

Some
statistics on
 λ -terms

Danièle
GARDY

Motzkin trees
and λ -terms

Motzkin trees

Bounded number of
unary nodes

Bounded unary
height

λ -terms with
bounded
number of
unary nodes

λ -terms of
bounded
unary height

The classes $\mathcal{P}^{(1,k)}$

Some values of k are
special

Asymptotics

What next?

Some statistics on λ -terms

Danièle GARDY

PRiSM, Université Versailles St-Quentin en Yvelines

and

D.M.G., T.U. Wien

Joint work with O. Bodini and B. Gittenberger

- 1 Motzkin trees and λ -terms
- 2 Motzkin trees
 - Bounded number of unary nodes
 - Bounded unary height
- 3 λ -terms with bounded number of unary nodes
- 4 λ -terms of bounded unary height
 - The classes $\mathcal{P}^{(i,k)}$
 - Some values of k are special
 - Asymptotics
- 5 What next?

Some statistics on λ -terms

Danièle GARDY

Motzkin trees and λ -terms

Motzkin trees

Bounded number of unary nodes

Bounded unary height

λ -terms with bounded number of unary nodes

λ -terms of bounded unary height

The classes $\mathcal{P}^{(1,k)}$

Some values of k are special

Asymptotics

What next?

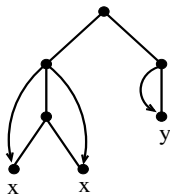
Motzkin trees and λ -terms

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

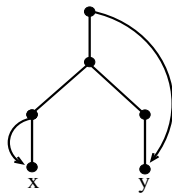
$(T * T)$: application

$\lambda a.T$: abstraction

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

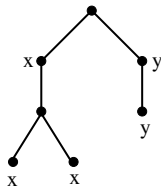
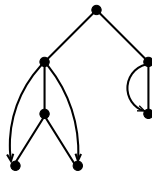
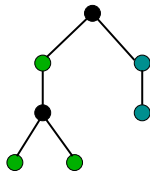


$$\lambda y.(\lambda x.x * \lambda x.y)$$



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

λ -terms as enriched Motzkin trees

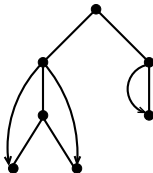


Labelling rules:

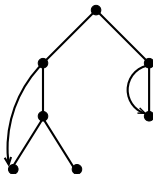
- Binary nodes are unlabelled
- Unary nodes get distinct labels (colors)
- Leaves get the label (color) of one of their unary ancestors

Free and bound variables in leaves

- 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

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

- 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

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 [this talk]
 - Restrict the number of abstractions *in a path from the root towards a leaf*: bounded unary height [Analo'11, this talk]
 - Restrict the number of pointers from an abstraction to a leaf [AofA11]

How do restricted λ -terms compare with Motzkin trees?

Some statistics on λ -terms

Danièle GARDY

Motzkin trees and λ -terms

Motzkin trees

Bounded number of unary nodes

Bounded unary height

λ -terms with bounded number of unary nodes

λ -terms of bounded unary height

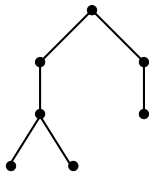
The classes $\mathcal{P}^{(1,k)}$

Some values of k are special

Asymptotics

What next?

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

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

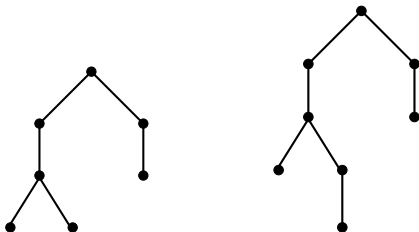
\Rightarrow there exist polynomials P_q s.t.

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

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 $(1 + 2z)^{\frac{1}{2^{2^{k+1}}}}$ (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!

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

Generating function

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

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

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

Some
statistics on
 λ -terms

Danièle
GARDY

Motzkin trees
and λ -terms

Motzkin trees

Bounded number of
unary nodes

Bounded unary
height

λ -terms with
bounded
number of
unary nodes

λ -terms of
bounded
unary height

The classes $\mathcal{P}^{(1,k)}$

Some values of k are
special

Asymptotics

What next?

λ -terms with bounded number of unary nodes

q unary nodes

$$S_q = (\mathcal{U} \times \text{subs}(\mathcal{F} \rightarrow \mathcal{F} + \mathcal{B}, S_{q-1})) + \sum_{\ell=0}^q (\mathcal{A}, S_\ell, S_{q-\ell})$$

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 $S_q(z, 0)$?

$$S_1(z, 0) = \frac{1}{2} - \frac{\sqrt{1-4z^2}}{2};$$

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

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 of R_q equal to $\prod_{0 \leq \ell < q} \sigma_\ell(f)^{\alpha_{\ell,q}}$
- $\alpha_{\ell,q} > 0$, either integer or $\frac{1}{2}$ + an integer

q unary nodes

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

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

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

(null if $n = q \pmod{2}$)

Some
statistics on
 λ -terms

Danièle
GARDY

Motzkin trees
and λ -terms

Motzkin trees

Bounded number of
unary nodes

Bounded unary
height

λ -terms with
bounded
number of
unary nodes

λ -terms of
bounded
unary height

The classes $\mathcal{P}^{(1,k)}$

Some values of k are
special

Asymptotics

What next?

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

- $i = k$

$$\mathcal{P}^{(k,k)} = k\mathcal{Z} + \mathcal{Z}\mathcal{P}^{(k,k)}{}^2$$

Generating function:

$$P^{(k,k)}(z) = kz + zP^{(k,k)}(z)^2$$

- $i < k$

$$\mathcal{P}^{(i,k)} = i\mathcal{Z} + \mathcal{Z}\mathcal{P}^{(i,k)}{}^2 + \mathcal{Z}\mathcal{P}^{(i+1,k)}$$

Generating function:

$$P^{(i,k)}(z) = iz + zP^{(i,k)}(z)^2 + zP^{(i+1,k)}(z)$$

Solve in $P^{(i,k)}$ 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!

Solve in $P^{(i,k)}$ 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!

- $H_{\leq k}$ is algebraic and written with $k + 1$ iterated radicands
- Its singularities are the values that cancel its radicands
- Which radicant has smallest root?

- $k = 1$

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

Dominant singularity: $\frac{1}{2}$ (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.

Values of k which give two dominant radicands?

- Define $(u_k)_{k \geq 0}$: $u_0 = 0$ and $u_k = u_{k-1}^2 + k$ for $k > 0$
- First values: $u_1 = 1$, $u_2 = 3$, $u_3 = 12$, $u_4 = 148$, $u_5 = 21909$, ...
- The sequence $(u_k)_{k \geq 0}$ is doubly exponential
- $\lim_{k \rightarrow \infty} u_k^{1/2^k} \simeq \chi = 1.36660956\dots$
- Define $N_k = u_k^2 - u_k + k = u_k^2 - u_{k-1}^2$.
 $N_1 = 1$, $N_2 = 8$, $N_3 = 135$, $N_4 = 21760$,
 $N_5 = 479982377$, ...

Theorem

Two asymptotic behaviours according to the value of k

- *For unary height N_i , the radicands with ranks i and $(i + 1)$ both cancel for the same value; both are dominant; the dominant singularity is algebraic of type $1/4$, and $[z^n]H_{\leq N_i} \sim C_i n^{-5/4} \rho_i^n$, with $\rho_i = 1/2u_i$.*
- *If $k \in]N_i, N_{i+1}[$, the dominant radicand of $H_{\leq k}(z)$ is the i -th radicand; the dominant singularity is algebraic of type $1/2$, and $[z^n]H_{\leq k} \sim C_k n^{-3/2} \rho_k^n$.*

(Radicands are ranked from the innermost to the outermost)

[Bodini-G-Gittenberger, Analco'11]

Observations

- The constants C_k become small very quickly.
Variation of C_k as a function of k ?
 - $[z^n]H_{\leq 1}(z) \sim 0.2426128012 \cdot \left(\frac{1}{n}\right)^{5/4} \cdot 2^n$
 - $[z^n]H_{\leq 8}(z) \sim 9.318885373 \cdot 10^{-5} \left(\frac{1}{n}\right)^{5/4} 6^n$
 - $[z^n]H_{\leq 135}(z) \sim 7.116999389 \cdot 10^{-158} \left(\frac{1}{n}\right)^{5/4} 24^n$
 - Doubly-exponential decay?
- We cannot hope to observe the asymptotic behaviour of $[z^n]H_{\leq k}$ from computations for “reasonable” n
- Yet we can randomly generate lambda-terms of bounded height and observe the behaviour of parameters...

The constant C_k for $k \in \{N_i\}$

$$[z^n]H_{\leq N_k} \sim \frac{D_k}{\Gamma(3/4) A_k} n^{-5/4} (2u_k)^n$$

Asymptotics for $k \rightarrow +\infty$?

$$D_k \sim \gamma u_{k-1} \quad \text{with} \quad \gamma = 1.2952778$$

$$A_k = \prod_{i=k}^{N_k-1} \sqrt{i + \sqrt{i-1 + \sqrt{i-2 + \sqrt{\dots + \sqrt{1}}}}}$$

$$\sim \frac{\varphi(N_k)}{\varphi(k)} \quad \text{with} \quad \varphi(k) = \frac{e^{\sqrt{k}}}{\sqrt{k}} \cdot \left(\frac{2k}{e}\right)^k$$

Some
statistics on
 λ -terms

Danièle
GARDY

Motzkin trees
and λ -terms

Motzkin trees

Bounded number of
unary nodes

Bounded unary
height

λ -terms with
bounded
number of
unary nodes

λ -terms of
bounded
unary height

The classes $\mathcal{P}^{(1,k)}$

Some values of k are
special

Asymptotics

What next?

What next?

Different types of asymptotic behaviour when enumerating restricted Motzkin trees and λ -terms

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; $C_k n^{-\frac{5}{4}} \rho_k^n$

Enumeration and properties of λ -terms

- How does a random restricted λ -term look like?
 - Number of nodes of each type?
 - Unary height? Total height?
 - Width? Profile?
- Enumerate (unrestricted) λ -terms according to their size
- Allow for free leaves; give a specific weight to the leaves
- Characterize the parameters of a random λ -term, their logical properties (strong normalizing, ...)

Number of λ -terms: $n \in [1, \dots, 198]$; unary height $k \in [1, \dots, 98]$

Motzkin trees and λ -terms

Motzkin trees

Bounded number of unary nodes

Bounded unary height

λ -terms with bounded number of unary nodes

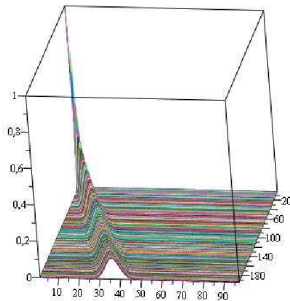
λ -terms of bounded unary height

The classes $\mathcal{P}^{(1,k)}$

Some values of k are special

Asymptotics

What next?

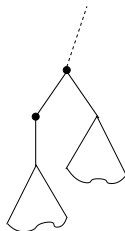


λ -terms and normal form

A term is in normal form



there is no pattern $\mathcal{A} \times (\mathcal{U}, \mathcal{T}) \times \mathcal{T}$



Forbidden pattern

Number of terms in normal form?

Normal form, bounded number of unary nodes

- Asymptotic number of closed, normal-form λ -terms with exactly q unary nodes and size n , $n \not\equiv q \pmod 2$

$$\frac{1}{2^q \sqrt{2\pi} n^3} \prod_{\ell=1}^q \frac{\sqrt{q} + \sqrt{\ell}}{\sqrt{\ell}} (4q)^{\frac{n+1-q}{2}}$$

- Asymptotic probability of closed, normal-form term with exactly q unary nodes and size n ($n \rightarrow +\infty$)

$$\pi_q = 2^{-q} \prod_{\ell=1}^q \left(1 + \sqrt{\frac{\ell}{q}} \right)$$

Normal form, bounded number of unary nodes

Asymptotic probability of closed normal-form term with exactly q unary nodes and size n for large q

$$\pi_q = \sqrt{2} \left(\frac{\sqrt{e}}{2} \right)^q (1 + o(1)) = \sqrt{2} \ 0.82436^q (1 + o(1)).$$

q	5	10	50	100	1000
Exact	0.496	0.193	$8.79 \cdot 10^{-5}$	$5.67 \cdot 10^{-9}$	$1.84 \cdot 10^{-84}$
Large q	0.538	0.205	$9.04 \cdot 10^{-5}$	$5.78 \cdot 10^{-9}$	$1.85 \cdot 10^{-84}$

Normal form, bounded number of unary nodes

Asymptotic probability of closed normal-form term with exactly q unary nodes and size n for large q

$$\pi_q = \sqrt{2} \left(\frac{\sqrt{e}}{2} \right)^q (1 + o(1)) = \sqrt{2} \ 0.82436^q (1 + o(1)).$$

q	5	10	50	100	1000
Exact	0.496	0.193	$8.79 \cdot 10^{-5}$	$5.67 \cdot 10^{-9}$	$1.84 \cdot 10^{-84}$
Large q	0.538	0.205	$9.04 \cdot 10^{-5}$	$5.78 \cdot 10^{-9}$	$1.85 \cdot 10^{-84}$

What about bounded unary height?

Some statistics on λ -terms

Danièle GARDY

Motzkin trees and λ -terms

Motzkin trees

Bounded number of unary nodes

Bounded unary height

λ -terms with bounded number of unary nodes

λ -terms of bounded unary height

The classes $\mathcal{P}^{(1,k)}$

Some values of k are special

Asymptotics

What next?

To be continued...

Thanks for your attention