = \Phi_m(\tilde{s}_m(z)).$$

For $z \in [0, \rho]$ we have

$$\tilde{s}_m(z) \leq s_m(z) \leq s_\infty(z).$$
Sketch of the proof II

Let $\tilde{s}_m(z)$ denote the fixed point of $\Phi_m$, i.e., $\tilde{s}_m(z)$ is defined as the solution of the equation

$$\tilde{s}_m(z) = \Phi_m(\tilde{s}_m(z)).$$

For $z \in [0, \rho]$ we have

$$\tilde{s}_m(z) \leq s_m(z) \leq s_\infty(z).$$

By the definition of $\tilde{s}_m(z)$,

$$z^2\tilde{s}_m^2(z) - (1 - z^2)^2\tilde{s}_m(z) + \frac{z^2(1 - z^m)}{1 - z} = 0.$$
Sketch of the proof II

Let $\tilde{s}_m(z)$ denote the fixed point of $\Phi_m$, i.e., $\tilde{s}_m(z)$ is defined as the solution of the equation

$$\tilde{s}_m(z) = \Phi_m(\tilde{s}_m(z)).$$

For $z \in [0, \rho]$ we have

$$\tilde{s}_m(z) \leq s_m(z) \leq s_\infty(z).$$

By the definition of $\tilde{s}_m(z)$,

$$z^2\tilde{s}_m^2(z) - (1 - z^2)^2\tilde{s}_m(z) + \frac{z^2(1 - z^m)}{1 - z} = 0.$$

Let us denote the main singularity of $\tilde{s}_m(z)$ by $\sigma_m$. We have

$$\sigma_m \geq \rho_m \geq \rho.$$
Sketch of the proof III

The discriminant of this equation is

$$\Delta_m = (1 - z^2)^2 - \frac{4z^4(1 - z^m)}{1 - z}.$$
Sketch of the proof III

The discriminant of this equation is

\[ \Delta_m = (1 - z^2)^2 - \frac{4z^4(1 - z^m)}{1 - z}. \]

The value of \( \sigma_m \) is equal to the root of smallest modulus of the polynomial

\[ P_m(z) := (z - 1)\Delta_m = 4z^4(1 - z^m) - (1 - z)^3(1 + z)^2. \]
The discriminant of this equation is

$$\Delta_m = (1 - z^2)^2 - \frac{4z^4(1 - z^m)}{1 - z}. $$

The value of $\sigma_m$ is equal to the root of smallest modulus of the polynomial

$$P_m(z) := (z - 1)\Delta_m = 4z^4(1 - z^m) - (1 - z)^3(1 + z)^2. $$

The sequence $(\sigma_m)_{m \in \mathbb{N}}$ of roots of polynomials $P_m(z)$ is decreasing and it converges to $\rho$. 
Enumerating binary terms

Sketch of the proof III

The discriminant of this equation is

$$\Delta_m = (1 - z^2)^2 - \frac{4z^4(1 - z^m)}{1 - z}.$$  

The value of $\sigma_m$ is equal to the root of smallest modulus of the polynomial

$$P_m(z) := (z - 1)\Delta_m = 4z^4(1 - z^m) - (1 - z)^3(1 + z)^2.$$  

The sequence $(\sigma_m)_{m \in \mathbb{N}}$ of roots of polynomials $P_m(z)$ is decreasing and it converges to $\rho$.

By $\sigma_m \geq \rho_m \geq \rho$, we get $\rho_m \to \rho$, as well.
Sketch of the proof III

The discriminant of this equation is

$$\Delta_m = (1 - z^2)^2 - \frac{4z^4(1 - z^m)}{1 - z}.$$

The value of $\sigma_m$ is equal to the root of smallest modulus of the polynomial

$$P_m(z) := (z - 1)\Delta_m = 4z^4(1 - z^m) - (1 - z)^3(1 + z)^2.$$

The sequence $(\sigma_m)_{m \in \mathbb{N}}$ of roots of polynomials $P_m(z)$ is decreasing and it converges to $\rho$.

By $\sigma_m \geq \rho_m \geq \rho$, we get $\rho_m \to \rho$, as well.

Since all the $\rho_m$’s are equal, we obtain that $\rho_m = \rho$ for every natural $m$. 
The number of closed binary lambda terms of size $n$ is of exponential order

$$S_{0,n} \asymp 1.963448 \ldots^n.$$
Conjecture

For every \( m, \quad S_{m,n} \sim o(1.963448 \ldots n \cdot n^{-3/2}). \)
Conjecture

For every \( m \), \( S_{m,n} \sim o(1.963448 \ldots n \cdot n^{-3/2}) \).

\[
\frac{S_{m,n}}{(\rho^n n^{-3/2})}
\]


R. David, K. Grygiel, J. Kozik, Ch. Raffalli, G. Theyssier, M. Zaionc: *Asymptotically almost all lambda terms are strongly normalizing*, Logical Methods in Computer Science Vol. 9(1:02) 2013, pp. 1-30


K. Grygiel, P. Lescanne: *Counting terms in the binary lambda calculus*, 2014
