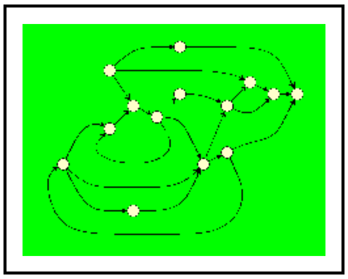
Torino, Oct. 1998 

ANALYTIC COMBINATORICS — SINGULAR COMBINATORICS

Philippe Flajolet, INRIA Rocquencourt

http://www-rocq.inria.fr/algo/flajolet/index.html

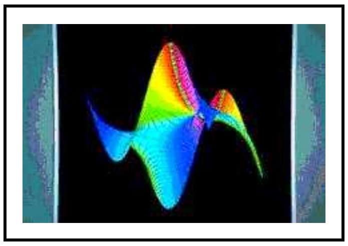
Decomposable objects have generating functions whose analytic and singular structure is the key to asymptotic & probabilistic properties.



$$\downarrow$$

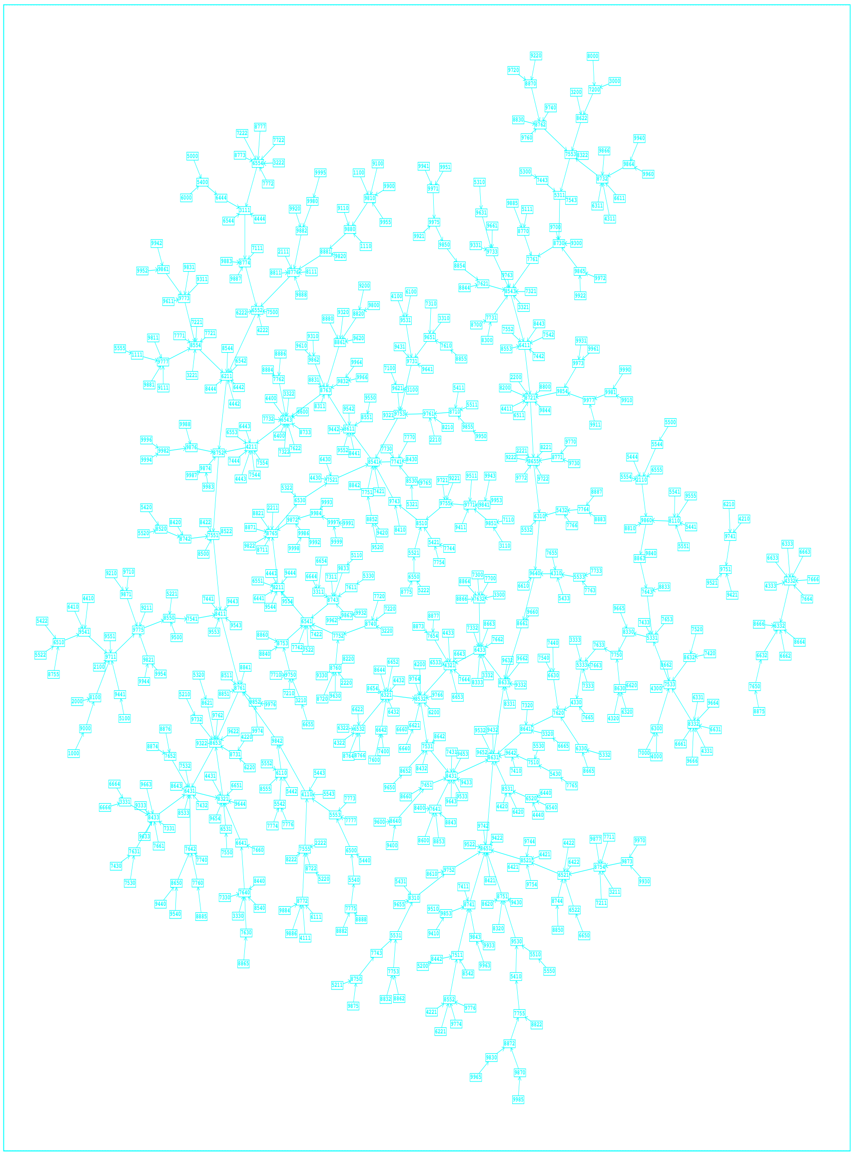
$$\log\left(1+\sum_{n=1}^{\infty}(1+q)^{n(n-1)/2}\frac{z^n}{n!}\right)$$

$$\downarrow$$



"Percolation" on the complete graph is modelled by generating functions and coalescent saddles.

Quantify the laws of random discrete structures



There are n^n functions on a set of n elements.

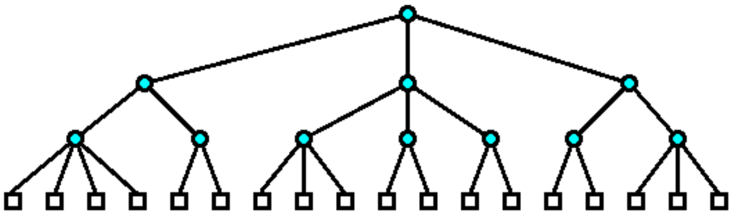
Here: a random functional graph (© F. Bertault)

1 Principles

Two basic principles $\mapsto$ "dictionaries"

SYMBOLIC METHODS

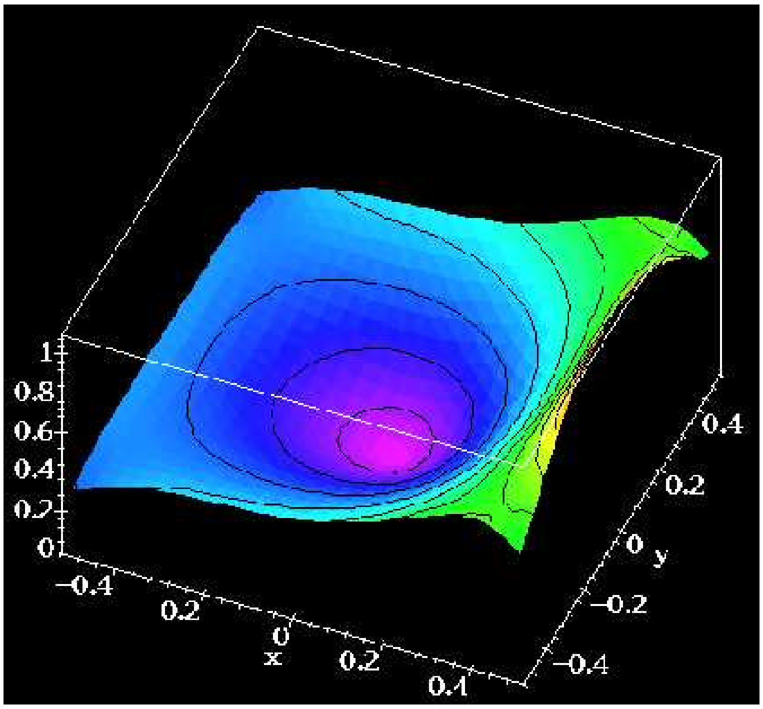
Generating functions

 $\mapsto$ $\mathbf{z}^{11}$

$$\mathbf{z} + \mathbf{z}^2 + \mathbf{z}^3 + 2\,\mathbf{z}^4 + 2\,\mathbf{z}^5 + 4\,\mathbf{z}^6 + 5\,\mathbf{z}^7 + 9\,\mathbf{z}^8 + \cdots$$

$$f(z) = z + f(z^2 + z^3 + z^4)$$

ANALYTIC FUNCTIONS AND SINGULARITIES



KERNEL CONSTRUCTIONS

Dictionary (I)

$$\mathcal{F} \quad \mapsto \quad \{f_n\} \quad \mapsto \quad f(z) = \sum_n f_n \frac{z^n}{n!}.$$

$$\frac{1}{1-f} = 1 + f + f^2 + f^3 + \cdots$$

$$\exp(f) = 1 + f + \frac{1}{2!}f^2 + \frac{1}{3!}f^3 + \cdots$$

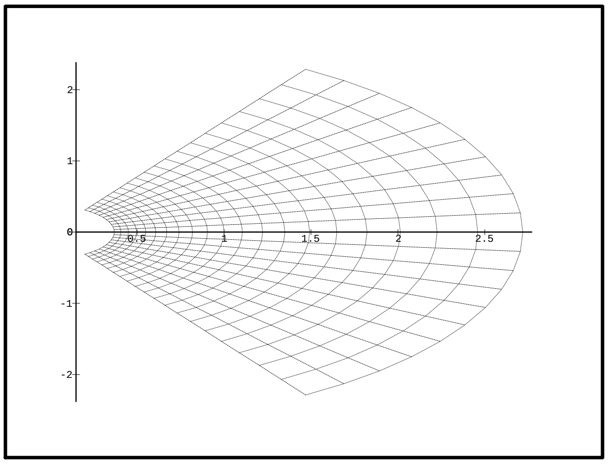
$$\begin{array}{lcl}
\mathbf{A} \cup \mathbf{B} & \mapsto & A(z) + B(z) \\
\mathbf{A} \times \mathbf{B} & \mapsto & A(z) \times B(z) \\
\mathsf{Seq}\ \mathbf{A} & \mapsto & \dfrac{1}{1-A(z)} \\
\mathsf{Set}\ \mathbf{A} & \mapsto & \exp(A(z)) \\
\mathsf{Cycle}\ \mathbf{A} & \mapsto & \log \dfrac{1}{1-A(z)}
\end{array}$$

+ Pólya Theory, Order constraints, Pointing, etc

COMPLEX ASYMPTOTICS
Dictionary (II)

Point of regularity. $f(z) \approx f(z_0) + f'(z_0)(z - z_0)$

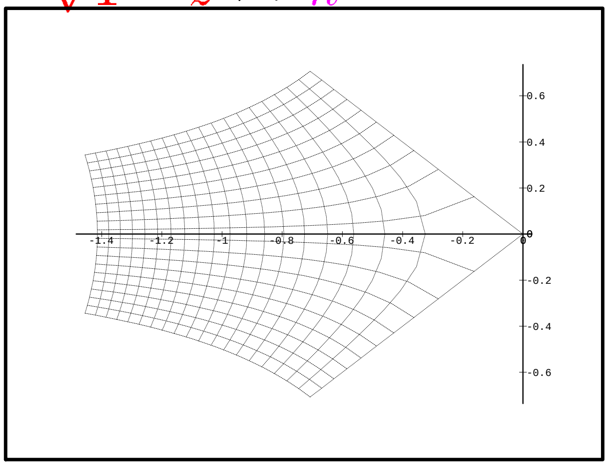
$\exp(z)$



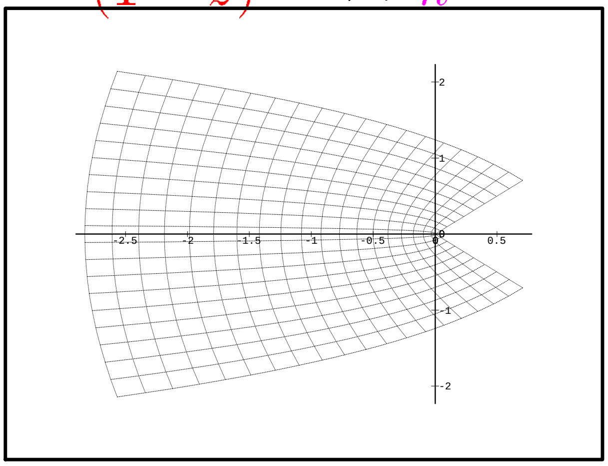
Point of singularity. $\neg\exists \left.\frac{d}{dz} f(z)\right|_{z_0}$

$$(1-z)^{\beta} \quad \mapsto \quad \frac{n^{\beta-1}}{\Gamma(\beta)}$$

$-\sqrt{1-z} \mapsto n^{-3/2}$



$-(1-z)^{3/2} \mapsto n^{-5/2}$



ALGEBRAIC ENUMERATIONS: An extended dictionary — with however a constraint of strong decomposability.

Theorem. *Counting sequences can be computed by top-down ("dynamic programming") approach in time* $O(n^2)$.
Random generation (simulation) cost almost linear $O(n \log n)$.

Fully implemented in Maple's package `Combstruct`.

ANALYTIC ENUMERATIONS: "Usually" decomposability implies isolated singularities.

Hence singularity analysis (saddle point, sometimes) is automatically applicable.

SINGULARITY ANALYSIS (details I)

Mapping between (singular) function scale and coefficient scale

$$[z^n]\frac{1}{(1-z)^\alpha} = \binom{n+\alpha-1}{n-1} \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}.$$

Cf. Darboux's method, Tauberian theorems.

1. *Basic function scale.* Asymptotic scale for **functions**

$$\mathcal{S} = \left\{ A(\frac{1}{1-z} \;\middle|\; A(u) = u^\beta (\log u)^\gamma (\log\log u)^\delta \right\}$$

gives asymptotic scale for **coefficients**

$$[z^n]A(\frac{1}{1-z}) \approx \frac{1}{\Gamma(\beta)}\frac{A(n)}{n}.$$

2. *Transfers.* For $\sigma(z)$ in $\mathcal{S}$

$$\begin{array}{ccccccc}
f(z) & = & O(\sigma(z)) & \Rightarrow & f_n & = & O(\sigma_n) \\
f(z) & = & o(\sigma(z)) & \Rightarrow & f_n & = & o(\sigma_n) \\
f(z) & \sim & \sigma(z) & \Rightarrow & f_n & \sim & \sigma_n \\
f(z) & \sim & \sum_j \sigma^{\langle j\rangle}(z) & \Rightarrow & f_n & \sim & \sum_j \sigma_n^{\langle j\rangle}
\end{array}$$

3. *Conditions.* The *Camembert Condition* (strong regularity!).

SINGULARITY ANALYSIS (details II)

Proof technique. Cauchy's coefficient formula and Hankel contours achieve singularity "capture"

$$f(z) = \frac{1}{(1-z)^\beta}\left(\frac{1}{z}\log(1-z)^{-1}\right)^{-1/2}.$$

Step 1. Hankel contour: $z = 1 - \frac{t}{n}$

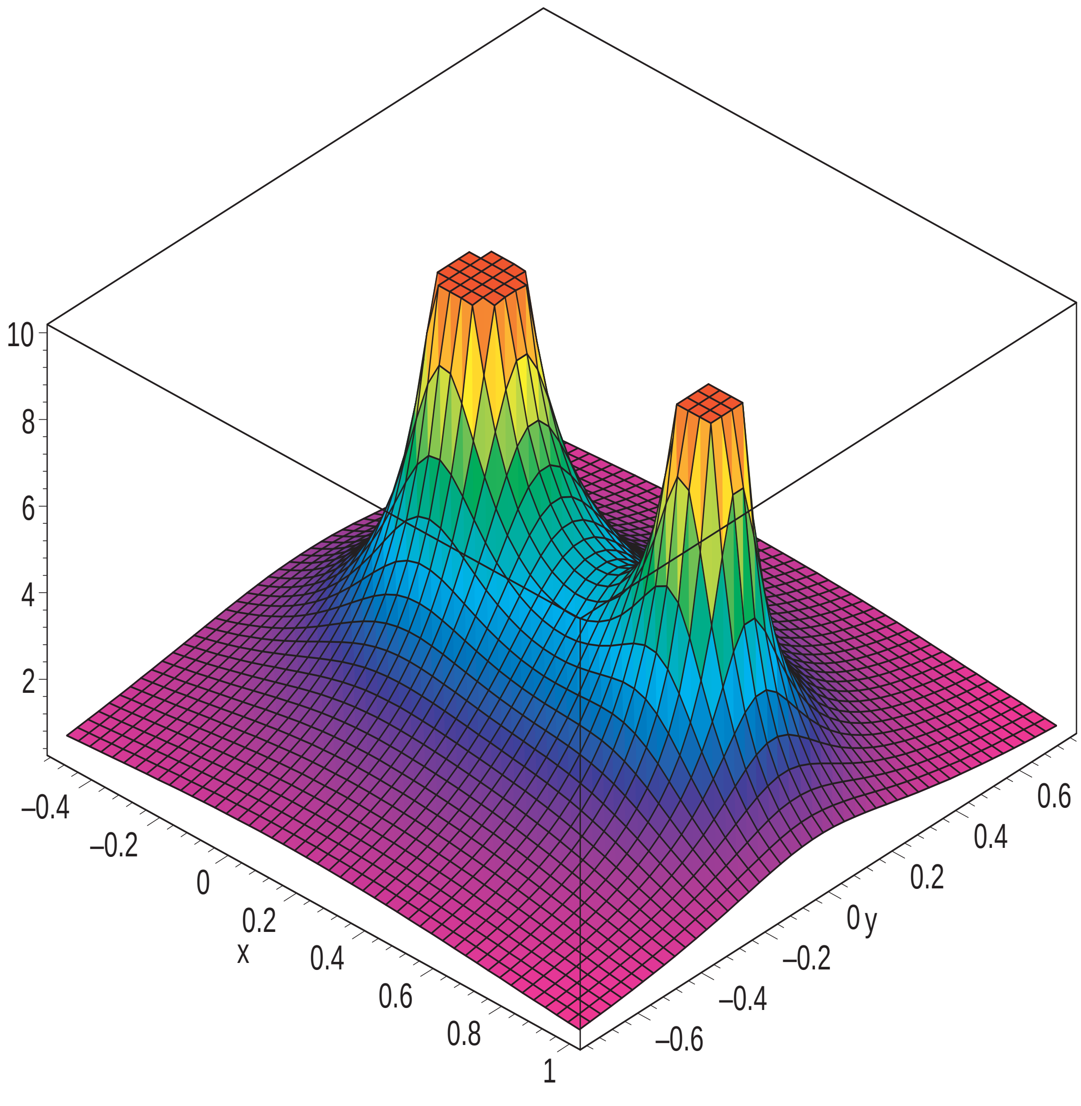
$$f_n = \frac{1}{2i\pi}\int_{\mathcal{H}}\left(\frac{n}{t}\right)^\beta (\log n - \log t)^{-1/2}\,(..)$$

$$\left(1-\frac{t}{n}\right)^{-n-1}\frac{dt}{n}$$

Step 2. Approximation.

$$n.f_n \approx \left[n^\beta(\log n)^{-1/2}\right]\cdot\frac{1}{2i\pi}\int_{\mathcal{H}}\frac{e^t}{t^\beta}\,dt.$$

Integral gives $1/\Gamma(\beta)$!!!


10
8
6
4
2
–0.4
–0.2
0
0.2
x
0.4
0.6
0.8
1
–0.6
–0.4
–0.2
0
y
0.2
0.4
0.6

2 Two examples

The example of HIERARCHIES (I)

→ A problem in statistical classification

```
> with(combstruct):
> hier:=[H,H=Union(Z,Set(H,card>1)),labelled]:
> draw(hier,size=10);
```

Set(Set(Set(Set(Z_6, Z_2, Set(Z_4, Z_8)), Z_7), Z_1), Set(Z_5, Z_3, Z_{10})):

```
> seq(count(hier,size=j),j=0..40);
```

0, 1, 1, 4, 26, 236, 2752, 39208, 660032, 12818912, 282137824, 6939897856, 188666182784, 5617349020544, 181790703209728, &c

```
> gfeqns(op(2..3,hier),z);
```

$$[Z(z) = z, H(z) = Z(z) + e^{H(z)} - 1 - H(z)]$$

$$\{Z(z) = z, H(z) = -\mathrm{LambertW}\left(-\frac{1}{2}e^{(1/2\,z - 1/2)}\right) + \frac{1}{2}z - \frac{1}{2}\}$$

Maple Combstruct

Maple Output

The example of HIERARCHIES (II)

$$F(z,H) := 2H + 1 - z - e^H = 0$$

- Dominant singularity from failure of the Implicit Function Theorem.

$$\rho := -1 + 2\log 2 = 0.3862943611198906188$$

Local quadratic relation implies square-root singularity.

$$H\_sing := \ln(2) - \sqrt{-1 + 2\ln(2)}\,\Delta + \left(\frac{1}{6} - \frac{1}{3}\ln(2)\right)\Delta^2 -$$

$$\Delta = \sqrt{\left(1 - \frac{z}{-1 + 2\ln(2)}\right)}$$

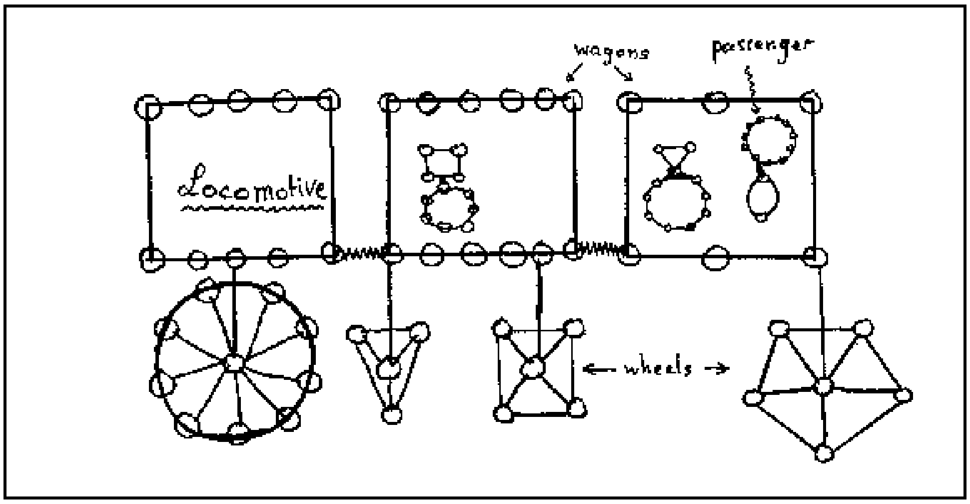
$$H\_n\_asympt := \frac{1}{2}\,\frac{n!\sqrt{-1 + 2\ln(2)}\left(\frac{1}{n}\right)^{3/2}}{\sqrt{\pi}\,(-1 + 2\ln(2))^n}$$

(Last approximation by singularity analysis.)

++ Statistics on Node Degrees (Poisson) and Node Types (Gaussian).

The example of TRAINS

- Cope with complex structural "specifications"



train.mws - [Server 1]

```
> with(combstruct);
```

[*allstructs*, *count*, *draw*, *finished*, *gfeqns*, *gfseries*, *gfsolve*, *iterstructs*, *nextstruct*]

```
> TRAINS:={T=Prod(Loco,Sequence(Wagon)),
     Wagon=Prod(Loco,Set(Passenger)),
     Loco=Sequence(Prod(Board,Wheel),card>=1),
  Passenger=Prod(Head,Belly),
     Board=Prod(Z,Z), Wheel=Prod(Z,Cycle(Z)),
     Head=Cycle(Z), Belly=Cycle(Z)}:
> seq(count([T,TRAINS,labelled],size=n),n=0..20);
```

0, 0, 0, 0, 24, 60, 240, 1260, 88704, 786240, 10800000, 151351200, 4295047680, 77837760000, 1696330097280, 37148330419200, 1032008553354240, 26762644285977600, 759278023249374720, 22080501641848366080, 710352165477065472000

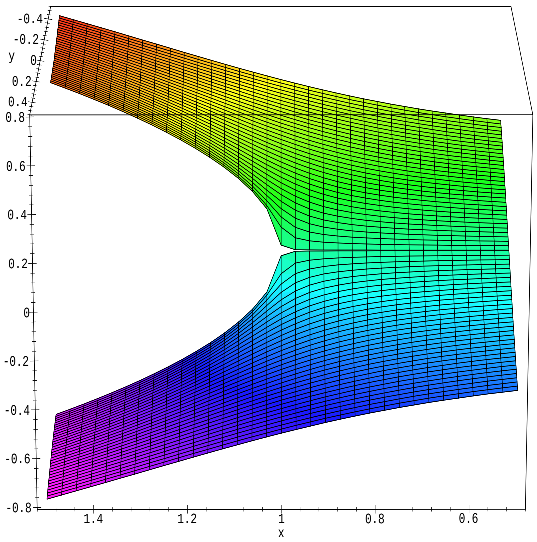
```
> subs(gfsolve(TRAINS,labelled,z),T(z));
```

$$-\frac{z^3 \ln\left(-\frac{1}{-1+z}\right)}{-1+z^3\ln\left(-\frac{1}{-1+z}\right)+z^3\ln\left(-\frac{1}{-1+z}\right)e^{\left(\ln\left(-\frac{1}{-1+z}\right)^2\right)}}$$

```
                                     (- n)
0.1008557594 (0.5180547070)      + ...
```

3 Schemas

“Critical Exponent” $= \frac{1}{2}$ & Implicit/Recursive objects



$\Im(1-\sqrt{1-z})$

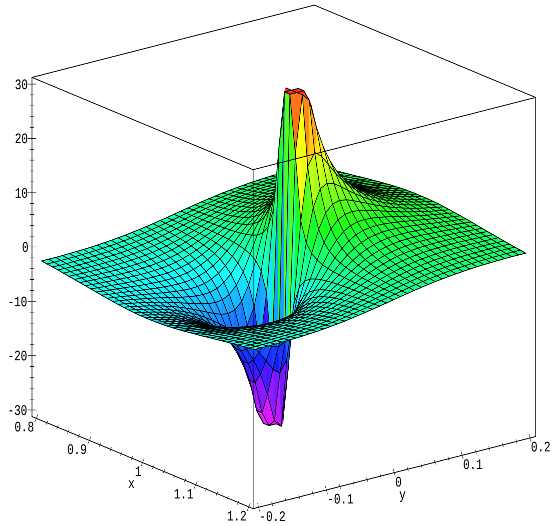
Square-root singularity

$$f(z) = c_0 - c_1\sqrt{1-z/\rho} + \cdots\,, \qquad f_n \approx \rho^{-n} n^{-3/2}$$

Arises from failure of Implicit Function Theorem

- “Most” recursive objects, e.g., non-crossing graphs, context-free grammars
- Hierarchies
- Expression trees $\sin(x+y) \mapsto \sin x \cos y + \cos x \sin y$
- Molecules, Pólya 1937 $f(z) = z e^{f(z)+\frac{1}{2}f(z^2)+\frac{1}{3}f(z^3)+\cdots}$.
- Parking allocations
- Lattice path combinatorics

Critical Exponent $= -1$ & Sequences



$\Im(1-z)^{-1}$

- Permutations as sequences of labelled atoms
- "Regular events" of theoretical C.S.
- Words and sequences. E.g. patterns in random strings, runs, etc DNA statistics(?)
- Balls in urns: distinguishable urns

$$\frac{1}{1-\frac{z}{1-z}}, \quad \frac{1}{1-(e^z-1)}$$

- Coin fountains

$$\cfrac{1}{1-\cfrac{z}{1-\cfrac{z^2}{\ddots}}}$$

- Complex objects driven by sequences, e.g., Trains (!)

Critical exponent $= -\alpha < 0$ & **Sets (exp-log)**

$$\boxed{K(z) \approx \alpha \log \frac{1}{1 - z/\rho}, \ \ G(z) = \exp(K(z)) \approx (1 - z/\rho)^{-\alpha}}$$

- Permutations = "pure case" $P(z) = 1/(1-z)$
- Mappings = set of components
- Polynomials over finite fields $GF(q)$
- 2-regular graphs, etc

$$\boxed{K_n \approx \rho^{-n} n^{-1}, \qquad G_n \approx \rho^{-n} n^{\alpha-1}}$$

EXAMPLE. A Prime Number Theorem for Polynomials over finite fields:

— density of primes $\approx \frac{1}{\log n}$,

— density of irreducibles $\approx \frac{1}{n}$

+ Number of components : Mean,Var: $\mu_n \sim \sigma_n^2 \sim \alpha \log n$

4 Distributions

LAWS, LIMIT DISTRIBUTIONS

Gaussian laws for $X_n^* = (X_n - \mu_n)/\sigma_n$

$$\Pr\{X_n^* \le x\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt.$$

Bivariate GF: Exp-log scheme. $G(z,u) = \exp(uK(z))$.

Singularity Perturbation

→ Movable singularity: $(1 - z/\rho(u))^{\beta}$

→ Movable exponent: $(1 - z/\rho)^{\beta(u)}$

$$\begin{cases} [z^n](1 - z/\rho(u))^{\beta} & = & (\rho(u))^{-n} n^{\alpha - 1} \\ [z^n](1 - z/\rho)^{\beta(u)} & = & (\rho^{-n} n^{-1} \left(e^{-\beta(u)}\right)^{\log n} \end{cases}$$

"QuasiPowers" $\Longrightarrow$ Central Limit Theorem

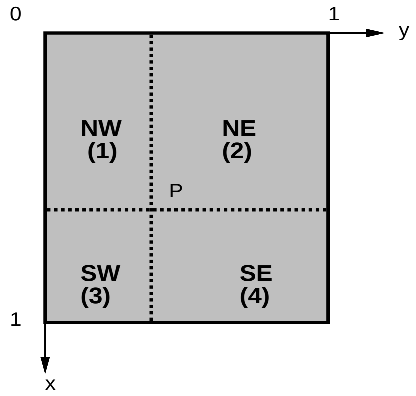
EXAMPLE. Exp-log scheme.

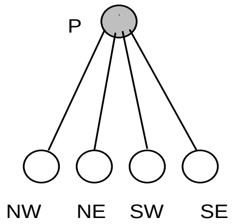
Cycles in Perms ≡ Factors in Polynomials ≡ Gauß

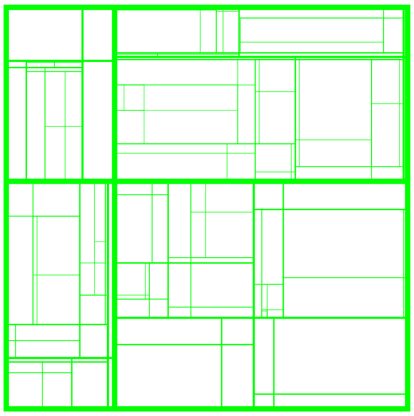
Application to Analysis of Algorithms —Polynomial factorization algorithms in computer algebra

The example of QUADTREES

Singularity Perturbation for functional equations







$$\mathbf{I}f(z) = \int_0^z f(u)\,\frac{du}{1-u}, \quad \mathbf{I}f(z) = \int_0^z f(u)\,\frac{du}{u(1-u)}$$

$$C(z) = t(z) + 2^d \mathbf{J}^{d-1}\mathbf{I}C(z).$$

for additive cost measures dim $= d$

Partial match: know s/d coordinates

$$\mathcal{L}f(z) = 0$$

where $\mathcal{L}$ is a linear differential operator.

Substitute $f(z) \approx (1-z)^{\alpha}$ and get indicial equation $I(\alpha) = 0$; for $s/d = 1/2$:

$$f_n \approx n^{(-3+\sqrt{17})/2} \approx n^{0.56}$$

Exact search: Use bivariate GF's

$$\mathcal{L}_u f(z) = 0$$

and get a variable exponent

$$f_u(z) \approx (1-z)^{\alpha(u)}.$$

The exp-log schema applies $\Longrightarrow$ Gaussian law.

5 Where are we?

A nicely decomposable world of combinatorial objects
+ A smooth complex-analytic world of generating fns.

= A world of regular probabilistic phenomena

♡ Scope: decomposability, divide-and-conquer paradigm and basic data structures in C.S.

♢ Techniques: Uniform Asymptotics from singularity analysis or saddle points with Perturbation theory.

Compare to: Probabilistic methods.

Research: A classification of schemas. Extensions to order constraints, pointing, bijections, etc.

♠ Limits: constrained objects (e.g. configurations on lattices)

♣ But: Approximate complex objects by decomposable models.
Make use of partial decomposability, e.g., graphs

References

— Flajolet & Sedgewick *Analytic Combinatorics*, see WWW

— Bruno Salvy and Paul Zimmermann on "automatic analysis"

— Knuth's *The Art of Computer Programming* for relations to CS.

— Herb Wilf's *Generatingfunctionology*

— Andrew Odlyzko's survey on asymptotic methods in *Handbook of Combinatorics*

— Ed Bender and Bruce Richmond for limit laws

— Michèle Soria on "schemas"

— Hsien-Kuei Hwang on limit laws

— Brigitte Vallée for AofA & dynamical systems

For pointers and detailed references:

http://www-rocq.inria.fr/algo/AofA/index.html
mailto: Philippe.Flajolet@inria.fr