

Automates et grammaires algébriques :
universalité de l'asymptotique
et
biodiversité des lois limites

Hanane Tafat Bouzid

University of Paris North,

03 février 2012

JCB 2012, Bordeaux

Plan

1 Introduction

2 Regular grammar

- Asymptotics
- Limits laws of motif
 - In strongly connected graph
 - In non-strongly connected graph
- Applications

3 Algebraic grammar

- asymptotics
 - In strongly connected graph
 - In non-strongly connected graph

4 Conclusion

A **strongly connected graph** is a directed graph that has a path from each vertex to every other vertex.

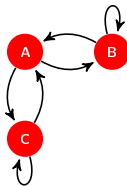
Language

Grammar

Dependency Graph

$$\mathcal{L} = \{(ab^*a \cup ba^*b)^*\}$$

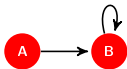
$$\begin{aligned} A &\rightarrow aB \mid bC \mid \epsilon \\ B &\rightarrow aA \mid bB \\ C &\rightarrow aC \mid bA \end{aligned}$$



strongly connected graph.

$$\mathcal{L} = \{a^n b^m a^m b^n, n \geq 0, m \geq 0\}$$

$$\begin{aligned} A &\rightarrow aAb \mid B \\ B &\rightarrow bBa \mid \epsilon \end{aligned}$$



non-strongly connected graph.

Plan

1 Introduction

2 Regular grammar

- Asymptotics
- Limits laws of motif
 - In strongly connected graph
 - In non-strongly connected graph
- Applications

3 Algebraic grammar

- asymptotics
 - In strongly connected graph
 - In non-strongly connected graph

4 Conclusion

Rational generating function

- $\mathcal{L} = \{(a \cdot b)^n, n \geq 0\}$

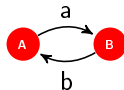
Right regular grammar

$$L \rightarrow a \cdot A \mid \epsilon$$
$$A \rightarrow b \cdot L$$

Left regular grammar

$$L \rightarrow A \cdot b \mid \epsilon$$
$$A \rightarrow L \cdot a$$

Automate



- A language is regular iff it is recognized by a finite deterministic automaton.

Theorem [Kleene, 1956]

The generating function of a regular language belongs to the set $\mathbb{Q}(z)$ of rational functions:

$$\mathbb{N}^{rat}[[z]] \subseteq \mathbb{Q}(z).$$

- $\mathbb{N}^{rat}[[z]]$ is also characterized as the smallest set of generating functions closed by addition, product, quasi-inverse.

Closure properties

$$\mathcal{A} = \{a^n, n \geq 0\}, \mathcal{B} = \{b^n, n \geq 0\}$$

- **Hadamard Product** $\mathcal{A} \odot \mathcal{B} := \{u.v, \quad |u| = |v|, \quad u \in \mathcal{A} \text{ and } v \in \mathcal{B}\}$

$$\mathcal{A} \odot \mathcal{B} = \{a^n \cdot b^n, n \geq 0\} \quad L \rightarrow a \cdot L \cdot b \mid \epsilon$$

- **Shuffle** $\mathcal{A} \sqcup \mathcal{B} := \bigcup_{w_1 \in \mathcal{A}, w_2 \in \mathcal{B}} (w_1 \sqcup w_2)$

$$\mathcal{A} \sqcup \mathcal{B} = \{(a \cup b)^n, n \geq 0\} \quad L \rightarrow a \cdot L \mid b \cdot L \mid \epsilon$$

- **Derivation** $\mathcal{A}' := \frac{\partial}{\partial \Sigma} \mathcal{A}$

$$\mathcal{A} = \{a^n, n \geq 0\} \quad \mathcal{A}' = \{a^n \cdot a^m, n \geq 0, m \geq 0\}$$

$$A \rightarrow a \cdot A \mid \epsilon \quad A' \rightarrow a \cdot A' \mid A$$
$$A \rightarrow a \cdot A \mid \epsilon$$

- **Complement, intersection, square root...**

Asymptotics of rational series (folklore)

- Let $F(z)$ be a rational function analytic at 0 then

$$\begin{aligned} [z^n]F(z) &= [z^n] \frac{P(z)}{Q(z)} = [z^n] \sum_{\rho \in \text{pole}(F)} \sum_{r=1}^{\text{order}(\rho)} \frac{C_{\rho,r}}{(z-\rho)^r} \\ &= \sum_{\rho \in \text{pole}(F)} \sum_{r=1}^{\text{order}(\rho)} C_{\rho,r} (-1)^r \binom{n+r-1}{r-1} \rho^{-n-r} \end{aligned}$$

- Let ρ be **the unique dominant singularity** and let r be its order, then

$$[z^n]F(z) \sim C_{\rho,r} (-1)^r \binom{n+r-1}{r-1} \rho^{-n-r}$$

- However, some difficult problems: $\exists f_n = 0$? (Pisot's problem), fluctuations of the positivity of the (f_n) 's (conjugate roots of same modulus, e.g. $F(z) = \frac{1}{1-6z+25z^2}$)

- Let $F(z, u)$ be a **bivariate rational generating function**

$$F(z, u) = \sum_{n \geq 0} f_n(u) z^n = \sum_{n \geq 0} \sum_{k=0}^n f_{n,k} u^k z^n,$$

where the number of words of length n with k “ b ” is $f_{n,k}$.

- X_n : a **random variable** counting the number of letters “ b ” in words of length n

$$\Pr(X_n = k) = \frac{f_{n,k}}{f_n(1)}.$$

What is the limit distribution of X_n ?

Limit law in strongly connected graph

- [Edward E. Bender, 1973]: Central and Local Limit Theorems Applied to Asymptotic Enumeration.
- [Guibas–Odlyzko, 1981]: String overlaps, pattern matching, and nontransitive games.
- [Régnier–Szpankowski, 1998]: On pattern frequency occurrences in a Markovian sequence.
- [Bertoni–Choffrut–Goldwurm–Lonati, 2006]: Asymptotics in a regular language of the number of words with a fixed number of occurrences.

Theorem [Nicodème–Salvy–Flajolet, 2002]

If the language \mathcal{L} is rational with strongly connected dependency graph, then X_n is asymptotically **Gaussian** with mean μ_n and variance σ_n^2 :

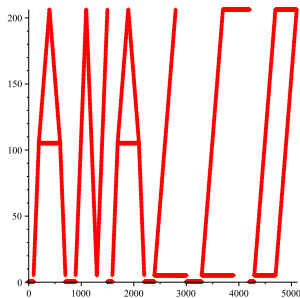
$$\Pr \left[\frac{X_n - \mu_n}{\sigma_n} \leq x \right] \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt.$$

Sketch of proof: Perron–Frobenius and Hwang quasi-power theorem. 

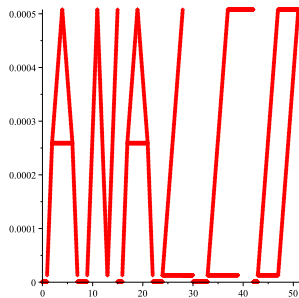
Non-Strongly Connected Graph

Theorem [BaBoPoTa, 2011]

For any continuous curve C , there exists a regular language in which X_n (number of "b" in words of length n) has a limit distribution which ϵ -near from C .



$$(x, y) \rightarrow \left(\frac{x}{n}, \frac{y}{f_n(1)} \right)$$



Theorem [BaBoPoTa, 2011]

For any polynomial $P(n, k)$ in $\mathbb{R}^{\geq 0}[n, k]$, there exists a power series $F(z, u) \in \mathbb{N}^{\text{rat}}[[z]]$ such that

$$\Pr(X_n = k) = \frac{[z^n][u^k]F(z, u)}{[z^n]F(z, 1)} = \frac{P(n, k)}{\sum_{i=0}^n P(n, i)}.$$

Sketch of proof

- $P(n, k)$ is a linear combination of $n^\alpha k^\beta$
- $n^\alpha k^\beta = [z^n][u^k] \sum_{n \geq 0} n^\alpha z^n \sum_{k \geq 0} k^\beta u^k.$



Lemma

The power series $\sum_{n \geq 0} n^i z^n$ belongs to $\mathbb{N}^{rat}[[z]]$ and more precisely

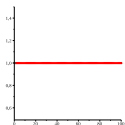
$$\sum_{n \geq 0} n^i z^n = \sum_{j=0}^i j! \left\{ \begin{matrix} i \\ j \end{matrix} \right\} \frac{z^j}{(1-z)^{j+1}} = \frac{\sum_{j=0}^i \left[\begin{matrix} i \\ j \end{matrix} \right] z^{j+1}}{(1-z)^{i+1}}.$$

Proof (by induction):

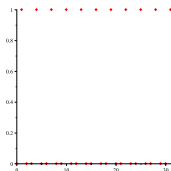
- apply $z\partial z$ to the equality (equivalent to pointing \ominus in the symbolic method),
- use the recurrence defining:

- ▶ the Stirling numbers of the second kind $\left\{ \begin{matrix} i+1 \\ j \end{matrix} \right\} = \left\{ \begin{matrix} i \\ j \end{matrix} \right\} + j \left\{ \begin{matrix} i \\ j-1 \end{matrix} \right\}$
- ▶ the Eulerian numbers $\left[\begin{matrix} i+1 \\ j \end{matrix} \right] = (j+1) \left[\begin{matrix} i \\ j \end{matrix} \right] + (i+1-j) \left[\begin{matrix} i \\ j-1 \end{matrix} \right].$ \square

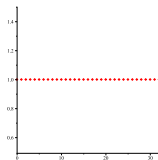
Examples



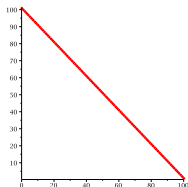
$$\frac{1}{(1-z)} \frac{1}{(1-uz)}$$



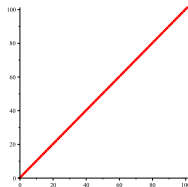
$$\frac{1}{(1-z)} \frac{1}{(1-u^3z)}$$



$$\frac{(u+u^2+u^3)}{(1-z)(1-u^3z)}$$



$$\frac{1}{(1-z)^2} \frac{1}{(1-uz)}$$



$$\frac{1}{(1-z)} \frac{1}{(1-uz)^2}$$

- $\mathcal{C} = \mathcal{A} \cup \mathcal{B}$

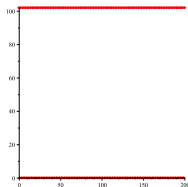
- $\Pr(X_n^{\mathcal{C}} = k) = \frac{[u^k][z^n]C(z,u)}{[z^n]C(z,1)} = \frac{[u^k][z^n](A(z,u)+B(z,u))}{[z^n](A(z,1)+B(z,1))}$

- ▶ If $(\rho_A < \rho_B)$ or $(\rho_A = \rho_B$ and $r_A < r_B)$ then

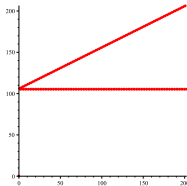
$$\Pr(X_n^{\mathcal{C}} = k) \sim \Pr(X_n^{\mathcal{A}} = k)$$

- ▶ If $\rho_A = \rho_B$ and $r_A = r_B$ then

$$\Pr(X_n^{\mathcal{C}} = k) = c_1 \Pr(X_n^{\mathcal{A}} = k) + c_2 \Pr(X_n^{\mathcal{B}} = k)$$

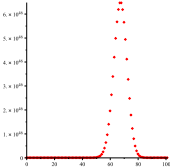
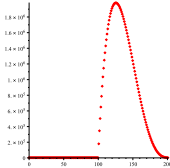
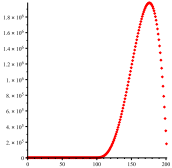
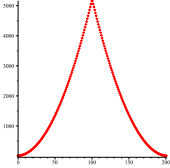
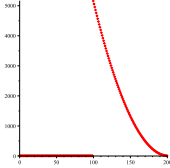
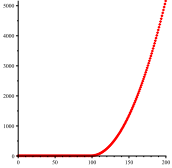


$$\frac{1}{(1-z)^2(1-u^2z)} + \frac{1}{(1-z)(1-u^2z)^2}$$



$$\frac{1}{(1-z)^2(1-u^2z)} + \frac{1}{(1-z)(1-u^2z)^2} + \frac{u}{(1-z)^2(1-u^2z)}$$

Examples



Boltzmann Sampling

- Boltzmann samplers allows the generation of uniform random decomposable combinatorial objects.
- [P. Duchon–P. Flajolet–G. Louchard–G. Shaeffer, 2004]: Boltzmann Samplers for the random generation of combinatorial structures.
- [O. Bodini–Y. Ponty, 2010]: Multidimensional Boltzmann Sampling of languages.
- The Boltzmann sampler returns an object γ of size n and parameter k with

$$\Pr_{x,u}(\gamma) = \frac{x^n u^k}{F(x, u)}.$$

Boltzmann Sampling

Theorem [BaBoPoTa, 2011] intrinsic efficiency for the size

For any fixed tolerance $\varepsilon > 0$, the Boltzmann sampler size generates, in average, in $O(n)$ independently in u_0 , an object of size N in the interval $[(1 - \varepsilon)n, (1 + \varepsilon)n]$.

Theorem [BaBoPoTa, 2011] intrinsic impossibility for the frequency

For any fixed tolerance $0 \leq \delta_u < 1$, the Boltzmann sampler size-frequency could be unable to generate an object with a frequency of u in an interval $[(\mu_u - \delta_u), (\mu_u + \delta_u)]$ where μ_u is the mean of u .

Plan

1 Introduction

2 Regular grammar

- Asymptotics
- Limits laws of motif
 - In strongly connected graph
 - In non-strongly connected graph
- Applications

3 Algebraic grammar

- asymptotics
 - In strongly connected graph
 - In non-strongly connected graph

4 Conclusion

Algebraic Branches

Combinatorial object	Algebraic grammar	Functional Equation	Branches
Binary trees	$T \rightarrow a \cdot T \cdot T a$	$z - T(z) + zT^2(z)$	$\left\{ \begin{array}{l} \frac{1 + \sqrt{1 - 4z^2}}{2z} \\ \frac{1 - \sqrt{1 - 4z^2}}{2z} \end{array} \right.$
Dyck words	$D \rightarrow a \cdot D \cdot b \cdot D \epsilon$	$1 - D(z) + z^2 D^2(z)$	$\left\{ \begin{array}{l} \frac{1 + \sqrt{1 - 4z^2}}{2z^2} \\ \frac{1 - \sqrt{1 - 4z^2}}{2z^2} \end{array} \right.$

Algebraic grammar with m non terminals:

- 1 Polynomial system of equations $Y_j = P_j(z, Y_1, \dots, Y_m)$,
- 2 P_j have non-negative coefficients,
- 3 $\vec{Y}^{(0)} = (0, \dots, 0)$, $\vec{Y}^{(h+1)} = P_i(\vec{Y}^{(h)})$, $\vec{Y} = \lim_{h \rightarrow \infty} \vec{Y}^{(h)}$

A **unique branch** analytic with non-negative Taylor coefficients around 0.

Asymptotics of algebraic grammars

Theorem of Newton-Puiseux

Let $f(z)$ the branch of an algebraic equation $P(z, f(z)) = 0$, near its singularities ρ , $f(z)$ admits a convergent fractional series expansion of the form: $f(z) = \sum_{k \geq k_0} c_k \cdot (z - \rho)^{\frac{k}{K}}$ with $k_0 \in \mathbb{Z}$ and $K \geq 1$.

asymptotics:

$$f_n \sim c_{k_0} \frac{1}{\Gamma(-\alpha)} \rho^{-n} n^{-\alpha-1} \text{ with } \alpha = \frac{k_0}{K}$$

What are the possible values of α ?

Strongly connected graph

Theorem of [Drmota–Lalley–Woods, 97–93–97]

Let $\vec{Y} = P_i(\vec{Y})$ a system satisfying:

- 1 All polynomials P_i have positive coefficients .
- 2 The system admits a fixed point \vec{Y} .
- 3 The dependency graph of the system is strongly connected.

Then:

- 1 all solutions behave like $Y_j(z) = C_{0j} - C_j \cdot \sqrt{1 - \frac{z}{\rho}} + O(1 - \frac{z}{\rho})$ in neighborhood $z = \rho$,
- 2 $[z^n]Y_j(z) \sim C_j \cdot \frac{1}{2\sqrt{\pi}} \cdot \rho^{-n} \cdot n^{-\frac{3}{2}}$ with C_{0j}, C_j, ρ algebraic numbers.

Example

$$\begin{cases} \mathcal{A} \rightarrow a \cdot \mathcal{A} | b \cdot \mathcal{B} \cdot \mathcal{B} | c \\ \mathcal{B} \rightarrow b \cdot \mathcal{A} | a \cdot \mathcal{B} \cdot \mathcal{B} | c \end{cases}$$

- $$\begin{cases} A(z) = zA(z) + zB(z)^2 + z \\ B(z) = zA(z) + zB(z)^2 + z \end{cases}$$

- Algebraic equation:** $zA(z)^2 + (z-1)A(z) + z = 0$

- Branches** of $A(z)$:

$$\frac{z-1 + \sqrt{-3z^2 - 2z + 1}}{2z}, \quad \frac{-z+1 + \sqrt{-3z^2 - 2z + 1}}{2z}$$

- The singularities:** $-1, 0, 1/3$

- Puiseux expansion:** $A(z) = \frac{5}{2} - \sqrt{3(1-3z)} + O(1-3z)$

non-strongly connected graph

general trees

$$\mathcal{T} = z \text{Seq}(\mathcal{T})$$

$$\begin{cases} \mathcal{T} \rightarrow a \cdot \mathcal{T} \cdot \mathcal{A} | a \\ \mathcal{A} \rightarrow \mathcal{T} \cdot \mathcal{A} | \epsilon \end{cases}$$



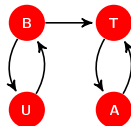
strongly connected.

super bicolore trees:

$$\mathcal{B} = \mathcal{T} [z, \mathcal{T}]$$

$$\mathcal{B} = 2z\mathcal{T} \text{Seq}(\mathcal{B})$$

$$\begin{cases} \mathcal{B} \rightarrow a \cdot \mathcal{T} \cdot \mathcal{U} | b \cdot \mathcal{T} \cdot \mathcal{U} \\ \mathcal{U} \rightarrow \mathcal{B} \cdot \mathcal{U} | \epsilon \\ \mathcal{T} \rightarrow a \cdot \mathcal{T} \cdot \mathcal{A} | a \\ \mathcal{A} \rightarrow \mathcal{T} \cdot \mathcal{A} | \epsilon \end{cases}$$



non-strongly connected.

Branch: $B(z) = \frac{1}{2} - \frac{1}{2} \sqrt{1 - 4z + 4z \sqrt{1 - 4z}}$

Puiseux expansion: $B(z) = \frac{1}{2} - \frac{1}{2} (1 - 4z)^{1/4} + O((1 - 4z)^{3/4})$

- The behavior asymptotics of **bicolored trees** \mathcal{T} is $1/2$
- **Super-bicolored trees**: $\mathcal{B} = \mathcal{T}[(z + z).\mathcal{T}]$
 - ▶ schema of **composition critique**,
 - ▶ $\frac{1}{2^2}$ is the behavior asymptotics.
- **k -Super-bicolored trees**: $\mathcal{B}_k = \mathcal{T}[(z + z).\mathcal{B}_{k-1}]$
 - ▶ schema of **k - composition critique**,
 - ▶ $\frac{1}{2^{k+1}}$ is the behavior asymptotics.

Using **Maple** (**Algolib:Combstruct**, **gdev**) we obtain: for k compositions de \mathcal{B} $\frac{1}{2^{k+1}}$ behavior.

Theorem

Suppose that $\vec{Y} = P(z, \vec{Y})$ is a positive polynomial system such that there is a unique solution $(Y_1(z), \dots, Y_m(z))$ that is analytic at $z = 0$.

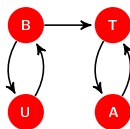
What is more, if $\frac{\partial^2 P_j}{\partial^2 Y_j} \neq 0$ then for every j there exists an integer $k_j \geq 1$ such that locally

$$Y_j(z) = \sum_{n \geq 0} c_n \cdot (z - \rho)^{n/2^{k_j}}$$

super bicolore trees:

$$\mathcal{B} = 2z\mathcal{T}\text{Seq}(\mathcal{B})$$

$$\left\{ \begin{array}{l} \mathcal{B} \rightarrow a \cdot \mathcal{T} \cdot \mathcal{U} \mid b \cdot \mathcal{T} \cdot \mathcal{U} \\ \mathcal{U} \rightarrow \mathcal{B} \cdot \mathcal{U} \mid \epsilon \\ \mathcal{T} \rightarrow a \cdot \mathcal{T} \cdot \mathcal{A} \mid a \\ \mathcal{A} \rightarrow \mathcal{T} \cdot \mathcal{A} \mid \epsilon \end{array} \right.$$



non-strongly connected.

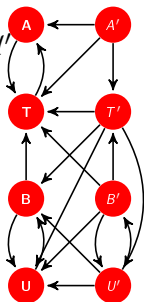
the derivative

- The generating function of Super-bicolored trees:

$$B(z) = \frac{1}{2} - \frac{1}{2}(1 - 4z)^{1/4} + O((1 - 4z)^{3/4})$$

$$\frac{\partial}{\partial z} B(z) = \frac{1}{2}(1 - 4z)^{-3/4} + O((1 - 4z)^{-1/4})$$

$$\left\{ \begin{array}{l} B' \rightarrow T \cdot U | a \cdot T' \cdot U | a \cdot T \cdot U' | b \cdot T' \cdot U | b \cdot T \cdot U' \\ B \rightarrow a \cdot T \cdot U | b \cdot T \cdot U \\ T' \rightarrow T \mathcal{A} | a \cdot T' \cdot \mathcal{A} | a \cdot T \cdot \mathcal{A}' | \epsilon \\ T \rightarrow a \cdot T \cdot \mathcal{A} | a \\ U' \rightarrow B' \cdot U | B \cdot U' | \epsilon \\ U \rightarrow B \cdot U | \epsilon \\ \mathcal{A}' \rightarrow T' \cdot \mathcal{A} | T \cdot \mathcal{A}' | \epsilon \\ \mathcal{A} \rightarrow T \cdot \mathcal{A} | \epsilon \end{array} \right.$$



non-strongly connected.

The dependency graph still non-strongly connected

Theorem

Suppose that $\vec{Y} = P(z, \vec{Y})$ is a positive polynomial system such that there is a unique solution $(Y_1(z), \dots, Y_m(z))$ that is analytic at $z = 0$.

What is more, if for all (i, j) $\frac{\partial P_j}{\partial Y_i} \neq 0$ and $\frac{\partial^2 P_j}{\partial Y_i \partial Y_h} \neq 0$ then for every j there exists an integer $k_j \geq 1$ such that locally

$$Y_j(z) = \sum_{n \geq 0} c_n \cdot (z - \rho)^{n/2^{k_j}}$$

or

$$Y_j(z) = \sum_{n \geq 0} c_n \cdot (z - \rho)^{n \frac{1}{2^{k_j}} - 1}$$

$$\left\{ \begin{array}{l} B' \rightarrow T \cdot U | a \cdot T' \cdot U | a \cdot T \cdot U' | b \cdot T' \cdot U | b \cdot T \cdot U' \\ B \rightarrow a \cdot T \cdot U | b \cdot T \cdot U \\ \vdots \end{array} \right.$$

Theorem

Suppose that $\vec{Y} = P(z, \vec{Y})$ is a positive polynomial system of entire functions such that there is a unique solution $(Y_1(z), \dots, Y_m(z))$ that is analytic at $z = 0$.

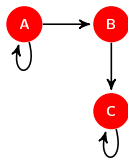
$$Y_j(z) = \sum_{n \geq 0} c_n \cdot (z - \rho)^{n/2^{k_j}}$$

or

$$Y_j(z) = \sum_{n \geq 0} c_n \cdot (z - \rho)^{n \frac{1}{2^{k_j}} - m}$$

Example

$$\begin{cases} A \rightarrow a \cdot A \cdot A \cdot A \mid b \cdot A \\ B \rightarrow a \mid a \cdot BC \mid b \cdot BC \\ C \rightarrow a \mid b \cdot CC \end{cases}$$



$$\begin{cases} A(z) = \frac{z^4(1-4z^2)^{-3/2}}{1-z} \\ B(z) = z(1-4z^2)^{-1/2} \\ C(z) = \frac{1-(1-4z^2)^{1/2}}{2z} \end{cases}$$

$$\begin{cases} A(z) = \frac{\sqrt{2}}{32}(1-2z)^{-3/2} + \dots \\ B(z) = \frac{\sqrt{2}}{4}(1-2z)^{-1/2} + \dots \\ C(z) = 1 - \sqrt{2}(1-2z)^{1/2} + \dots \end{cases}$$

$$[z^n]A(z) \sim \frac{\sqrt{2}}{56\sqrt{\pi}} 2^n n^{1/2}$$

Plan

1 Introduction

2 Regular grammar

- Asymptotics
- Limits laws of motif
 - In strongly connected graph
 - In non-strongly connected graph
- Applications

3 Algebraic grammar

- asymptotics
 - In strongly connected graph
 - In non-strongly connected graph

4 Conclusion

Conclusion

- We gave distributions which can be reached as a pattern occurrence distribution (+ application: random generation).
- Work in progress: the asymptotic behavior and the possible limit laws in non strongly connected context free-grammars?
- Implementation (in SageMath):
 - ▶ Input: (language, pattern),
 - ▶ Output: (limit law, multivariate Boltzmann sampler).