

Enumerating λ -terms of fixed arity

Danièle GARDY

PRISM, Université Versailles St-Quentin en Yvelines

et

DMG, T.U. Wien, Autriche

daniele.gardy@prism.uvsq.fr

Journées ALÉA, Mars 2013

(avec O. Bodini, B. Gittenberger, A. Jacquot)

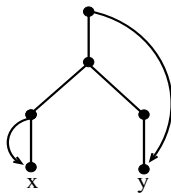
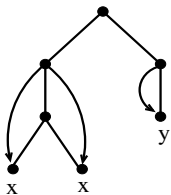
$$T ::= a \mid (T * T) \mid \lambda a.T$$

$(T * T)$: application

$\lambda a.T$: abstraction

$$(\lambda x.(x * x) * \lambda y.y)$$

$$\lambda y.(\lambda x.x * \lambda x.y)$$



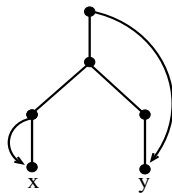
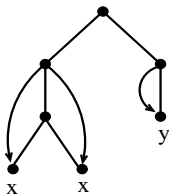
$$T ::= a \mid (T * T) \mid \lambda a.T$$

 $(T * T)$: application

 $\lambda a.T$: abstraction

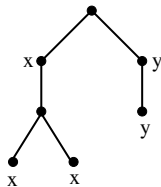
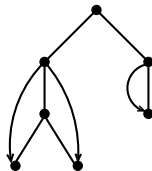
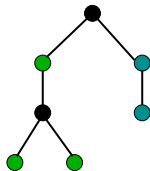
$$(\lambda x.(x * x) * \lambda y.y)$$

$$\lambda y.(\lambda x.x * \lambda x.y)$$

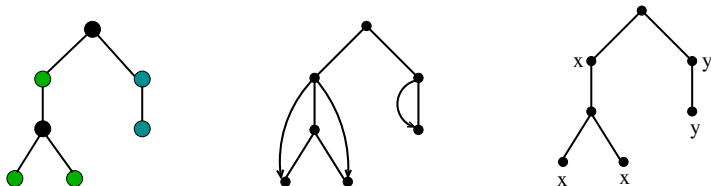


These λ -terms are **closed** (no free variable)

λ -terms as enriched Motzkin trees



λ -terms as enriched Motzkin trees

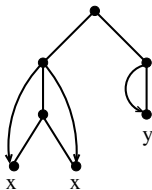


Labelling rules:

- Binary nodes are unlabelled
- Unary nodes get distinct labels (colors)
- Leaves get the label (color) of one of their unary ancestors

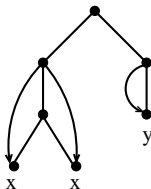
Free and bound variables

- Here all variables are bound

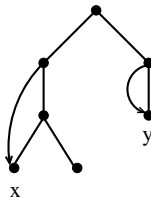


Free and bound variables

- Here all variables are bound



- Some variables may be free



- Recursive definition for λ -terms?

- \mathcal{L} : class of λ -terms with free variables
- \mathcal{N} atomic class of binary node
- \mathcal{U} atomic class of unary node
- \mathcal{F} atomic class of free leaf
- \mathcal{B} atomic class of bound leaf

$$\mathcal{L} = \mathcal{F} + \left(\mathcal{N} \times \mathcal{L}^2 \right) + \left(\mathcal{U} \times \text{subs}(\mathcal{F} \rightarrow \mathcal{F} + \mathcal{B}, \mathcal{L}) \right)$$

- Recursive definition for λ -terms?
 - \mathcal{L} : class of λ -terms with free variables
 - \mathcal{N} atomic class of binary node
 - \mathcal{U} atomic class of unary node
 - \mathcal{F} atomic class of free leaf
 - \mathcal{B} atomic class of bound leaf

$$\mathcal{L} = \mathcal{F} + \left(\mathcal{N} \times \mathcal{L}^2 \right) + \left(\mathcal{U} \times \text{subs}(\mathcal{F} \rightarrow \mathcal{F} + \mathcal{B}, \mathcal{L}) \right)$$

- Generating function

$$L(z, f) = fz + zL(z, f)^2 + zL(z, f + 1).$$

with $z \leftrightarrow$ size of the λ -term and $f \leftrightarrow$ free leaves
(size = total number of nodes)

- Generating function enumerating closed λ -terms (without free variables): $L(z, 0)$
- Generating function enumerating all λ -terms:
 $L(z, 1) = \frac{1}{z}L(z, 0) - L(z, 0)^2$
- $L(z, 0) = \frac{1}{2z} \left(1 - \sqrt{\Lambda(z)} \right)$ with $\Lambda(z)$ equal to

$$1 - 2z + 2z \sqrt{1 - 2z - 4z^2} + 2z \sqrt{\dots} \sqrt{1 - 2z - 4nz^2} + 2z \sqrt{\dots}$$

- $L(z, 0)$ has null radius of convergence \Rightarrow standard tools of analytic combinatorics **fail**

BCI(p)
enumeration

Danièle
GARDY

What can we do?

λ -terms and
enriched trees

λ -terms with
fixed arity

Equations on
the generating
functions

Asymptotics
for $BCI(p)$

Larger
classes of
 λ -terms

What can we do?

- Try to find a way to deal with null radius of convergence?

What can we do?

- Try to find a way to deal with null radius of convergence?
- *Ad hoc* methods?

$$\left(\frac{(4 - \epsilon)n}{\log n} \right)^{n(1 - 1/\log n)} \leq L_n \leq \left(\frac{(12 + \epsilon)n}{\log n} \right)^{n(1 - 1/3 \log n)}$$

[David et al. 00; here leaves have size 0]

What can we do?

- Try to find a way to deal with null radius of convergence?
- *Ad hoc* methods?

$$\left(\frac{(4 - \epsilon)n}{\log n} \right)^{n(1-1/\log n)} \leq L_n \leq \left(\frac{(12 + \epsilon)n}{\log n} \right)^{n(1-1/3 \log n)}$$

[David et al. 00; here leaves have size 0]

- Consider sub-classes of terms?
 - Restrict the *total* number of abstractions
 - Restrict the number of abstractions *in a path from the root towards a leaf*: bounded unary height [Analco'11, Aléa'11]
 - Restrict the number of pointers from an abstraction to a leaf

λ -terms with fixed arity

Two classes of closed λ -terms:

- $BCI(p)$: each abstraction binds *exactly* p variables
- $BCK(p)$: each abstraction binds *at most* p variables

Consider first $p = 1$, then generalize...

$BCI(1)$ and $BCK(1)$

Class of λ -terms when each abstraction binds exactly one variable: $BCI(1)$

- Size is always $3n + 2$
- Bijection with triangular pointed diagrams enumerated according to the number of edges (Vidal)
- Equate number of terms with coefficients of $z^3 \frac{d}{dz} \ln \left(e^{z^3/3} \odot e^{z^2/2} \right)$
- Get asymptotic equivalent $BCI(1)_{3n+2} \sim C\sqrt{n} \left(\frac{6n}{e}\right)^n$

Extension of approach for $BCI(1)$ gives

$$BCK(1)_n \sim \frac{C_1}{n^{1/6}} \left(\frac{2n}{e}\right)^{n/3} e^{\frac{(2n)^{2/3}}{2} - \frac{(2n)^{1/3}}{6}}$$

How can we generalize this approach to $BCI(p)$? to $BCK(p)$? to unrestricted λ -terms?

How can we generalize this approach to $BCI(p)$? to $BCK(p)$? to unrestricted λ -terms?

- Generalize the differential equation on the bivariate generating function for $BCI(1)$ terms with free leaves?

$$T(z, f) = zf + zT^2(z, f) + z\frac{\partial T}{\partial f}(z, f)$$

How can we generalize this approach to $BCI(p)$? to $BCK(p)$? to unrestricted λ -terms?

- Generalize the differential equation on the bivariate generating function for $BCI(1)$ terms with free leaves?

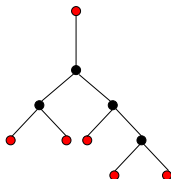
$$T(z, f) = zf + zT^2(z, f) + z\frac{\partial T}{\partial f}(z, f)$$

- Go back to the recursive definition of (unrestricted) λ -terms and adapt?

$$\mathcal{L} = \mathcal{F} + (\mathcal{N} \times \mathcal{L}^2) + (\mathcal{U} \times \text{subs}(\mathcal{F} \rightarrow \mathcal{F} + \mathcal{B}, \mathcal{L}))$$

Recursive construction of a $BCI(p)$ term

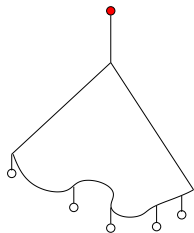
- A $BCI(p)$ term with j abstraction nodes has size $(2p + 1)j - 1$
- Smallest terms: $j = 1$. There are C_{p-1} such terms



- All other terms are obtained either by taking two terms as left and right children of a binary root, or by taking a term and adding an abstraction node at root

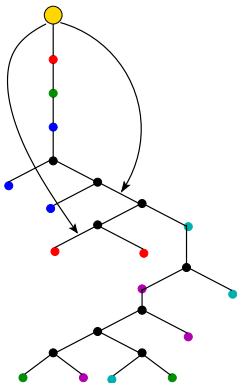
Adding an abstraction as root of a $BCI(p)$ term

- In the term whose root is the new abstraction node, all other abstraction nodes already have p pointer to leaves
- The term below the root must have p free leaves...



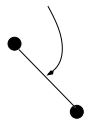
- ... but a $BCI(p)$ term is closed!

How do we get new, free leaves?

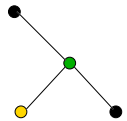
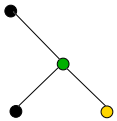


How do we get new, free leaves?

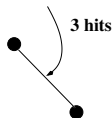
- If an edge is hit once...



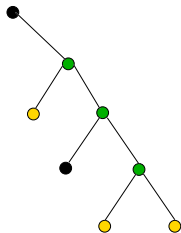
... we create a new leaf; there are 2 choices



- If an edge is hit i times....



... we create i new leaves and i new binary nodes.
There are $\binom{2i}{i}$ ways to do it (build a sequence of binary trees with a total of i leaves; each new tree goes either right or left)



The differential operator Δ_p

- p hits
- some edges can be hit repeatedly
- l different edges are hit

$$\alpha_{l,p} = \sum_{\sum_i s_i=l; \sum_i i s_i=p} \binom{l}{s_1! \dots s_p!} \prod_{m=1}^p \binom{2m}{m}^{s_m}$$

$$\Delta_p = \sum_{1 \leq l \leq p} \frac{\alpha_{l,p}}{l!} z^{l+2p+1} D^l$$

Generating function for $BCI(p)$

A $BCI(p)$ term either

- 1 is a Catalan tree with p leaves, topped by an abstraction node,
- 2 or has a binary node as root with two $BCI(p)$ children ,
- 3 or has abstraction node as root, pointing to p new free leaves

Univariate generating function for $BCI(p)$ is solution of a differential equation:

$$Y(z) = C_{p-1} z^{2p} + zY(z)^2 + \Delta_p Y(z)$$

What about $BCK(p)$?

Generating function $F_p(z) = Y\left(\frac{z}{1-z}\right)$, with Y satisfying

$$Y(z) = \sum_{l=1}^p C_{l-1} z^{2l} + zY(z)^2 + \left(\sum_{l=1}^p \Delta_p\right) Y(z)$$

Alternative form:

$$M(z, u) = \left(1 - z - \sqrt{(1-z)^2 - 4uz}\right) / (2z);$$

$$F_p(z) = z[u^p] \frac{M(z, u)}{1-u} \\ + zF_p(z)^2 + z[u^p] \frac{1}{1-u} F_p\left(\frac{z}{1-2zM(z, u)}\right)$$

And for unrestricted λ -terms?

Generating function for closed terms is $\Lambda(z) = \tilde{\Lambda}\left(\frac{z}{1-z}\right)$ with

$$\tilde{\Lambda}(z) = C(z) + z\tilde{\Lambda}(z) + z\tilde{\Lambda}\left(\frac{z}{1-2C(z)}\right) - z\tilde{\Lambda}(z)$$

and $C(z) = (1 - \sqrt{1 - 4z^2})/2$

Alternative form:

$$\Lambda(z) = zM(z, 1) + z\Lambda(z)^2 + z\Lambda\left(\frac{z}{1-2zM(z, 1)}\right)$$

Solving the differential equation for $BCI(p)$?

$$Y = C_{p-1}z^{2p} + zY^2 + \Delta_p Y$$

We cannot solve explicitly this differential equation, nor find asymptotics by singularity analysis (radius of convergence is null again)...

Solving the differential equation for $BCI(p)$?

$$Y = C_{p-1}z^{2p} + zY^2 + \Delta_p Y$$

We cannot solve explicitly this differential equation, nor find asymptotics by singularity analysis (radius of convergence is null again)...

... but we can do asymptotics for an *approximate* equation

$$Y = C_{p-1}z^{2p} + 2C_{p-1}zY + \Delta_p Y$$

with same asymptotic behaviour!

Asymptotic enumeration of $BCI(p)$

Theorem

For $p \geq 2$, the number of λ -terms of $BCI(p)$ of size $(2p+1)n-1$ is asymptotically

$$a_p B_p \beta_p^{n-1} n^{\frac{p(p-2)}{2p+1}} n^{np}$$

with $a_p = 1 + \mathcal{O}(1/(pe^p))$, $\beta_p = \frac{(4p+2)^p}{p!}$ and

$$B_p = \frac{C_{p-1}}{\prod_{1 \leq j \leq p} \Gamma\left(1 + \frac{2(p-j)-1}{2p+1}\right)}$$

$$\sim C_{p-1} (1.0844375\dots)^{(2p+1)/2} (1 + \mathcal{O}(1/p)).$$

Number λ_n of unrestricted λ -terms

Theorem

For all ε and for large n

$$c_1 \left(\frac{4n}{e \log n} \right)^{n/2} \frac{\sqrt{\log n}}{n} \leq \lambda_n \leq c_2 \left(\frac{9(1+\varepsilon)n}{e \log n} \right)^{n/2} \frac{(\log n)^{n/(2 \log n)}}{n^{3/2}}$$

[Proof: by counting a subclass, and a larger set]

Recurrence relations for λ_n

$$\lambda_n = M_{n-1} + \sum_{\ell+q=n-1} \lambda_\ell \lambda_q + \sum_{1 \leq \ell \leq n-1} \delta_{n,\ell} \lambda_\ell$$

D-finite recurrences for $\delta_{n,\ell}$:

$$\begin{aligned} & (n-\ell)(n+1-\ell)(n-2\ell-2)\delta_{n+2,\ell} \\ & - (n-\ell)(2n^2-6n\ell-5n+2\ell^2+3\ell+1)\delta_{n+1,\ell} \\ & - (n-1)(3n^2-2n\ell+n-\ell^2-9\ell-8)\delta_{n,\ell} \\ & + 20(n-1)\ell(\ell+1)\delta_{n,\ell+2} \\ & + 2(n-1)(5n-9\ell-12)\ell\delta_{n,\ell+1} = 0 \end{aligned}$$

and

$$\begin{aligned} & (n-\ell)(\ell-n-1)\delta_{n+2,\ell} + (n-\ell)(2n-\ell)\delta_{n+1,\ell} - \ell(n-1)\delta_{n+1,\ell+1} \\ & - 4\ell(n-1)\delta_{n,\ell+1} - (n-1)(3n-2\ell+1)\delta_{n,\ell} = 0. \end{aligned}$$

Summary: number of λ -terms for several classes

- q abstractions: $\frac{q^{\frac{q}{2}}}{\sqrt{q!} \sqrt{2\pi n^3}} (4q)^{\frac{n+1-q}{2}}$
- Bounded unary height h : (usually) $C \rho_h^{-n} n^{-3/2}$
- $BCI(1)_{3n+2} \sim C \sqrt{n} \left(\frac{6n}{e}\right)^n$
- $BCI(p)_{(2p+1)n-1} \sim a_p B_p \beta_p^{n-1} n^{\frac{p(p-2)}{2p+1}} n^{np}$
- Unrestricted terms:

$$c_1 \left(\frac{4n}{e \log n} \right)^{n/2} \frac{\sqrt{\log n}}{n} \leq \lambda_n \leq c_2 \left(\frac{9(1+\varepsilon)n}{e \log n} \right)^{n/2} \frac{(\log n)^{\frac{n}{2 \log n}}}{n^{3/2}}$$