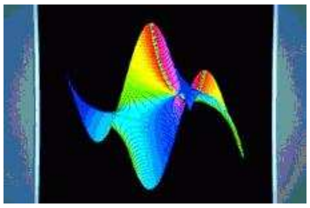
Colloquium Junior
INRIA Rocquencourt
11 mai 1999

Combinatoire analytique pour juniors

Cyril.Banderier@inria.fr

http://algo.inria.fr/banderier/
Projet Algorithmes



Projet Algorithmes

étude de l'aléa dans les structures combinatoires
dégager des lois universelles
automatisation de ces études via le calcul formel

- combinatoire du génome
  recherche de motifs (ADN)
- modélisation des réseaux de communication
  méthodes probabilistes (pertes/cache)
- théorie des nombres
  tests de primalité, récurrences
- calcul formel
  fonctions spéciales ; packages Maple :
  Groebner, Gfun, Mgfun, Combstruct
- combinatoire analytique
  complexité en moyenne d'algorithmes

Précurseurs de la combinatoire analytique

• Euler (1707-1783)
séries génératrices (triangulations 1753)
• Cauchy (1789-1857)
analyse complexe (résidus 1825)

• Ramanujan (1887-1920)
asympotique, séries à convergence rapide
• Knuth (né en 1938)
analyse d'algorithmes (fin des 60's)

Du discret au continu... et réciproquement !

Combinatoire analytique $\Longrightarrow$ dénombrement, établissement des comportements génériques, de structures finies/récursives (physique, chimie, mathématiques, biologie, informatique)

discret (structure combinatoire)

$\Downarrow$

continu (série de variable complexe)

$\Downarrow$

analyse ("Majorer, minorer, approcher", éq. fonctionnelle, forme close)

$\Downarrow$

asymptotique (singularités, cols...)

$\Downarrow$

loi limite

$\Downarrow$

discret

Il existe un dictionnaire qui traduit une structure combinatoire en "série génératrice".

Le dictionnaire

principe : manipuler structure combinatoire $\mathcal{A}$
= manipuler série génératrice $A(z) = \sum_n a_n z^n$

union	$\mathcal{A} \dot{\cup} \mathcal{B}$	$\mapsto$	$A(z) + B(z)$
produit	$\mathcal{A} \times \mathcal{B}$	$\mapsto$	$A(z) \times B(z)$
suite	$\mathrm{Seq}\,\mathcal{A}$	$\mapsto$	$\frac{1}{1-A(z)}$
ensemble	$\mathrm{Set}\,\mathcal{A}$	$\mapsto$	$\exp(A(z))$ (étiqueté)
cycle	$\mathrm{Cyc}\,\mathcal{A}$	$\mapsto$	$\ln \frac{1}{1-A(z)}$ (étiqueté)
substitution	$\mathcal{B} \circ \mathcal{A}$	$\mapsto$	$B(A(z))$
pointage	$\Theta\mathcal{A}$	$\mapsto$	$z\frac{d}{dz}A(z)$

Arbres binaires et nombres de Catalan

arbres binaires = mots bien parenthésés = marche sur $\mathbb{N}$ à pas ± 1.

b_n nombre d'arbres binaires avec n nœuds (internes) sont comptés par $B(z)$, la série génératrice

$$B(z) = \sum_{n \geq 0} b_n z^n$$

lecture directe de la série génératrice à partir de la relation $\mathcal{B} = \mathcal{F} + \mathcal{B} \times \mathcal{N} \times \mathcal{B}$.

$$B(z) = 1 + zB^2(z)$$

$$B(z) = \frac{1 - \sqrt{1 - 4z}}{2z} = 1 + z + 2z^2 + 5z^3 + 14z^4 + \ldots$$

On obtient les nb de Catalan : $\frac{\binom{2n}{n}}{n+1} \sim 4^n$ $\square$

Quelques résultats classiques

- Graphes

  - étude de diverses propriétés (composantes connexes...) des graphes aléatoires (Erdös-Rényi)
  - dénombrement des cartes (Tutte)

- Arbres
  hauteur moyenne h_n de familles d'arbres

  - arbres planaires généraux : $h_n \sim \sqrt{\pi n}$ (de Bruijn/Knuth/Rice 72)
  - arbres binaires et d'autres classes : $h_n \sim C\sqrt{\pi n}$ (Flajolet/Odlyzko 82)

- Permutations
  Ex : études d'algorithmes de tris

  - insertion sort, bubble sort : $O(n^2)$ comparaisons et échanges

- selection sort : $O(n^2)$ comp. et n éch.
- shellsort : $n^{3/2}$ comparaisons
- quicksort : $2n \ln n$ comparaisons

Preuve de la complexité en $2n \ln n$ de quicksort :

$$c_n = n + 1 + \frac{1}{n} \sum_{1 \leq j \leq n} (c_{j-1} + c_{n-j})$$

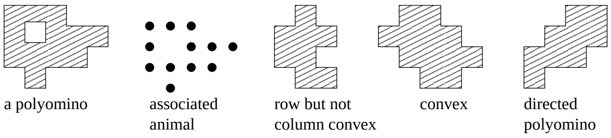
$$\Rightarrow C(z) = \sum_{n \geq 0} c_n z^n = \frac{2}{(1-z)^2} \ln \frac{1}{1-z}$$

$$\Rightarrow c_n = 2(n+1)(H_{n+1} - 1)$$

$$\text{or } H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \sim \ln n \quad \square$$

Combinatoire et physique

- molécules d'alcools (Pólya 1937), diverses structures cristallines en chimie
- le modèle d'Ising, les tas de sable en physique statistique



Modèle courant en percolation est celui des polyominos : le dénombrement (et même l'asymptotique) sur un réseau carré est toujours non résolu.

Promesse de mariage fructueux :
combinatoire analytique / physique statistique
$\rightarrow$ phénomènes de transitions de phases, seuils.

Wanted : $érie génératrice bivariée de marches aléatoires

But : généralisation de l'exemple marches à pas (+1,-1) à d'autres types de pas.

mots engendrés grammaire infinie ("régulière")
≈ traces automate infini
≈ puissances itérées matrice infinie
≈ arbre de possibilités
≈ marches (aléatoires)

Série génératrice d'une "marche aléatoire sur les entiers" :

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

n code la longueur $\qquad$ k code la hauteur

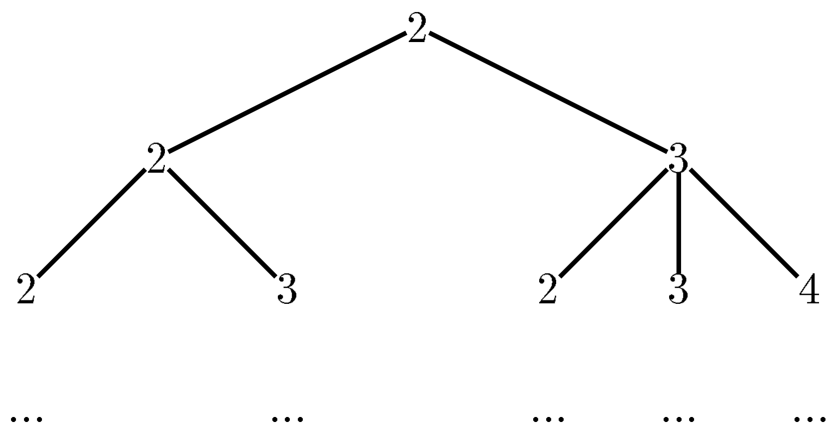
•$f_{n,k}$ est le nombre de chemins de longueur n qui vont de l'état initial à l'état k.
•$F(z,1)$ compte tous les chemins.
•$F(z,0)$ compte les excursions (hauteur finale=0).

Les *ECO*systèmes

énumération d'objets combinatoires
génération aléatoire

confer RR-3661 de C. Banderier, M. Bousquet-Mélou, A. Denise, P. Flajolet, D. Gardy et D. Gouyou-Beauchamps :
"On generating functions on generating trees".

Arbre de possibilités : un nœud étiqueté k a k fils.



règle de récriture ci-dessus :
$(k) \mapsto (2)(3)\ldots(k+1)$

Combien de nœuds à hauteur n ?
$=$ combien de marches de longueur n ?

Résultats Explicites

application de la méthode du noyau à :

$$F(z,u)(1-zP(u)) = 1 - z\sum_{i=0}^{c-1} p_i(u)\partial_i F(z,0)$$

Théorème 1. *La s.g. d'une marche sur $\mathbb{N}$ à pas bornés est algébrique.*

Théorème 2. *Une marche sur $\mathbb{N}$ à pas arrière "continus" et à pas avant bornés a une série génératrice bivariée algébrique*

Preuve : méthode du noyau

Exemples :

Catalan $[(2),(k) \to (2)(3)\ldots(k)(k+1)]$

arbres t-aires $[(t),(k) \to (t)\ldots(k)\ldots(k+t-1)]$

Motzkin $[(1),(k) \to (1)(2)\ldots(k-1)(k+1)]$

$$F(z,1) = \frac{1-z-\sqrt{1-2z-3z^2}}{2z^2} = \{1,1,2,4,9,21\}$$

Schröder $[(3),(k) \to (3)\ldots(k)(k+1)^2]$

$$F(z,1) = \frac{1-3z-\sqrt{1-6z+z^2}}{4z^2} = \{1,3,11,45,197\}$$

Marches en dimension 2

pas du cavalier codés par

$$P(u,v) := uv^2 + uv^{-2} + u^2v + u^2v^{-1}$$

$$F(z,u,v) = 1 + z\big(P(u,v)F(z,u,v)$$
$$-uv^{-1}\partial_v F(z,u,0) - (u^2v^{-1} + uv^{-2})F(z,u,0)\big)$$

"nombre de chemins allant de $(0,0)$ à $(k,0)$" :

$$a_k = [u^k v^0]F(1,u,v).$$

$$F(1,u,0) = \frac{1+2u+\Delta}{4u} - \frac{\sqrt{2+2\Delta-8u^2+4u\Delta}}{4u}$$

$\Delta := \sqrt{1-4u+4u^2-4u^4}$ $\qquad a_k \sim (1+\sqrt{3})^k \approx 2.73^k \quad \square$

autre approche : Labelle et Yeh

Fractions continuées

naturel (pour sauts -1,0,1) de par la traduction structure combinatoire / analyse :

$$\mathcal{E} = \mathrm{Seq}(\nearrow \mathcal{E} \searrow)$$

$$\Leftrightarrow E(z) = \frac{1}{1 - z^2 E(z)} = \cfrac{1}{1 - \cfrac{z^2}{1 - \ddots}}.$$

Liens polynômes orthogonaux.
Arrangements $[2, (k) \to (k)(k+1)^{k-1}]$
Laguerre $[2, (k) \to (k+1)^{k-1}(k+2)]$
Hermite $[2, (k) \to (k-1)^{k-1}(k+1)]$
(involutions en marquant les points fixes)
Stirling $[2, (k) \to (k)^{k-1}(k+1)]$ (nb de partages de $\{1, \dots n\}$ en k ss-ens. non vides)
Bessel $[0, (k) \to (k-1)(k)^k(k+1)]$
(permutations sans croisements)

Nombre moyen d'effleurements

$$F(z,u,t) = \sum_{n,k,j} f_{nkj} z^n u^k t^j$$

$u \leftrightarrow$ hauteur,$z \leftrightarrow$ longueur, $t \leftrightarrow$ nombre de touchers de $y = 0$.

Exemple avec les mots de Dyck $(P(u) = u + \frac{1}{u})$:

$$F(z,u,t) = \frac{u^2(1-t)(u_1 - z) + zu_1}{((1-t)u_1 + zt)(zu^2 - u + z)}$$

Le nombre moyen d'effleurements est

$$\mu_{2k} = \frac{[z^{2k}]\partial_t F(z,0,1)}{[z^{2k}]F(z,0,1)} = \frac{[z^{2k}]\frac{u_1(z)(u_1(z)-z)}{z^2}}{[z^{2k}]\frac{u_1(z)}{z}}$$

$$= \frac{\frac{\binom{2(k+1)}{k+1}}{k+2} - \frac{\binom{2k}{k}}{k+1}}{\frac{\binom{2k}{k}}{k+1}} = 3\frac{k}{k+2} = 3\sum_{p\geq 0} \frac{(-1)^p 2^p}{k^p}$$

$$= 3 - \frac{6}{k} + \frac{12}{k^2} + O(\frac{1}{k^3}).$$

$\square$

Asymptotique des excursions

Combien y a-t-il, pour n “grand”, d’excursions de longueur n ?

Théorème 3. *Asymptotique en* $\frac{A}{\rho^n\sqrt{\pi n^3}}$. *$A$ et ρ (le rayon de convergence) sont algébriques.*

Preuve (par analyse de singularité).
$F = \prod U_i$. Or singularité(s) de U_i dicte(nt) la croissance des coeff de F.

U_i compliqués mais asymptotique simple : comportement en $\sqrt{}$ au voisinage de la singularité.

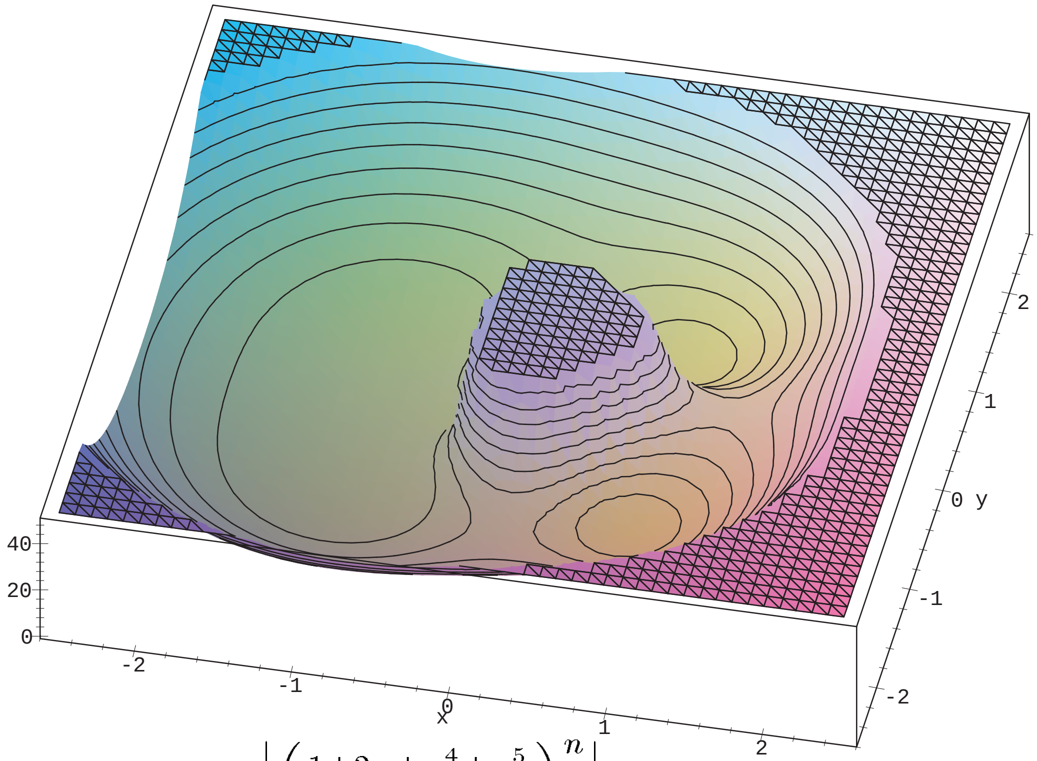
$$\text{i.e.} \quad U - \tau = d_1\sqrt{\rho - z} + d_2(\rho - z) + \dots$$

Or $[z^n]\sqrt{1-z} = \frac{-1}{2\sqrt{\pi n^3}}(1 + O(\frac{1}{n}))$ $\square$

Autre preuve. Méthode du col.

Morale : coeff=intégrale, c'est ça la formule intégrale de Cauchy : $f_n = \frac{1}{2i\pi}\int_\gamma \frac{F(z)}{z^{n+1}}dz$

Idée : jouer avec le contour γ et obtenir l'asymptotique quand n est grand. $\square$



graphe de $\left|\left(\frac{1+2u+u^4+u^5}{u^3}\right)^n\right|$, *i.e.* marche (-3,-2,-2,+1,+2).

Quand n est grand, tout est concentré autour du point col.

Conclusion

Combinatoire analytique $\Longrightarrow$ dénombrement, établissement des comportements génériques, de structures finies/récursives (physique, chimie, mathématiques, biologie, informatique)

discret (structure combinatoire)

$\Downarrow$

continu (série de variable complexe, via dictionnaire)

$\Downarrow$

analyse ("Majorer, minorer, approcher", éq. fonctionnelle, forme close)

$\Downarrow$

asymptotique (singularités, cols...)

$\Downarrow$

loi limite

$\Downarrow$

discret