Asymptotique des grammaires régulières et algébriques.

Hanane Tafat

Équipe : CALIN
Laboratoire d'Informatique de Paris Nord

Encadrée par : Cyril Banderier.

20 juillet 2010

Plan

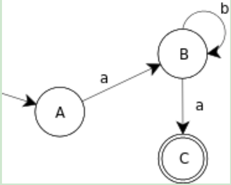
Plan

Définitions

- **Une grammaire régulière** : Grammaire dont les règles de productions sont de la forme : $A \rightarrow a.B|B.b|\epsilon$.

- **Un automate d'états fini** : Un quintuplet $(\Sigma, E, E \times \Sigma \rightarrow E, D, F)$.

Exemple



- La notion de grammaire régulière est équivalente à celle d'automate d'états fini.

Théorème (Kleene, 1952)

La série génératrice d'un langage régulier est rationnelle.

$$A(z) = \sum_{w \in \mathcal{L}} z^{|w|} = \sum_{n \in \mathbb{N}} a_n z^n = \frac{P(z)}{Q(z)}.$$

Preuve :

Grammaire	Série génératrice
ϵ	1
$a \in \Sigma$	z
$\mathbb{A} \cup \mathbb{B}$	$A(z) + B(z)$
$\mathbb{A}.\mathbb{B}$	$A(z) \times B(z)$
$\mathbb{A}^*$	$\frac{1}{1-A(z)}$

Théorème

Théorème (Kleene, 1952)

La série génératrice d'un langage régulier est rationnelle.

$$A(z) = \sum_{w \in \mathcal{L}} z^{|w|} = \sum_{n \in \mathbb{N}} a_n z^n = \frac{P(z)}{Q(z)}.$$

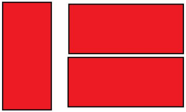
Preuve :

Grammaire	Série génératrice
ϵ	1
$a \in \Sigma$	z
$\mathbb{A} \cup \mathbb{B}$	$A(z) + B(z)$
$\mathbb{A}.\mathbb{B}$	$A(z) \times B(z)$
$\mathbb{A}^*$	$\frac{1}{1-A(z)}$

$A(z) = \frac{P(z)}{Q(z)} = \sum_{\rho,r} \frac{C}{(z-\rho)^r} \Rightarrow a_n \sim C.\rho^{-n-r}.\frac{n^{r-1}}{(r-1)!}.$

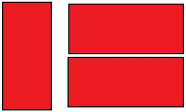
Exemple 1 : pavage

Quels est le nombre de manière de paver un rectangle de hauteur n et de largeur 2 par des dominos ?



Exemple 1 : pavage

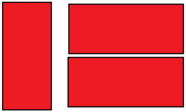
Quels est le nombre de manière de paver un rectangle de hauteur n et de largeur 2 par des dominos ?



$\mathcal{P}_0 = 0,\ \mathcal{P}_1 = 1$ et $\mathcal{P}_n = \mathcal{P}_{n-1} + \mathcal{P}_{n-2}$ (Suite de Fibonacci)
$\mathcal{P} = (z + z^2)^* \Rightarrow P(z) = \frac{1}{1-(z+z^2)}$

Exemple 1 : pavage

Quels est le nombre de manière de paver un rectangle de hauteur n et de largeur 2 par des dominos ?

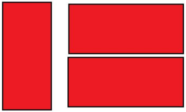


$\mathcal{P}_0 = 0,\ \mathcal{P}_1 = 1$ et $\mathcal{P}_n = \mathcal{P}_{n-1} + \mathcal{P}_{n-2}$ (Suite de Fibonacci)
$\mathcal{P} = (z + z^2)^* \Rightarrow P(z) = \frac{1}{1-(z+z^2)}$

$P(z) = \sum_{n\geq 0} a_n * z^n = 1 + z + 2 * z^2 + 3 * z^3 + 5 * z^4 + 8 * z^5 + O(z^6)$

Exemple 1 : pavage

Quels est le nombre de manière de paver un rectangle de hauteur n et de largeur 2 par des dominos ?



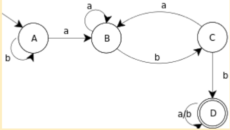
$\mathcal{P}_0 = 0,\ \mathcal{P}_1 = 1$ `et` $\mathcal{P}_n = \mathcal{P}_{n-1} + \mathcal{P}_{n-2}$ `(Suite de Fibonacci)`
$\mathcal{P} = (z + z^2)^* \Rightarrow P(z) = \frac{1}{1-(z+z^2)}$

$P(z) = \sum_{n\geq 0} a_n * z^n = 1 + z + 2 * z^2 + 3 * z^3 + 5 * z^4 + 8 * z^5 + O(z^6)$

$a_n = \frac{1}{\sqrt{5}} * \left[\left(\frac{1+\sqrt{5}}{2} \right)^{n+1} - \left(\frac{1-\sqrt{5}}{2} \right)^{n+1} \right]$ `(Formule de Binet).`

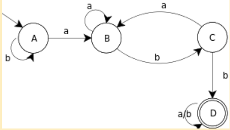
Exemple 2 : motif dans un langage

Quels est le nombre de mots de taille n ayant le motif $a.b.b$?



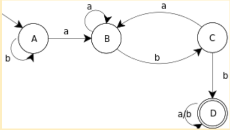
Exemple 2 : motif dans un langage

Quels est le nombre de mots de taille n ayant le motif $a.b.b$?



Exemple 2 : motif dans un langage

Quels est le nombre de mots de taille n ayant le motif $a.b.b$?

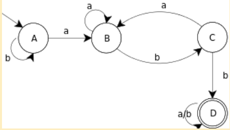


Le système d'équation linéaire lui correspondant est :

$$
\begin{aligned}
\mathcal{L}_0 &= a.\mathcal{L}_1 + b.\mathcal{L}_0 \\
\mathcal{L}_1 &= a.\mathcal{L}_1 + b.\mathcal{L}_2 \\
\mathcal{L}_2 &= a.\mathcal{L}_1 + b.\mathcal{L}_3 \\
\mathcal{L}_3 &= a.\mathcal{L}_3 + b.\mathcal{L}_3 + \epsilon
\end{aligned}
$$

Exemple 2 : motif dans un langage

Quels est le nombre de mots de taille n ayant le motif $a.b.b$?



Le système d'équation linéaire lui correspondant est :

$$\begin{aligned}
\mathcal{L}_0 &= a.\mathcal{L}_1 + b.\mathcal{L}_0 \\
\mathcal{L}_1 &= a.\mathcal{L}_1 + b.\mathcal{L}_2 \\
\mathcal{L}_2 &= a.\mathcal{L}_1 + b.\mathcal{L}_3 \\
\mathcal{L}_3 &= a.\mathcal{L}_3 + b.\mathcal{L}_3 + \epsilon
\end{aligned}$$

$L(z) = \frac{z^3}{(1-z)(1-2z)(1-z-z^2)} = \frac{1}{1-2z} - \frac{2+z}{1-z-z^2} + \frac{1}{1-z}$
$l_n = 2^n - \mathcal{P}_{n+3} + 1$

Modèle de Schelling

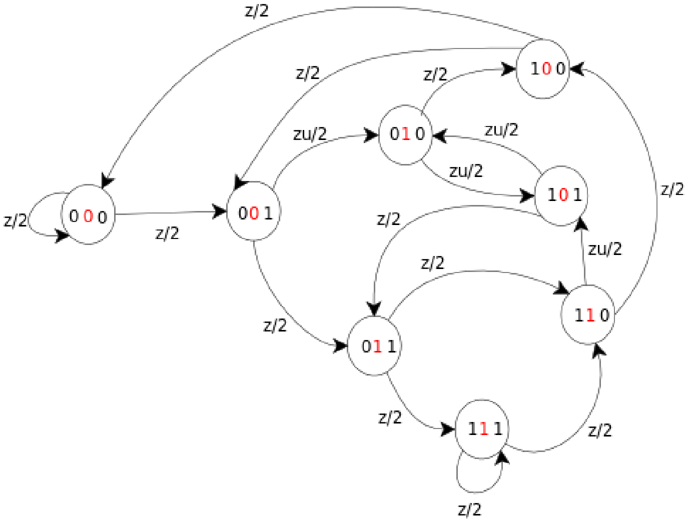
- [T.Schelling 71] propose un modèle dynamique de ségrégation dans le journal of Mathematical Sociology.
- Étant donnés
  - Une population de deux types $(0, 1)$,
  - Un voisinage de taille V,
  - Un individu est dit non satisfait s'il est minoritaire dans son voisinage.
  - Un individu non satisfait se déplace vers l'endroit le plus proche qui le satisfait.
- on a :
  - Stabilitée lorsque le nombre des non satisfaits est constant.
  - La majorité des individus sont satisfaits.
  - Alternances des groupes de même type.
  - Les individus minoritaires sont les plus ségrégationnistes.

Modèle de Schelling

- [T.Schelling 71] propose un modèle dynamique de ségrégation dans le journal of Mathematical Sociology.
- Étant donnés
  - Une population de deux types $(0, 1)$,
  - Un voisinage de taille V,
  - Un individu est dit non satisfait s'il est minoritaire dans son voisinage.
  - Un individu non satisfait se déplace vers l'endroit le plus proche qui le satisfait.
- on a :
  - Stabilitée lorsque le nombre des non satisfaits est constant.
  - La majorité des individus sont satisfaits.
  - Alternances des groupes de même type.
  - Les individus minoritaires sont les plus ségrégationnistes.

Question : Quel est le nombre moyen μ_n d'insatisfaits dans une population de taille n pour un voisinage V donné ?

Voisinage égal à 3



Méthodes appliquées

Pour un V fixé, formule explicite pour :
$F(z,u) = \sum_{n\geq 0} f_n(u) z^n = \sum_{n\geq 0} \sum_{k\geq 0} f_{n,k} z^n u^k$,

$$\mu_n = \frac{\sum_{k\geq 0} k.f_{n,k}}{f_n(1)} = \frac{f_n'(1)}{f_n(1)} = \frac{[z^n]F'(z,u)|_{u=1}}{[z^n]F(z,1)}.$$

Méthodes appliquées

Pour un V fixé, formule explicite pour :
$F(z,u) = \sum_{n\geq 0} f_n(u) z^n = \sum_{n\geq 0} \sum_{k\geq 0} f_{n,k} z^n u^k$,

$$\mu_n = \frac{\sum_{k\geq 0} k.f_{n,k}}{f_n(1)} = \frac{f_n'(1)}{f_n(1)} = \frac{[z^n]F'(z,u)|_{u=1}}{[z^n]F(z,1)}.$$

μ_n peut être obtenu en

- résolvant $[F_i(z,u)]$, $F(z,u) = \sum_i F_i(z,u)$ et $F(z,u)'|_{u=1}$.
- résolvant $[F_i(z,u), F_i'(z,u)]$, $F'(z,u) = \sum_i F_i'(z,u)$ et mettre $u=1$.
- résolvant $[F_i(z,1), F_i'(z,1)]$, $F'(z,1) = \sum_i F_i'(z,1)$.

Résultats obtenus

V	T_1	M_1	T_2	M_2	T_3	M_3
3	0.02	31415152	0.06	4455632	0.04	3276200
5	0.17	66883448	1.27	11925368	0.21	7600784
7	8.87	20443488	880.81	231168672	2.77	20181392

Résultats obtenus

V	T_1	M_1	T_2	M_2	T_3	M_3
3	0.02	31415152	0.06	4455632	0.04	3276200
5	0.17	66883448	1.27	11925368	0.21	7600784
7	8.87	20443488	880.81	231168672	2.77	20181392

les valeurs de μ_n obtenues sont :

Résultats obtenus

V	T_1	M_1	T_2	M_2	T_3	M_3
3	0.02	31415152	0.06	4455632	0.04	3276200
5	0.17	66883448	1.27	11925368	0.21	7600784
7	8.87	20443488	880.81	231168672	2.77	20181392

les valeurs de μ_n obtenues sont :

V	Nombre d'états	μ_n
3	2^3	$\frac{n}{4}$
5	2^5	$\frac{5n}{16}$
7	2^7	$\frac{11n}{32}$
9	2^9	$\frac{93n}{256}$
11	2^{11}	$\frac{193n}{512}$

Résultats obtenus

V	T_1	M_1	T_2	M_2	T_3	M_3
3	0.02	31415152	0.06	4455632	0.04	3276200
5	0.17	66883448	1.27	11925368	0.21	7600784
7	8.87	20443488	880.81	231168672	2.77	20181392

les valeurs de μ_n obtenues sont :

V	Nombre d'états	μ_n
3	2^3	$\frac{n}{4}$
5	2^5	$\frac{5n}{16}$
7	2^7	$\frac{11n}{32}$
9	2^9	$\frac{93n}{256}$
11	2^{11}	$\frac{193n}{512}$

Peut-on trouver une formule générale pour V quelconque ?

Plan

Grammaire algébrique

- **Une grammaire algébrique** : Une grammaire dont les règles de productions sont de la forme : $V \to w$.
  Exemple : La grammaire des mots de Dyck $S \to a.S.b.S|\epsilon$.
- À toute grammaire algébrique est associée une équation algébrique.
- La concaténation de deux langages algébriques est un langage algébrique.
- L'union de deux langages algébriques est un langage algébrique.

Grammaire algébrique

- Sans perte de généralité, on peut autoriser l'opérateur $Seq(a) := \epsilon + a + a.a + a.a.a + \ldots$ dans les grammaires algébriques. Comment représenter $A \rightarrow Seq(a)$ sous forme de grammaire algébrique ?

- Sans perte de généralité, on peut autoriser l'opérateur $Seq(a) := \epsilon + a + a.a + a.a.a + \ldots$ dans les grammaires algébriques. Comment représenter $A \rightarrow Seq(a)$ sous forme de grammaire algébrique ?
- La grammaire est :

$$A \rightarrow A.B|\epsilon$$
$$B \rightarrow a$$

Grammaire algébrique

- Sans perte de généralité, on peut autoriser l'opérateur $Seq(a) := \epsilon + a + a.a + a.a.a + \ldots$ dans les grammaires algébriques. Comment représenter $A \rightarrow Seq(a)$ sous forme de grammaire algébrique ?
- La grammaire est :

$$A \rightarrow A.B|\epsilon$$
$$B \rightarrow a$$

Grammaire	dérivation de la grammaire
$Seq(a)_{=0}$	$A \rightarrow \epsilon$
$Seq(a)_{=1}$	$A \rightarrow A.B \rightarrow \epsilon.a$
$Seq(a)_{=2}$	$A \rightarrow A.B \rightarrow A.B.B \rightarrow A.B.a \rightarrow \epsilon.a.a$

Graphe de dépendance

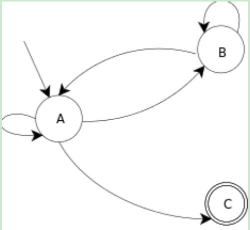
- Un système $\vec{Y} = \phi(\vec{Y})$ peut être représenté par un **graphe de dépendance**.

Exemple

Soit la grammaire :

$$A \to A.B|C,$$
$$B \to a.B|b.A,$$
$$C \to a.C|a$$

Le graphe de dépendance correspondant est :



Universalité de l'asymptotique en $n^{-\frac{3}{2}}$

Théorème Drmota–Lalley–Woods

Soit $\vec{Y} = \phi(\vec{Y})$ un système.
Si

1. Tous les polynômes ϕ_j ont des coefficients positifs.
2. Solution du système : $\vec{Y}^{(0)} = (0 \ldots 0)^t$, $\vec{Y}^{(h+1)} = \phi(\vec{Y})^h$, $\vec{Y} = \lim_{h \to \infty} \vec{Y}^{(h)}$.
3. le graphe de dépendance du système est fortement connexe.

alors :

1. toutes les solutions se comportent comme $Y_j = C_{0_j} - C_j.\sqrt{1 - \frac{z}{\rho}} + O(1 - \frac{z}{\rho})$ au voisinage de $z = \rho$,
2. $[z^n]Y_j(z) \sim C_j.\frac{1}{2\sqrt{\pi}}.\rho^{-n}.n^{-\frac{3}{2}}$ avec C_{0_j}, C_j, ρ des nombres algébriques.

Au-delà de Drmota–Lalley–Woods

Quand les conditions du théorème ne sont pas vérifiées, quels sont les autres comportements possibles ?

A-t-on universalité d'un comportement en n^α ?

On va montrer différents α possibles, via différentes constructions combinatoires.

Analyse de singularité

Si a la singularité dominante de $A(z)$ et b la singularité dominante de $B(z)$ alors la singularité dominante de :

- $C(z) = A(z) + B(z)$ vaut $c = \min(a, b)$
- $C(z) = A(z) \times B(z)$ vaut $c = \min(a, b)$
- schéma de composition $C(z) = A(B(z))$ vaut b ou z_a avec $B(z_a) = a$ selon les cas :
  - Schéma critique : $b = z_a$.
  - Schéma sur-critique : $b > z_a$.
  - Schéma sous-critique : $b < z_a$.

Comportement asymptotique

Théorème de Newton-Puiseux

Soit $f(z)$ branche d'une fonction algébrique $P(z, f(z)) = 0$, au voisinage de sa singularité ρ, $f(z)$ admet un développement en série fractionnelle convergente de la forme : $f(z) = \sum_{k \geq k_0} c_k (z - \rho)^{\frac{k}{K}}$ avec $k_0 \in \mathbb{Z}$ et $K \geq 1$.

Comportement asymptotique

Théorème de Newton-Puiseux

Soit $f(z)$ branche d'une fonction algébrique $P(z, f(z)) = 0$, au voisinage de sa singularité ρ, $f(z)$ admet un développement en série fractionnelle convergente de la forme : $f(z) = \sum_{k \geq k_0} c_k (z-\rho)^{\frac{k}{K}}$ avec $k_0 \in \mathbb{Z}$ et $K \geq 1$.

Pour $k > 1$, si $A(z) = \sum_n a_n z^{n/K}$ et $B(z) = \sum_n b_n z^{n/K}$ alors :

- $A(z) + B(z) = \sum_n (a_n + b_n) z^{n/K}$.
- $A(z) \times B(z) = \sum_n c_n z^{n/K}$ avec $c_n = \sum_{1<=p<=n} a_p \times b_{n-p}$.

Schéma de composition

- À l'aide de la méthode symbolique, on a :
  - $\mathbb{T} = z.Seq(\mathbb{T})$ = les arbres.
  - $\mathbb{B} = (z + z).Seq(\mathbb{B})$ = les arbres bicolores.
- Substitution $\mathbb{H} = \mathbb{T}[z.\mathbb{T}]$: les "super-arbres" (composition critique).
- À l'aide du package gfun et combstruct de Maple, on récupère les séries génératrices et leur comportement asymptotique : cf session Maple

Dérivée (ou "pointage")

- $(a_{n+1}z^{n+1})' = (n+1)a_{n+1}z^n$.
- donc dériver un mot w de taille $n+1$ revient à :
  1. Copier le mot $n+1$ fois,
  2. Pour chaque copie éliminer une lettre.
- On obtient donc $n+1$ copies toutes différentes de taille n.

Exemple

1. Mots de Dyck : $S \rightarrow a.S.b.S|\epsilon$
2. Dyck derivée :

$$\begin{cases} S \rightarrow a.S.b.A|a.A.b.S|A.b.A|a.A.A \\ A \rightarrow a.A.b.A|\epsilon \end{cases}$$

Théorème

- Pour avoir un comportement en $1/2^k$, il suffit de faire k composition de $\mathbb{T}$.
- Pour avoir un comportement en $(1/2^k) - 1$, il suffit de dériver $\mathbb{T}$.

Théorème

- Pour avoir un comportement en $1/2^k$, il suffit de faire k composition de $\mathbb{T}$.
- Pour avoir un comportement en $(1/2^k) - 1$, il suffit de dériver $\mathbb{T}$.

Les autres comportements s'expliquent par somme et produit de ces exposants.

Plan

Perspectives

1. Généralisation de la formule du nombre moyen d'insatisfaits pour n'importe quel voisinage.
2. librairie Maple (puis Sage) faisant l'asymptotique de n'importe quelle grammaire algébrique.

Lois limites

- $\mathcal{L} = (a+b)^*.(a+c)^*$,
- $L(z,u) = \sum l_{n,k} u^k z^n$,
- $l_{n,k} = \Pr(X_n = k)$.

Qu'obtient-on lorsque $n \to \infty$?

- $\mathcal{L}$ est rationel,
- Graphe connexe $\Rightarrow$ loi gausienne.

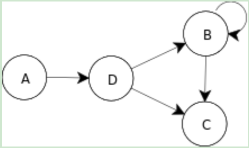
Dans notre cas, le graphe est-il connexe ?

$\mathcal{L} = (a+b)^*.(a+c)^*$

Soit la grammaire :

$A \to D|b.B|c.C|\epsilon$
$D \to B|C$
$B \to a.B|b.B|\epsilon$
$C \to a.C|c.C|\epsilon$

Le graphe de dépendance correspondant est :



- $L(z) = Seq(z_a + z_b).Seq(z_a + z_c) = \frac{1}{1-z_a-z_b} \times \frac{1}{1-z_a-z_c}$
- Pour $z_a = z_c = 1$ on a $L(z) = \frac{1}{1-z(1+w_b)} \times \frac{1}{1-2z}$
- Selon les différentes valeures de w_b on a : cf session Maple

Plus généralement, que peut-on dire dans le cas des grammaires régulières et algébriques avec un graphe non fortement connexe ? ? ? ?

Merci de votre attention

Questions ?