Properties of random lambda and combinatory logic terms

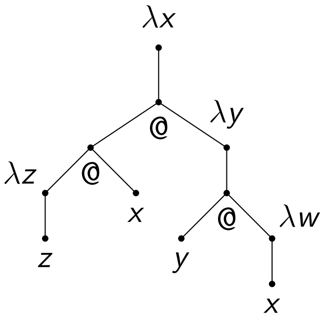
Maciej Bendkowski

Theoretical Computer Science

Jagiellonian University

CLA 2017
Göteborg, May 2017

λ-terms in the classic notation



$$\lambda x.((\lambda y.y)x)(\lambda z.z(\lambda w.x))$$

Canonical representation

David, Grygiel, Kozik, Raffalli, Theyssier, Zaionc (2013)

Let $\varepsilon \in (0, 4)$ and $\delta > 0$. Then, the number L_n of closed λ-terms of size n (modulo α-conversion) satisfies:

$$\left(\frac{(4-\varepsilon)}{\log n}\right)^{n-\frac{n}{\log n}} \lesssim L_n \lesssim \left(\frac{(12+\delta)}{\log n}\right)^{n-\frac{n}{3\log n}}$$

Canonical representation (II)

David, Grygiel, Kozik, Raffalli, Theyssier, Zaionc (2013)

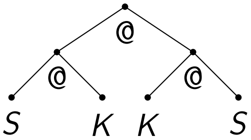
Asymptotically almost all λ-terms are strongly normalising.

Moral:

- Large random λ-terms represent 'safe' computations;
- ... however their analysis is quite difficult and technical
- ... and moreover we cannot efficiently generate them†.

†except for some restricted classes of linear and affine λ-terms, see [Bodini, Gardy, Jacquot (2013) and Bodini, Gardy, Gittenberger Jacquot (2013)].

Combiantors in the classic notation



$SK(KS)$

Combiantors in the classic notation (II)

David, Grygiel, Kozik, Raffalli, Theyssier, Zaionc (2013)

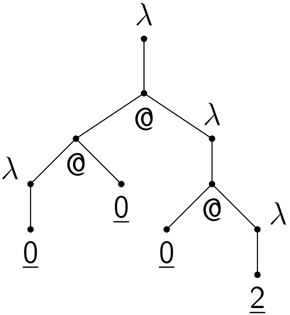
Asymptotically almost *no* combinator is strongly normalising.

Moral:

- Large random combinators represent 'unsafe' computations;
- Fortunately their analysis is moderately easy
- . . . and moreover we can efficiently generate them[†].

[†]for instance, using Rémy's exact-size sampler for binary trees, see [Rémy, Un procédé itératif de dénombrement d'arbres binaires et son application a leur génération aléatoire, 1985].

λ-terms in the de Bruijn notation



$$\lambda((\lambda\underline{0})\underline{0})(\lambda\underline{0}(\lambda\underline{2}))$$

De Bruijn representation

Bendkowski, Grygiel, Lescanne, Zaionc (2016)

The number L_n of plain (closed or open) λ-terms of size n in the unary de Bruijn representation satisfies:

$$L_n \sim C\rho^n n^{-3/2} \quad \text{where} \quad \rho \approx 0.2955 \quad C \approx 0.6067$$

Combinatorial specification:

$$\begin{aligned} \mathcal{D} &= \underline{0} \mid \mathrm{succ}(\mathcal{D}) \\ \mathcal{L} &= \lambda\mathcal{L} \mid \mathcal{L}\,\mathcal{L} \mid \mathcal{D}\,. \end{aligned}$$

De Bruijn representation (II)

Bendkowski, Grygiel, Lescanne, Zaionc (2016)

Asymptotically almost *no* is strongly normalising.

Proof sketch (idea dates back to [DGKRTZ'13]):

- Show that λ-terms exhibit the fixed-subterm property, i.e. for a fixed T, asymptotically almost all λ-terms contain T as a subterm.

Notable consequences:

- Large random λ-terms are not typeable;
- Generalises to properties spanning 'upwards' in terms.

De Bruijn representation (III)

Some notable statistical properties of random λ-terms:[1]

- Constant number of head abstractions, ≈ 0.4196 (sic!);
- Constant average index value, ≈ 1.41964 (sic!).

[1] ongoing work with Olivier Bodini and Sergey Dovgal.

Random generation: techniques

Available methods for random generation of λ-terms:

- ad-hoc bijection-based methods;
- Boltzmann models and rejection sampling techniques.

Status quo:

- Closed λ-terms can be effectively sampled using Boltzmann samplers and rejection techniques (achievable sizes $\geq 100,000$);
- Typeable λ-terms can be effectively sampled… up to sizes of ≈ 140 combining Boltzmann models and logic programming techniques[2].

[2] see [Bendkowski, Grygiel, Tarau (2017)]

Random generation: software

- github.com/fredokun/arbogen
- github.com/Lysxia/generic-random
- github.com/maciej-bendkowski/lambda-sampler
- github.com/maciej-bendkowski/boltzmann-brain

Open problems

Bias selection of open problems (I)

Effective random generation of closed typeable λ-terms in either the de Bruijn representation or the canonical one.

Current challenges:

- We don't know how to specify combinatorially the set of closed typeable λ-terms nor an asymptotically significant portion of them.
- How to overcome the context-sensitive nature of typeability with intrinsically context-free methods?

Bias selection of open problems (II)

What is the average 'computational complexity' of large random λ-terms in the de Bruijn notation?

Bendkowski (2017)

There exists an effective, infinite hierarchy of regular tree grammars capturing the set of normalising combinators.

Notable consequence:

- We can analyse the structure of large classes of normalising combinators. For instance, roughly 34% reduce under seven reduction steps[3].

[3] See [Bendkowski, Grygiel, Zaionc (2017)]

Bias selection of open problems (IIa)

Main challenge:

There exists no effective method of answering the following question – what is the number of terminating computations (λ-terms, combinators, etc.) of size n?

Proof idea:

- Assume the contrary and solve the halting problem.

In consequence, there's no computable specification for terminating computations...

Bias selection of open problems (IIb)

Current plan:

- Consider variants of λ-calculus with explicit substitution, for instance $\lambda\upsilon$[†] (read lambda upsilon).

Interesting things might happen…[4]

… and many more intriguing problems remain.

[†] see [Lescanne, From lambda-sigma to lambda-upsilon a journey through calculi of explicit substitutions, 1994].

[4] Spoiler: $\lambda\upsilon$ terms are counted by Catalan numbers! (ongoing work)

Some references

- R. David, K. Grygiel, J. Kozik, C. Raffalli, G. Theyssier, M. Zaionc. *Asymptotically almost all λ-terms are strongly normalizing*. Logical Methods in Computer Science, 2017.
- O. Bodini, D. Gardy, B. Gittenberger, A. Jacquot. "Enumeration of Generalized BCI Lambda-terms". Electronic Journal of Combinatorics 2013.
- O. Bodini, D. Gardy, A. Jacquot. "Asymptotics and random sampling for BCI and BCK lambda terms". Theoretical Computer Science, 2013.
- M. Bendkowski, K. Grygiel, P. Lescanne, M. Zaionc. *Combinatorics of λ-terms*. Journal of Logic and Computation 2017.
- M. Bendkowski, K. Grygiel, M. Zaionc. *On the likelihood of normalisation in combinatory logic*. Journal of Logic and Computation 2017.
- M. Bendkowski, K. Grygiel, P. Tarau. *Boltzmann Samplers for Closed Simply-Typed Lambda Terms*. PADL 2017.

Conclusions

Thank you for your attention!