Enumerating Lambda Terms By Weighted Length of Their de Bruijn Representation

Bernhard Gittenberger

joint work with Olivier Bodini and Zbigniew Gołębiewski

Institut für Diskrete Mathematik und Geometrie
Technische Universität Wien

and

LIPN
Université Paris Nord

Department of Computer Science
Wrocław University of Technology

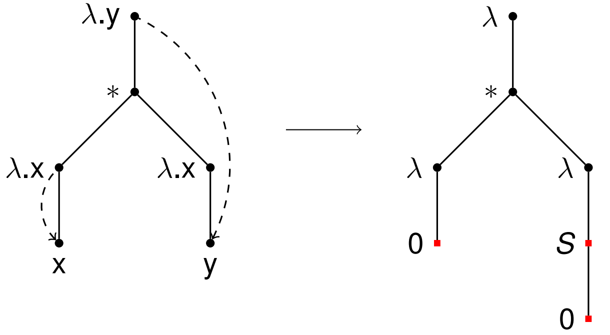
AofA17, Princeton, June 18, 2017

Definition of lambda terms

$$T ::= x \mid \lambda x.T \mid T * T \quad \rightarrow \quad T ::= S^n 0 \mid \lambda T \mid T * T$$

$\lambda x.T$: abstraction, unary node $\qquad (T * T)$: application, binary node

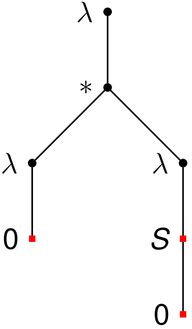
$$\lambda y.((\lambda x.x) * (\lambda x.y)) \rightarrow \lambda(\lambda 1 * \lambda 2) \rightarrow \lambda((\lambda 0) * (\lambda(S0)))$$



m-open lambda terms

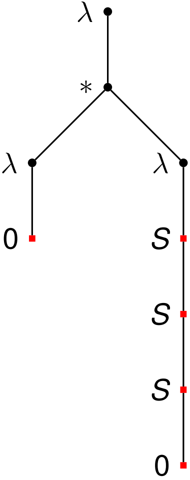
closed lambda term (0-open)

$\lambda((\lambda 0) * (\lambda(S0)))$



2-open lambda term

$\lambda((\lambda 0) * (\lambda(SSS0)))$



General notion of size

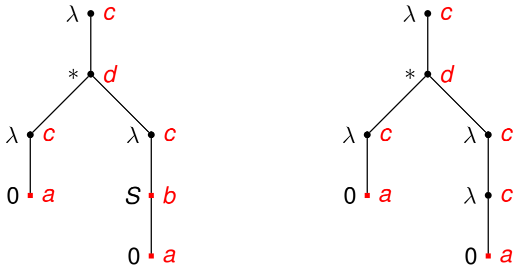
$$\begin{aligned} |0| &= a \\ |S| &= b \\ |\lambda M| &= |M| + c \\ |MN| &= |M| + |N| + d. \end{aligned}$$

Assumptions

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

General notion of size

$$\lambda((\lambda 0) * (\lambda(S0))) \quad \neq \quad \lambda((\lambda 0) * (\lambda(\lambda 0)))$$



size: $2a + b + 3c + d$ size: $2a + 4c + d$

- natural counting (Bendkowski, Grygiel, Lescanne, Zaionc 2015): $a = b = c = d = 1$
- less natural counting (Bendkowski, Grygiel, Lescanne, Zaionc 2015): $a = 0,\ b = c = 1,\ d = 2$
- binary lambda calculus (Tromp 2006): $b = 1,\ a = c = d = 2$

Combinatorial specification and lambda terms

$$\mathcal{L} = \textsc{Seq}(\mathcal{S}) \times \mathcal{Z} \;\cup\; \mathcal{U} \times \mathcal{L} \;\cup\; \mathcal{A} \times \mathcal{L}^2$$

- $\mathcal{L}$ – the class of lambda terms,
- $\mathcal{Z}$ – the class of zeros,
- $\mathcal{S}$ – the class of successors,
- $\mathcal{U}$ – the class of abstractions,
- $\mathcal{A}$ – the class of applications.

Remark: $\mathcal{Z}, \mathcal{S}, \mathcal{U}, \mathcal{A}$ contain only one atomic object.
Thus

$$L(z) = z^a \sum_{j=0}^{\infty} z^{bj} + z^c L(z) + z^d L(z)^2,$$

$[z^n]L(z)$=number of lambda terms of size n.

m-open terms and functional equations

Let

$$\mathcal{L}_m = \textsc{Seq}_{\leq m-1}(\mathcal{S}) \times \mathcal{Z} \;\cup\; \mathcal{U} \times \mathcal{L}_{m+1} \;\cup\; \mathcal{A} \times \mathcal{L}_m^2.$$

- $L_{m,n}$ – the number of m-open lambda terms of size n,
- $L_m(z) = \sum_{n\geq 0} L_{m,n} z^n \quad ([z^n]\, L_m(z) = L_{m,n})$

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

- $L_0(z)$ is the gen. fun. of the set $\mathcal{L}_0$ of closed lambda terms,
- $L_\infty(z)$ is the gen. fun. of the set $\mathcal{L}_\infty = \mathcal{L}$ of all lambda terms.

$L_\infty(z)$ – all terms

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

Proposition

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

$$L_\infty(z) = a_\infty - b_\infty \sqrt{1 - \frac{z}{\rho}} + \mathrm{O}\left(\left|1 - \frac{z}{\rho}\right|\right),$$

for some constants $a_\infty > 0, b_\infty > 0$ *that depend on* a, b, c, d. *Thus the coefficients of* $L_\infty(z)$ *satisfy*

$$L_{\infty,n} \sim \frac{b_\infty}{2\sqrt{\pi}} \rho^{-n} n^{-3/2}, \text{ as } n \to \infty.$$

Idea: Replace $\mathcal{L}_m$ by pruning de Bruijn indices

We have

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

$\mathcal{L}_m^{(h)}$ – lambda terms in $\mathcal{L}_m$ where the length of each string of successors is bounded by h

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

because for $m \geq h$ we have $L_m^{(h)}(z) = L_h^{(h)}(z)$.
$\rightsquigarrow$ upper and lower bounds.

Theorem (G., Gołębiewski 2016)

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

$$\liminf_{n\to\infty} \frac{L_{m,n}}{\underline{C} n^{-\frac{3}{2}} \rho^{-n}} \geq 1 \quad \mathit{and} \quad \limsup_{n\to\infty} \frac{L_{m,n}}{\overline{C} n^{-\frac{3}{2}} \rho^{-n}} \leq 1,$$

Remark

In case of given a, b, c, d *and* m *we can compute numerically such constants* $\underline{C}$ *and* $\overline{C}$*.*

Theorem (G., Gołębiewski 2016)

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

$$\liminf_{n\to\infty} \frac{L_{m,n}}{\underline{C} n^{-\frac{3}{2}} \rho^{-n}} \geq 1 \quad \mathit{and} \quad \limsup_{n\to\infty} \frac{L_{m,n}}{\overline{C} n^{-\frac{3}{2}} \rho^{-n}} \leq 1,$$

Remark

In case of given a, b, c, d *and* m *we can compute numerically such constants* $\underline{C}$ *and* $\overline{C}$.
For instance, for natural counting we have

$$\underline{C}^{(nat)} \approx 0.0779099{\color{red}5266}\ldots,$$
$$\overline{C}^{(nat)} \approx 0.0779099{\color{red}8229}\ldots.$$

Theorem

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

$$L_{m,n} \sim C n^{-\frac{3}{2}} \rho^{-n}$$

Theorem

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

$$L_{m,n} \sim Cn^{-\frac{3}{2}}\rho^{-n}$$

Proof idea: We know that

$$\lim_{m\to\infty} L_m(z) = L_\infty(z) = a_\infty - b_\infty\sqrt{1-\frac{z}{\rho}}.$$

Theorem

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

$$L_{m,n} \sim Cn^{-\frac{3}{2}}\rho^{-n}$$

Proof idea: We know that

$$\lim_{m\to\infty} L_m(z) = L_\infty(z) = a_\infty - b_\infty\sqrt{1-\frac{z}{\rho}}.$$

Replace $L_m(z)$ by $L_\infty(z)$ and trace back:

$$a_m := a_\infty, \qquad b_m := b_\infty; \qquad L_{m,m}(z) := L_\infty(z).$$

$$L_{m,m}(z) = L_\infty(z) = a_\infty - b_\infty\sqrt{1 - \frac{z}{\rho}},$$

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

Then, eventually we obtain

$$L_{0,m}(z) = a_{0,m} - b_{0,m}\sqrt{1 - \frac{z}{\rho}}$$

$$L_{0,m}(z) = a_{0,m} - b_{0,m}\sqrt{1 - \frac{z}{\rho}}$$

Lemma

The sequences $(a_{0,m})_{m\geq 0}$ and $(b_{0,m})_{m\geq 0}$ are convergent.

Proof.

We know that $\lim_{m\to\infty} L_{0,m}(z) = L_0(z)$, uniformly in $[0,\rho]$ and that $L_{0,m}(z)$ is decreasing. Thus

$$a_{0,m} = L_{0,m}(\rho) \longrightarrow L_0(\rho) =: a_0, \text{ as } m \to \infty.$$

b_0, m is increasing and bounded by b_∞, thus converges to b_0. $\square$

The theorem follows now from the uniform convergence of $L_{0,m}(z)$ and the local shape of these functions.

Boltzmann sampling

Singular Boltzmann output size according to Boltzmann distribution

$$\mathbb{P}\{N = n\} = \frac{a_n \rho^n}{A(\rho)}$$

where $A(\rho)$ is the generating function and ρ its dominant singularity.

Boltzmann sampling

Singular Boltzmann output size according to Boltzmann distribution

$$\mathbb{P}\{N = n\} = \frac{a_n \rho^n}{A(\rho)}$$

where $A(\rho)$ is the generating function and ρ its dominant singularity.

The sampler: Construct superclass $\mathcal{L}_{N,0} \supseteq \mathcal{L}_0$ tending to $\mathcal{L}_0$ and reject unwanted results:

$$\begin{cases} L_{N,0} & = zL_{N,1} + zL_{N,0}^2, \\ L_{N,1} & = zL_{N,2} + zL_{N,1}^2 + z, \\ L_{N,2} & = zL_{N,3} + zL_{N,2}^2 + z + z^2 \quad , \\ \dots & = \dots, \\ L_{N,N-1} & = zL_{N,N} + zL_{N,N-1}^2 + z\dfrac{1 - z^{N-1}}{1 - z}, \\ L_{N,N} & = zL_{N,N} + zL_{N,N}^2 + \dfrac{z}{1 - z} \end{cases}$$

Rejection if de Bruijn index larger than N drawn in ΓL_N.

Boltzmann sampling – Costs

Without extra rejection:

- traditional framework of singular Boltzmann sampling
- linear if target size is in $((1-\varepsilon)n, (1+\varepsilon)n)$

Boltzmann sampling – Costs

Without extra rejection:

- traditional framework of singular Boltzmann sampling
- linear if target size is in $((1-\varepsilon)n, (1+\varepsilon)n)$

Extra rejection:

- Costs bounded by object size

Boltzmann sampling – Costs

Without extra rejection:

- traditional framework of singular Boltzmann sampling
- linear if target size is in $((1-\varepsilon)n, (1+\varepsilon)n)$

Extra rejection:

- Costs bounded by object size
- Unwanted terms are open terms in $\mathcal{L}_{0,N}$:

$$\frac{[z^n]L_0(z)}{[z^n]L_{N,0}(z)} \longrightarrow 1.$$

Speed is exponential: For $N = 20$, the proportion of closed terms is 0.999999998.

Boltzmann sampling – Experiments

Experiments with $N = 20$:

$\mathbb{P}\{\text{unary}\} \approx 0.2955977425, \qquad \mathbb{P}\{\text{binary}\} = \mathbb{P}\{\text{leaf}\} \approx 0.3522011287.$

Boltzmann sampling – Experiments

Experiments with $N = 20$:

$\mathbb{P}\{\text{unary}\} \approx 0.2955977425, \qquad \mathbb{P}\{\text{binary}\} = \mathbb{P}\{\text{leaf}\} \approx 0.3522011287.$

Average number of leaves in a term of size n is asymptotically $0.3522011287 \cdots \cdot n$.

Boltzmann sampling – Experiments

Experiments with $N = 20$:

$\mathbb{P}\{\text{unary}\} \approx 0.2955977425, \qquad \mathbb{P}\{\text{binary}\} = \mathbb{P}\{\text{leaf}\} \approx 0.3522011287.$

Average number of leaves in a term of size n is asymptotically $0.3522011287 \cdots n$.
$X_N := \#$ leaves in $\mathcal{L}_{N,0} \stackrel{d}{=} \#$ leaves in $\mathcal{L}_{N,N}$.

Thus, all moments of X_N and $X := \#$ leaves in a closed term are asymptotically equal.

Theorem

Let X_n the number of variables in a lambda-term of size n. Then

$$X_n \sim \mathcal{N}(\mu n, \sigma^2 n)$$

where μ and σ^2 tend to $\frac{1-\rho}{2} \approx 0.3522011287$.

Boltzmann sampling – Experiments

Sampler has same complexity as sampler for trees (linear in approximate size)

On laptop with CPU i7-5600U, clock rate 2.6 GHz, it is possible to draw a lambda-term of size in the range [1 000 000, 2 000 000] in less than 10 minutes.

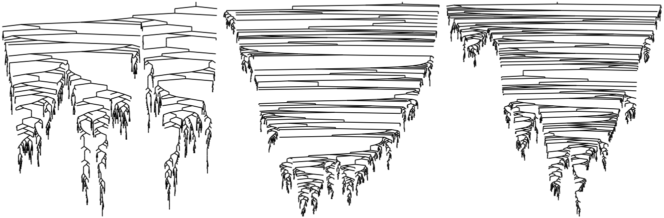


Figure: Three uniform random lambda-terms of size 2098, 2541, 2761.

Related work and perspectives

- Study further properties of these terms
- terms of bounded unary height
  cf. Bodini, Gardy, G. 2011 and Bodini, Gardy, G, Gołębiewski 2016
  - Shape characteristics of terms with bounded unary height (G., Larcher, in progress)
- restricting the number of variables bound by an abstraction
  cf. Bodini, Gardy, Jacquot 2010 (BCI/BCK), Bodini, Gardy, G., Jacquot 2013(gen. BCI), Bodini, G. 2014 (BCK_2)
  - Shape characteristics of BCI/BCK/gen BCI terms (G., Larcher, in progress)

Thank you!