Random lambda terms: how they look like and how they behave

Katarzyna Grygiel

Joint work with: René David, Jakub Kozik, Christophe Raffalli, Guillaume Theyssier, Marek Zaionc

CLA, Lyon, 5th June 2009

Outline

1. Preliminaries: quest for the number

Outline

1. Preliminaries: quest for the number
2. Let's make it easier: combinatory logic

Outline

1. Preliminaries: quest for the number
2. Let's make it easier: combinatory logic
3. Measuring trees

Outline

1. Preliminaries: quest for the number
2. Let's make it easier: combinatory logic
3. Measuring trees
4. How does a random term look like?

Outline

1. Preliminaries: quest for the number
2. Let's make it easier: combinatory logic
3. Measuring trees
4. How does a random term look like?
5. One can judge a random term by its cover

Outline

Slogan

Probabilistic methods appear to be very powerful in computer science. A point of view of those methods is that we investigate a typical object chosen from the set.

Enumerating lambda terms

We are dealing with untyped lambda calculus terms.

Enumerating lambda terms

We are dealing with untyped lambda calculus terms.
Size of a lambda term:

$$
\begin{aligned}
|x| &= 0 \\
|\lambda x.M| &= 1 + |M| \\
|MN| &= 1 + |M| + |N|
\end{aligned}
$$

Enumerating lambda terms

We are dealing with untyped lambda calculus terms.
Size of a lambda term:

$$\begin{aligned} |x| &= 0 \\ |\lambda x.M| &= 1 + |M| \\ |MN| &= 1 + |M| + |N| \end{aligned}$$

Problem

How many lambda terms of a given size are there?

Density

Let $n \in \mathbb{N}$ and let $\Lambda_n(\mathcal{P})$ be a set of lambda terms of size n with a given property $\mathcal{P}$. We say that the density of terms with the property $\mathcal{P}$ is equal to $d \in [0,1]$ iff the limit

$$\lim_{n\to\infty} \frac{\#\Lambda_n(\mathcal{P})}{\#\Lambda_n}$$

exists and is equal to d.

Motivation

Question

How many programs have the halting property?

Motivation

Question

How many programs have the halting property?

In terms of lambda calculus: compute the following limit

$$\lim_{n\to\infty} \frac{\#\Lambda_n(SN)}{\#\Lambda_n} =?$$

where SN denotes the property of strong normalization.

Combinatorial view

Lambda tree

A lambda tree is a unary-binary tree in which each leaf can be labelled in as many ways as many vertices with exactly one child occur on the path from the leaf to the root of the tree.

Combinatorial view

Lambda tree

A lambda tree is a unary-binary tree in which each leaf can be labelled in as many ways as many vertices with exactly one child occur on the path from the leaf to the root of the tree.

Problem (once again)

How many lambda trees of a given size are there?

Enumerating combinatory logic terms

The set CL of combinators is defined by the following grammar:

$$M := \mathbf{S} \mid \mathbf{K} \mid \mathbf{I} \mid (MM)$$

Enumerating combinatory logic terms

The set CL of combinators is defined by the following grammar:

$$M := \mathbf{S} \mid \mathbf{K} \mid \mathbf{I} \mid (MM)$$

Reduction rules defined on CL:

$$\begin{aligned} \mathbf{I}\ P &= P \\ \mathbf{K}\ P\ Q &= P \\ \mathbf{S}\ P\ Q\ R &= (P\ R)(Q\ R) \end{aligned}$$

Enumerating combinatory logic terms

Size of a CL term:

$$
\begin{aligned}
|\mathbf{S}|, |\mathbf{K}|, |\mathbf{I}| &= 0 \\
|MN| &= 1 + |M| + |N|
\end{aligned}
$$

Enumerating combinatory logic terms

Size of a CL term:

$$|\mathbf{S}|, |\mathbf{K}|, |\mathbf{I}| = 0$$
$$|MN| = 1 + |M| + |N|$$

Number of CL terms

$$\#CL_n = 3^{n+1} \cdot C(n+1)$$

A random combinator does not terminate

Theorem

Let t_0 be a combinator. The density of combinators with t_0 as a subterm is 1.

A random combinator does not terminate

Theorem

Let t_0 be a combinator. The density of combinators with t_0 as a subterm is 1.

Let us take $t_0 = \mathbf{S}\ \mathbf{I}\ \mathbf{I}\ (\mathbf{S}\ \mathbf{I}\ \mathbf{I})$.

A random combinator does not terminate

Theorem

Let t_0 be a combinator. The density of combinators with t_0 as a subterm is 1.

Let us take $t_0 = \mathbf{S}\ \mathbf{I}\ \mathbf{I}\ (\mathbf{S}\ \mathbf{I}\ \mathbf{I})$.

Since t_0 is not strongly normalizing, we get

Corollary

The density of strongly normalizing combinators is 0.

Bounds

For $\varepsilon > 0$ and for n big enough we have the following bounds:

Theorem

Bounds

For $\varepsilon > 0$ and for n big enough we have the following bounds:

Theorem

1. lower bound:

$$\#\Lambda_n \geq \left(\frac{(4-\varepsilon)n}{\ln(n)}\right)^{n-\frac{n}{\ln(n)}}$$

Bounds

For $\varepsilon > 0$ and for n big enough we have the following bounds:

Theorem

1. lower bound:

$$\#\Lambda_n \geq \left(\frac{(4-\varepsilon)n}{\ln(n)}\right)^{n-\frac{n}{\ln(n)}}$$

2. upper bound:

$$\#\Lambda_n \leq \left(\frac{(12+\varepsilon)n}{\ln(n)}\right)^{n-\frac{n}{3\ln(n)}}$$

Number of lambdas

Applying both bounds, we can estimate the number of lambdas in a random term:

Number of lambdas

Applying both bounds, we can estimate the number of lambdas in a random term:

Theorem

1. The density of terms with more than $\frac{3n}{\ln(n)}$ λ's is 0.

Number of lambdas

Applying both bounds, we can estimate the number of lambdas in a random term:

Theorem

1. The density of terms with more than $\frac{3n}{\ln(n)}$ λ's is 0.
2. The density of terms with less than $\frac{n}{3\ln(n)}$ λ's is 0.

Depth, width. . .

Let t be a lambda tree.

1. The unary height of a term t is the maximal number of unary nodes in a branch of t.

Depth, width. . .

Let t be a lambda tree.

1. The unary height of a term t is the maximal number of unary nodes in a branch of t.
2. The binary height of a term t is the maximal number of binary nodes in a branch of t.

Depth, width. . .

Let t be a lambda tree.

1. The unary height of a term t is the maximal number of unary nodes in a branch of t.
2. The binary height of a term t is the maximal number of binary nodes in a branch of t.
3. Two λ's in t are called incomparable if there is no branch containing both of them.
   The λ-width of t is the maximal number of pairwise incomparable binding λ's.

How high is a lambda tree?

Heights

How high is a lambda tree?

Heights

1. A random lambda tree has the unary height greater than $\frac{n}{3\ln(n)}$.

How high is a lambda tree?

Heights

1. A random lambda tree has the unary height greater than $\frac{n}{3\ln(n)}$.
2. A random lambda tree has the binary height less than $\frac{n}{\sqrt{\ln(\ln(n))}}$.

Head lambdas

Let t be a term and $u = \lambda x.a$ be a subterm of t. We say that λx is in head position if $t = \lambda \overrightarrow{y}.u$ for some, possibly empty, sequence of abstractions $\lambda \overrightarrow{y}$.

Head lambdas

Let t be a term and $u = \lambda x.a$ be a subterm of t. We say that λx is in head position if $t = \lambda \overrightarrow{y}.u$ for some, possibly empty, sequence of abstractions $\lambda \overrightarrow{y}$.

Theorem

Let $g(n) \in o\left(\sqrt{n/\ln(n)}\right)$. The density of terms with less than $g(n)$ head lambdas is 0.

Head lambdas

Many head lambdas are useful, i.e. binding:

Head lambdas

Many head lambdas are useful, i.e. binding:

Theorem

Let $g(n) \in o\left(\sqrt{n/\ln(n)}\right)$. The density of terms in which there is at least one lambda among $g(n)$ head lambdas that does not bind any variable is 0.

Head lambdas

What is more, altogether they bind many occurences of variables:

Head lambdas

What is more, altogether they bind many occurences of variables:

Theorem

Let $g(n) \in o\big(\ln(n)\big)$. The density of terms in which the total number of occurrences of variables bound by the first three lambdas is at most $g(n)$ is 0.

Head lambdas

Actually, many head lambdas bind many occurences:

Head lambdas

Actually, many head lambdas bind many occurences:

Theorem

For any fixed integers k and k_0, the density of terms in which each of the first k lambdas binds more than k_0 variables is 1.

No closed subterms

Theorem

Let t_0 be a term of size k_0 with k occurrences of free variables. Assume $k_0 \geq k+1$. Then the density of terms with t_0 as a subterm is 0.

No closed subterms

Theorem

Let t_0 be a term of size k_0 with k occurrences of free variables. Assume $k_0 \geq k + 1$. Then the density of terms with t_0 as a subterm is 0.

Corollary

Let t_0 be a term. If

No closed subterms

Theorem

Let t_0 be a term of size k_0 with k occurrences of free variables. Assume $k_0 \geq k + 1$. Then the density of terms with t_0 as a subterm is 0.

Corollary

Let t_0 be a term. If

1. t_0 is closed or

No closed subterms

Theorem

Let t_0 be a term of size k_0 with k occurrences of free variables. Assume $k_0 \geq k + 1$. Then the density of terms with t_0 as a subterm is 0.

Corollary

Let t_0 be a term. If

1. t_0 is closed or
2. there are at least two lambdas in t_0

No closed subterms

Theorem

Let t_0 be a term of size k_0 with k occurrences of free variables. Assume $k_0 \geq k + 1$. Then the density of terms with t_0 as a subterm is 0.

Corollary

Let t_0 be a term. If

1. t_0 is closed or
2. there are at least two lambdas in t_0

then the density of terms with t_0 as a subterm is 0.

Width of a tree

Theorem

The lambda width of a random term is at most 2.

Width of a tree

Theorem

The lambda width of a random term is at most 2.

Now we know that a random lambda tree:

Width of a tree

Theorem

The lambda width of a random term is at most 2.

Now we know that a random lambda tree:
— starts with many useful unary nodes,

Width of a tree

Theorem

The lambda width of a random term is at most 2.

Now we know that a random lambda tree:
— starts with many useful unary nodes,
— ends with many binary nodes,

Width of a tree

Theorem

The lambda width of a random term is at most 2.

Now we know that a random lambda tree:

— starts with many useful unary nodes,

— ends with many binary nodes,

— has many leaves.

Width of a tree

Theorem

The lambda width of a random term is at most 2.

Now we know that a random lambda tree:
— starts with many useful unary nodes,
— ends with many binary nodes,
— has many leaves.
Thus, it looks a bit like. . .

. . . a stone pine

Safe terms

1. Let t be a term of width 1. We say that t is unfair if the highest binding lambda in t occurs in its leftmost branch.

Safe terms

1. Let t be a term of width 1. We say that t is unfair if the highest binding lambda in t occurs in its leftmost branch.
2. We say that a term t is safe if
   - either its width is less than 2
   - or its width is equal to 2 and its smallest subterm $(u\ v)$ of width 2 is such that either u or v is not unfair.

Safe terms

1. Let t be a term of width 1. We say that t is unfair if the highest binding lambda in t occurs in its leftmost branch.
2. We say that a term t is safe if
   - either its width is less than 2
   - or its width is equal to 2 and its smallest subterm $(u\ v)$ of width 2 is such that either u or v is not unfair.

Theorem

A safe term is strongly normalizing.

Different language, different results

Theorem

A random term is safe.

Different language, different results

Theorem

A random term is safe.

Corollary

A random term is strongly normalizing.

What next?

- Other notions of size

What next?

- Other notions of size
- Typable terms

What next?

- Other notions of size
- Typable terms
- Other patterns, e.g. $\lambda x.y$

What next?

- Other notions of size
- Typable terms
- Other patterns, e.g. $\lambda x.y$
- . . . **Number of terms**

The end.