Enumerating (restricted) λ-terms

Danièle GARDY

PRiSM, Université Versailles St-Quentin en Yvelines
and
D.M.G., T.U. Wien

In collaboration with O. Bodini, B. Gittenberger and A. Jacquot

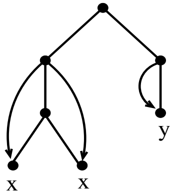
λ-terms and enriched (Motzkin) trees

Definition of λ-terms

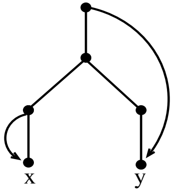
$$T ::= a \mid (T * T) \mid \lambda a.T$$

$(T * T)$: application $\lambda a.T$: abstraction

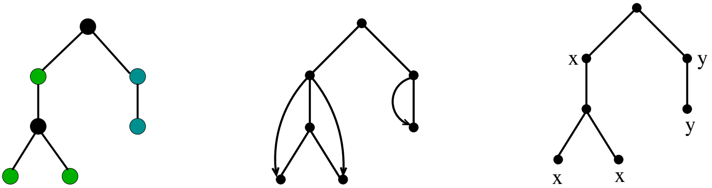
$(\lambda x.(x * x) * \lambda y.y)$



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



Enriched Motzkin trees

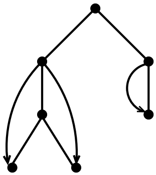


Labelling rules:

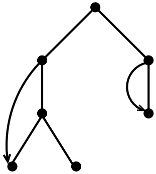
- Binary nodes are unlabelled
- Unary nodes get distinct labels (colors)
- Leaves get the label (color) of one of their unary ancestors

Free and bound variables

- Here all variables are bound: closed terms



- Some variables may be free



Enumeration?

- 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

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

- $L_{\ell,n}$ number of λ-terms of size n (total number of nodes) with ℓ free leaves
- Generating function $L(z, f) = \sum_{\ell,n} L_{\ell,n} f^\ell z^n$ satisfies a functional equation

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

Analytic combinatorics

- Generating function of a sequence a_n: $A(z) = \sum_n a_n z^n$
- $A(z)$ considered as a function of complex variable z: domain of analycity? radius of convergence ρ?
- Type and location of dominant singularity determine the asymptotic behaviour of the sequence a_n
- E.g., ρ algebraic of type $(1 - \frac{z}{\rho})^{\alpha}$ $(\alpha \notin N)$ gives

$$[z^n]A(z) \sim \frac{\rho^n \; n^{-\alpha-1}}{\Gamma(-\alpha)}$$

- Extensions to multivariate cases, asymptotic distributions

Enumeration???

- 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

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

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

What can we do?

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

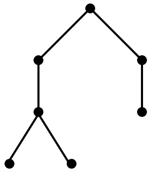
$$\left(\frac{(4-\epsilon)n}{\log n}\right)^{n(1-1/\log n)} \leq L_n \leq \left(\frac{(12+\epsilon)n}{\log n}\right)^{n(1-1/3\log n)}$$

[David et al. 10; here leaves have size 0]

- Consider sub-classes of terms?
  - Restrict the *total* number of abstractions [Bodini-G-Gittenberger'14]
  - Restrict the number of abstractions *in a path from the root towards a leaf:* bounded unary height [Bodini-G-Gittenberger'11, Bodini-G-Gittenberger'14]
  - Restrict the number of pointers from an abstraction to a leaf [Bodini-G-Jacquot'10; Bodini-G-Gittenberger-Jacquot'13, Bodini-Gittenberger'15]

Motzkin trees

Motzkin trees



$$\mathcal{M} = \mathcal{Z} + (\mathcal{U} \times \mathcal{M}) + (\mathcal{Z} \times \mathcal{M}^2)$$

$$M(z) = \frac{1}{2z}\left(1 - z - \sqrt{1 - 2z - 3z^2}\right)$$

Dominant singularity at $z = 1/3$ of square-root type

$$[z^n]M(z) \sim \frac{3^{n+\frac{1}{2}}}{2n\sqrt{\pi n}}$$

q unary nodes

$$\mathcal{M}_q = \mathcal{U} \times \mathcal{M}_{q-1} + \sum_{\ell=0}^{q} \mathcal{A} \times \mathcal{M}_\ell \times \mathcal{M}_{q-\ell}.$$

Recurrence equation on the generating functions

$$M_q(z) = \frac{zM_{q-1}(z) + z\sum_{1\leq\ell\leq q-1} M_\ell(z)\, M_{q-\ell}(z)}{1 - 2zM_0(z)}.$$

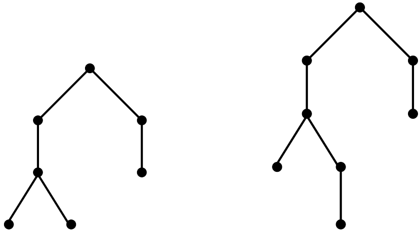
$\Rightarrow$ there exist polynomials P_q s.t.

$$M_q(z) = \frac{z^{q+1}\, P_q(z^2)}{(1-4z^2)^{q-\frac{1}{2}}},$$

Straightforward computations give

$$[z^n]M_q(z) \sim [z^n]\mathcal{M}_{\leq q} \sim \frac{\sqrt{2}\, P_q(1/4)}{\Gamma(q-\frac{1}{2})}\, 4^n\, n^{q-\frac{3}{2}}$$

Leaves at same unary height



- Tree on the left: all leaves have unary height 1
- Tree on the right: leaves have unary heights 1, 2 and 1

Leaves at same unary height

$$\mathcal{MH}_k = \mathcal{U} \times \mathcal{MH}_{k-1} + \mathcal{A} \times \mathcal{MH}_k^2$$

On generating functions

$$\mathcal{MH}_k = \frac{1}{2}\left(1 - \sqrt{1 - 2z + 2z\sqrt{1 - 2z + 2z\sqrt{\dots + 2z\sqrt{1 - 4z^2}}}}\right)$$

Two singularities

- $z = -\frac{1}{2}$ of type $(1+2z)^{\frac{1}{2}}$ (negligible)
- $z = \frac{1}{2}$ of type $(1+2z)^{\frac{1}{2^{k+1}}}$ (dominant, comes from the innermost radicand)

$$\Rightarrow [z^n]\mathcal{MH}_k(z) \sim \frac{2^{\frac{1}{2^{k+1}}}\, 2^n\, n^{-1-\frac{1}{2^{k+1}}}}{2^{k+1}\Gamma(1 - \frac{1}{2^{k+1}})}$$

Bounded unary height

Here leaves can have different unary height!

$$\mathcal{MH}_{\leq k} = \mathcal{Z} + \mathcal{U} \times \mathcal{MH}_{\leq k-1} + \mathcal{A} \times \mathcal{MH}^2_{\leq k}$$

Generating function

$$\mathcal{MH}_{\leq k} = \frac{1}{2}\left(1 - \sqrt{1 - 2z - 4z^2 + 2z\sqrt{1 - 2z - 4z^2 + 2z\sqrt{\ldots + 2z\sqrt{1-4z^2}}}}\right)$$

Dominant singularity ρ_k comes from outermost radicand, decreases towards $\frac{1}{3}$

$$\Rightarrow [z^n]\mathcal{MH}_{\leq k} \sim \frac{\sqrt{1+4\rho_k^2}}{4\,\rho_k^{n+1} n\sqrt{\pi n}}$$

λ-terms with bounded number of unary nodes

q unary nodes

$$\mathcal{S}_q = (\mathcal{U} \times \textit{subs}(\mathcal{F} \to \mathcal{F} + \mathcal{B}, \mathcal{S}_{q-1})) + \sum_{\ell=0}^{q} (\mathcal{A}, \mathcal{S}_\ell, \mathcal{S}_{q-\ell})$$

Generating function

$$S_q(z,f) = zS_{q-1}(z,f+1) + z\sum_{\ell=0}^{q} S_\ell(z,f)\ S_{q-\ell}(z,f).$$

G.F. for closed terms $S_q(z,0)$?

$$\begin{aligned} S_1(z,0) &= \frac{1}{2} - \frac{\sqrt{1-4z^2}}{2}; \\ S_2(z,0) &= \frac{z}{2}(1-2z^2) + \frac{2z^3}{\sqrt{1-4z^2}} - \frac{z\sqrt{1-8z^2}}{2\sqrt{1-4z^2}}; \end{aligned}$$

(no terms of size $n = q \ mod \ 2$)

q unary nodes

$$S_q(z,f) = -\frac{z^{q-1}\sigma_q(f)}{2\prod_{\ell=0}^{q-1}\sigma_\ell(f)} + R_q(z,\sigma_0(f),\dots,\sigma_{q-1}(f))$$

where

- $\sigma_q(f) = \sqrt{1-4(f+q)z^2}$
- R_q rational, denominator $\prod_{0\leq\ell<q}\sigma_\ell(f)^{\alpha_{\ell,q}}$
- $\alpha_{\ell,q} > 0$, either integer or $\frac{1}{2}+$ an integer

q unary nodes

$$S_q(z,f) = -\frac{z^{q-1}\sigma_q(f)}{2\prod_{\ell=0}^{q-1}\sigma_\ell(f)} + R_q(z,\sigma_0(f),\dots,\sigma_{q-1}(f))$$

where

- $\sigma_q(f) = \sqrt{1-4(f+q)z^2}$
- R_q rational, denominator $\prod_{0\le\ell<q}\sigma_\ell(f)^{\alpha_{\ell,q}}$
- $\alpha_{\ell,q} > 0$, either integer or $\frac{1}{2}+$ an integer

$$\Rightarrow S_q(z,0) = -\frac{z^{q-1}\sqrt{1-4qz^2}}{2\prod_{\ell=0}^{q-1}\sqrt{1-4\ell z^2}} + R_q(z,1,\sqrt{1-4z^2},\dots,\sqrt{1-4(q-1)z^2})$$

Dominant singularities at $\pm\frac{1}{2\sqrt{q}}$ of square-root type

$$\Rightarrow [z^n]S_q(z,0) \sim \frac{q^{\frac{q}{2}}}{\sqrt{q!}\,\sqrt{2\,\pi\,n^3}}\,(4q)^{\frac{n+1-q}{2}}$$

λ-terms of bounded unary height

The classes $\mathcal{P}^{(i,k)}$

k: maximal number of abstractions on a path from the root to a leaf

- $\mathcal{P}^{(0,k)}$: λ-terms with bound variables and unary height $\leq k$
- $\mathcal{P}^{(1,k)}$: λ-terms with bound variables, 1 kind of free variables, and unary height $\leq k - 1$
- ...
- $\mathcal{P}^{(i,k)}$: λ-terms with bound variables, i kinds of free variables, and unary height $\leq k - i$
- ...
- $\mathcal{P}^{(k,k)}$: λ-terms with bound variables, k kinds of free variables, and no unary node

The classes $\mathcal{P}^{(i,k)}$

Set up equations on generating functions $P^{(i,k)}$, solve, and take $H_{\leq k}(z) = P^{(0,k)}(z)$:

$$H_{\leq k} = \frac{1 - \sqrt{1 - 2z + 2z\sqrt{1 - 2z - 4z^2 + 2z\sqrt{\dots\sqrt{1 - 4kz^2}}}}}{2z}$$

We can start the asymptotic study of its coefficients!

The classes $\mathcal{P}^{(i,k)}$

Set up equations on generating functions $P^{(i,k)}$, solve, and take $H_{\leq k}(z) = P^{(0,k)}(z)$:

$$H_{\leq k} = \frac{1 - \sqrt{1 - 2z + 2z\sqrt{1 - 2z - 4z^2 + 2z\sqrt{\ldots\sqrt{1 - 4kz^2}}}}}{2z}$$

We can start the asymptotic study of its coefficients!

- $H_{\leq k}$ is algebraic and written with $k+1$ iterated radicands
- Its singularities are the values that cancel its radicands
- Which radicant has smallest root? (We rank radicands from the innermost to the outermost)

- $k = 1$

$$H_{\leq 1}(z) = \frac{1 - \sqrt{1 - 2z + 2z\sqrt{1 - 4z^2}}}{2z}$$

Dominant singularity: $\rho = \frac{1}{2}$ (cancels both radicands)

$$[z^n]H_{\leq 1}(z) \sim \frac{1}{4} \frac{2^{\frac{1}{4}} 2^n n^{-\frac{5}{4}}}{\Gamma(\frac{3}{4})}$$

- $k = 2$

$$H_{\leq 2}(z) = \frac{1 - \sqrt{1 - 2z + 2z\sqrt{1 - 2z - 4z^2 + 2z\sqrt{1 - 8z^2}}}}{2z}$$

Dominant singularity: $\rho = 0.3437999303$ (cancels the second innermost radicand)

$$[z^n]H_{\leq 2}(z) \sim \frac{C}{\Gamma(\frac{1}{2})} n^{-\frac{3}{2}} \rho^{-n}$$

Where is the dominant singularity when k grows?

Function	Radicand	Singularity
$H_{\leq 1}$	{1,2}	0.5
$H_{\leq 2}$	2	0.3438
$H_{\leq 3}$	2	0.2760
...	...	...
$H_{\leq 8}$	{2,3}	0.1667
$H_{\leq 9}$	3	0.1571
...	...	...
$SH_{\leq 134}$	3	0.0418
$H_{\leq 135}$	{3,4}	0.0417
$H_{\leq 136}$	4	0.0415
...	...	...

Sometimes, the same value cancels *two* consecutive radicands.

Define $u_k = u_{k-1}^2 + k$ for $k > 0$, $u_0 = 0$

- $u_1 = 1$, $u_2 = 3$, $u_3 = 12$, $u_4 = 148$, $u_5 = 21909$, ... Doubly exponential growth: $\lim_{k\to\infty} u_k^{1/2^k} \simeq \chi = 1.36660956...$
- Set $N_k = u_k^2 - u_{k-1}^2$: $N_1 = 1$, $N_2 = 8$, $N_3 = 135$, $N_4 = 21760$, $N_5 = 479982377$, ...

Theorem

i) *$\exists i, k = N_i$: radicands of ranks i and $(i+1)$ cancel for the same value, are both dominant. Dominant singularity $\rho_{N_i} = \frac{1}{2u_i}$ is algebraic of type $1/4$.*

$$[z^n]H_{\leq N_i} \sim C_i n^{-5/4} \rho_i^n$$

ii) *$k \in]N_i, N_{i+1}[$: dominant radicand has rank i. Dominant singularity ρ_k is algebraic of type $1/2$.*

$$[z^n]H_{\leq k} \sim C_k n^{-3/2} \rho_k^n$$

λ-terms of fixed arity

λ-terms of fixed arity

Two classes of closed λ-terms:

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

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

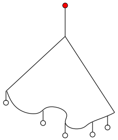
$BCI(1)$ and $BCK(1)$

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)
   - Asymptotic equivalent $BCI(1)_{3n+2} \sim C\sqrt{n}\left(\frac{6n}{e}\right)^n$
2. Adapt this to get

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

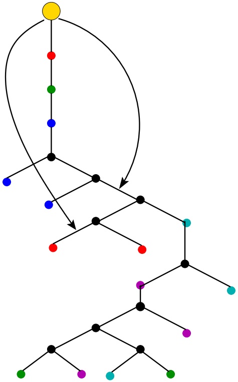
BCI(p)

- A *BCI*(p) term with j abstraction nodes has size $(2p+1)j-1$
- Smallest terms: $j=1$; *one* unary node above a binary tree
- Other terms:
  - binary root, two *BCI*(p) terms as left and right children
  - unary root, one child with p free leaves...



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

How do we get new, free leaves?



The differential operator Δ_p

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

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

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

Univariate generating function for $BCI(p)$ satisfies

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

Solving the differential equation for $BCI(p)$, $p \geq 2$?

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

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

Solving the differential equation for $BCI(p)$, $p \geq 2$?

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

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!

Theorem

Asymptotic number of λ-terms of size $(2p+1)n-1$:

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

(α_p and β_p are explicit)

Concluding remarks

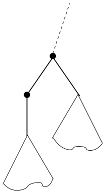
Varied asymptotic behaviours

1. Motzkin trees
    - Number of unary nodes = q: one radical, $C_q\, 4^n n^{q-\frac{3}{2}}$
    - Shared unary height of leaves = k: iterated radicals; *innermost* radical dominates; $C_k\, 2^n\, n^{-1-\frac{1}{2^{k+1}}}$
    - Bounded unary height = k: iterated radicals, *outermost* radical dominates; $C_k\, \rho_k^n\, n^{-\frac{3}{2}}$
2. λ-terms
    - Number of unary nodes = q: product of radicals; $C_q\, (4q)^{\frac{n+1-q}{2}}\ n^{-\frac{3}{2}}$
    - Bounded unary height = k: iterated radicals; dominant radical *fluctuates*
        - Standard case: $C_k n^{-\frac{3}{2}} \rho_k^n$
        - Special values: *two* dominant radicals; $C_k n^{-\frac{5}{4}} \rho_k^n$
    - Arity = p: $\alpha_p\, \beta_p^{n-1}\, n^{\frac{p(p-2)}{2p+1}}\ n^{np}$
    - Unrestricted terms:

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

Statistical properties

- Asymptotic enumeration of other classes? of unrestricted λ-terms?
- Number of nodes of various types?
- Unary/total height?
- Number of λ-terms in normal form?



Forbidden pattern

- ...

Work in progress

Thanks for your attention