Santiago de Chile

DEC 2006

information & randomness 2006 Santiago, Chile • December 11-15

SINGULAR COMBINATORICS

A. Symbolic Methods

Philippe Flajolet, INRIA, Rocquencourt

`http://algo.inria.fr/flajolet`

Based on *Analytic Combinatorics*, Flajolet & Sedgewick, C.U.P., 2007$^+$.

ANALYTIC COMBINATORICS

• Find quantitative properties of large discrete structures = random combinatorial structures.

• Identify the fundamental analytic structures $\neq$ probabilistic approaches.

Via complex analysis establish relationship

$$\text{Combinatorics} \rightsquigarrow \text{Analysis} \rightsquigarrow \text{Asymptotics}$$

• Organization into major schemas where chain can be worked out: "combinatorial processes" // stochastic processes.

Example: "bag" process (Set); "row" process" (Seq).

Universality: E.g. take a random tree of size n (large):

— Height is with high probabiliy (w.h.p.) $O(\sqrt{n})$;

— Any designated pattern ϖ occurs on average $C_{\varpi} \cdot n$, and distribution is asymptotically normal.

• Such properties hold for a very wide range of local construction rules (also Galton-Watson trees conditioned on size).

• Similar properties hold for "molecule trees", random mappings, etc. But labelled trees based on order properties belong to a different universality class, with e.g., logarithmic height.

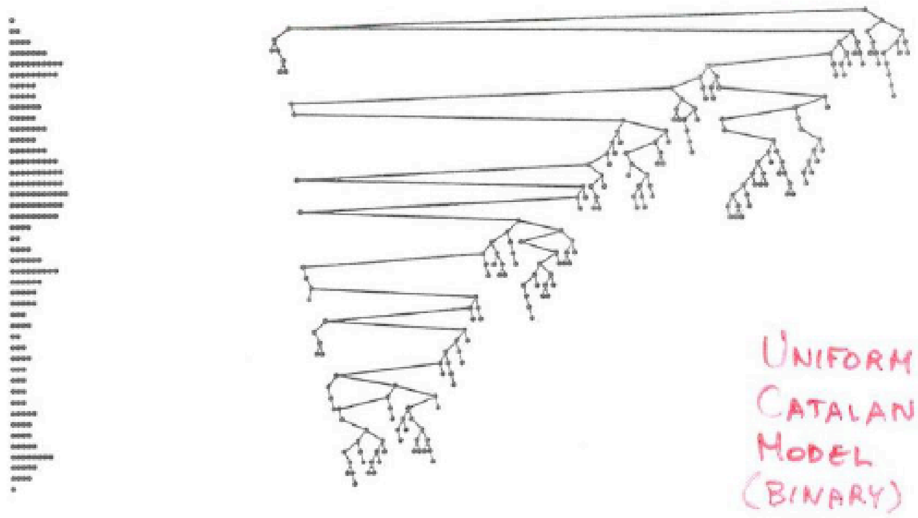


Figure 5.5 A random binary tree with 256 internal nodes

“$\sqrt{n}$–trees”

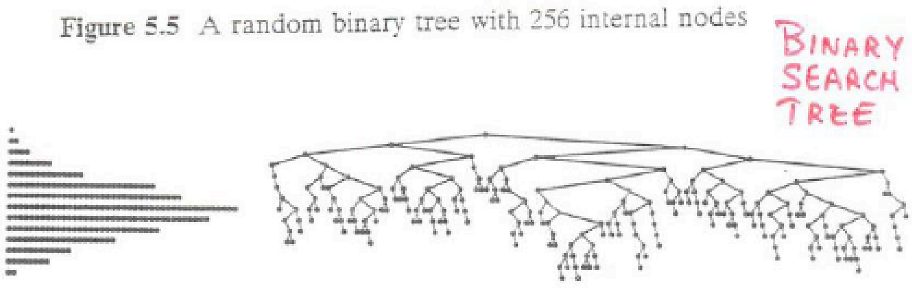


Figure 5.11 A binary search tree built from 256 randomly ordered keys

“log–trees”

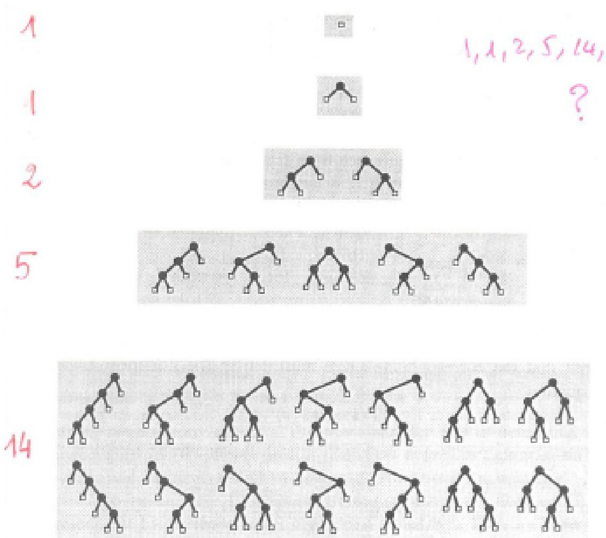


Figure 3.1 All binary trees with 1, 2, 3, 4, and 5 external nodes

Analytic combinatorics $\rightsquigarrow$

A. Counting Generating Function

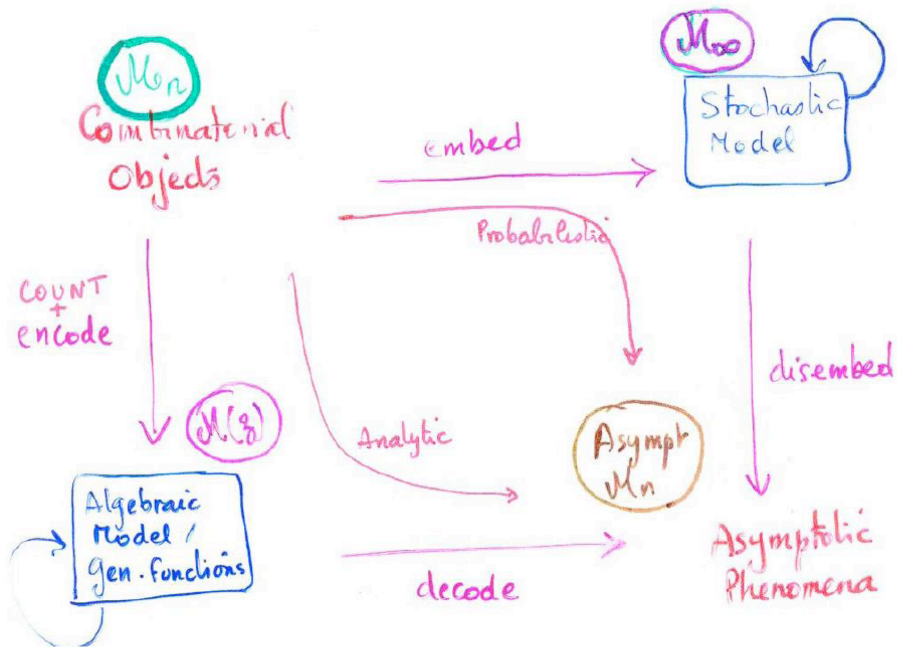
B. Analytic properties of GF

Singularities + transfer to coefficients

C. Perturbation for **distributions**.

SYMBOLIC METHODS + COMPLEX ASYMPTOTICS + PERTURBATION.

Duality: Combinatorics versus probability



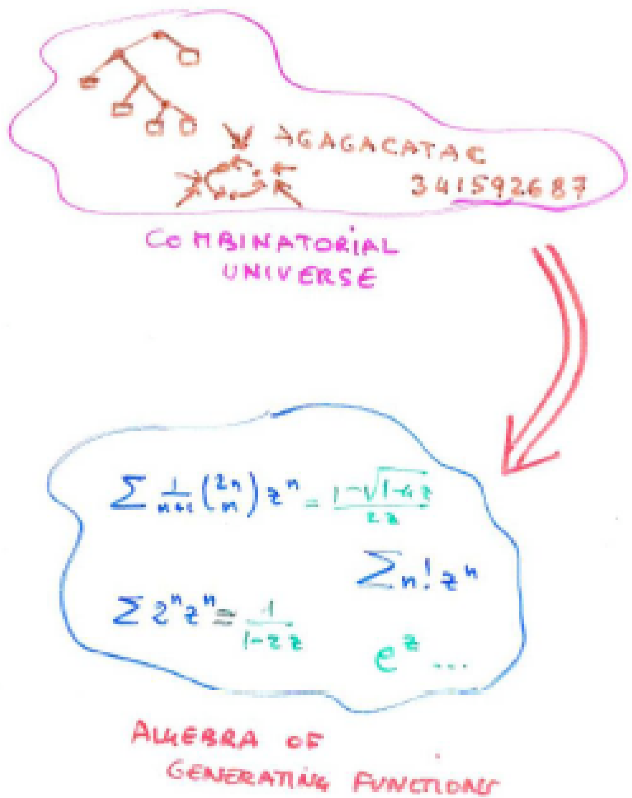
Brownian motion, continuum random tree, etc.

PART A. SYMBOLIC METHODS

Goal: develop generic tools to determine generating functions $\equiv$ GFs.

Approach: Formulate a programming language to specify combinatorial structures such that translation into GFs is **automatic**.

Parallels Joyal's theory of species (BLL's book). Similar in spirit to Jackson & Goulden's book. Cf Rota/Stanley. Formalizes recipes known to earlier combinatorialists.



Abstraction:

Embed a fragment of elementary set theory into a **language of constructions**. Map to algebra(s) of special functions.

1 UNLABELLED STRUCTURES AND OGFS

Ordinary Generating Function (OGF)

$$(f_n) \longrightarrow f(z) := \sum_{n=0}^{\infty} f_n z^n.$$

(f_n) is number sequence, e.g., counting sequence.

Later: Exponential Generating function (EGF): $(f_n) \longrightarrow f(z) := \sum_{n=0}^{\infty} f_n \frac{z^n}{n!}$.

$\mathcal{C}$ = a combinatorial class: at most denumerable set with size function.

$\mathcal{C}_n$ = subclass of objects of size n.

C_n = # objects of size n = $\mathrm{card}(\mathcal{C}_n)$.

$C(z)$ = OGF := $\displaystyle\sum_{n\geq 0} C_n z^n = \sum_{\gamma\in\mathcal{C}} z^{|\gamma|}$.

Count up to combinatorial isomorphism: $\mathcal{C} \cong \mathcal{D}$ iff $\exists$ size-preserving bijection.

Atom: $\mathcal{Z} \mapsto z$; neutral element: $\mathcal{E} \mapsto 1$.

How many binary trees B_n with n external nodes?

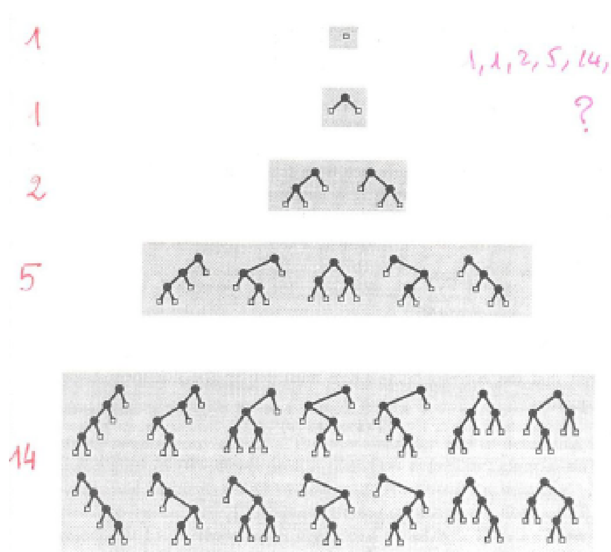


Figure 3.1 All binary trees with 1, 2, 3, 4, and 5 external nodes

$\mathcal{B} = \square + \bullet, (\mathcal{B} \times \mathcal{B})$.

Euler-Segner (1743): Recurrence

$$B_n = \sum_{k=1}^{n-1} B_k B_{n-k}.$$

Form OGF: $B(z) = z + (B(z) \times B(z))$.

Solve equation (quadratic):

$$B(z) = \tfrac{1}{2}(1 - \sqrt{1-4z}) = \tfrac{1}{2} - \tfrac{1}{2}(1-4z)^{1/2}.$$

Expand:

$$B_n = \frac{1}{n}\binom{2n-2}{n-1}$$ (Catalan numbers)

Analogy: $\boxed{\mathcal{B} = \square \; + \; (\bullet \mathcal{B} \times \mathcal{B})} \rightsquigarrow \boxed{B(z) = z \; + \; (B(z) \times B(z))}$

Outline

Define a collection of constructions

union, product, sequence, set, cycle, . . .

allowing for *recursive definitions*.

meta-THM1: *OGFs are automatically computable (equations!)*

meta-THM2: *Counting sequences are automatically computable in time* $O(n^2)$*, and even* $O(n^{1+\epsilon})$*.*

meta-THM3: *Random generation is fast in* $O(n \log n)$ *arithmetic op´ns.*

Theorem. *There exists a dictionary:*

Construction	OGF
$\mathcal{C} = \mathcal{A} + \mathcal{B}$	$C(z) = A(z) + B(z)$
$\mathcal{C} = \mathcal{A} \times \mathcal{B}$	$C(z) = A(z) \cdot B(z)$
$\mathcal{C} = \mathrm{S{\scriptsize EQ}}(\mathcal{A})$	$C(z) = \dfrac{1}{1 - A(z)}$
$\mathcal{C} = \mathrm{MS{\scriptsize ET}}(\mathcal{A})$	$C(z) = \mathrm{Exp}(A(z))$
$\mathcal{C} = \mathrm{PS{\scriptsize ET}}(\mathcal{A})$	$C(z) = \widehat{\mathrm{Exp}}(A(z))$
$\mathcal{C} = \mathrm{C{\scriptsize YC}}(\mathcal{A})$	$C(z) = \mathrm{Log}\,\dfrac{1}{1 - A(z)}$

$\mathcal{E}$ or $\mathbf{1}$: "neutral class" formed with element of size $0 \mapsto E(z) = 1$.
$\mathcal{Z}$: "atomic class" formed with element of size $1 \mapsto E(z) = 1$.

$\mathrm{Exp}(g(z)) = \exp\left(\sum_{k\geq 1} \frac{1}{k} g(z^k)\right)$; $\widehat{\mathrm{Exp}}(g(z)) = \exp\left(\sum_{k\geq 1} \frac{(-1)^k}{k} g(z^k)\right)$;

$\mathrm{Log}(g(z)) = \sum_{k\geq 1} \frac{\varphi(k)}{k} g(z^k)$ with $\varphi(k)$= Euler totient.

Proofs. $\mathcal{A} \mapsto A(z) = \sum A_n z^n = \sum_\alpha z^{|\alpha|}$.

— Union: $\mathcal{C} = \mathcal{A} + \mathcal{B}$; $\sum_\gamma = \sum_\alpha + \sum_\beta$. $\boxed{C(z) = A(z) + B(z)}$

— Product: $\mathcal{C} = \mathcal{A} \times \mathcal{B}$; $\sum_\gamma = \sum_\alpha \cdot \sum_\beta$. $\boxed{C(z) = A(z) \cdot B(z)}$

— Sequence: $\mathcal{C} = \textsc{Seq}(\mathcal{A})$ means $\mathcal{C} = \mathbf{1} + \mathcal{A} + (\mathcal{A} \times \mathcal{A}) + \cdots$. $\boxed{C(z) = \frac{1}{1 - A(z)}}$

— Multiset: $\mathcal{C} = \textsc{MSet}(\mathcal{A})$ means $\mathcal{C} \cong \prod_\alpha (\mathbf{1} + \{\alpha\})$, so that

$$C(z) = \prod_\alpha \frac{1}{1 - z^{|\alpha|}} = \prod_{n \geq 1} \frac{1}{(1 - z^n)^{A_n}},$$

and conclude by $C(z) = \exp(\log C(z)) \ldots$ $\boxed{C(z) = \mathrm{Exp}(A(z))}$.

— Cycle: (omitted) $\varphi(k)$ is Euler's totient function.

Example 1. Binary words

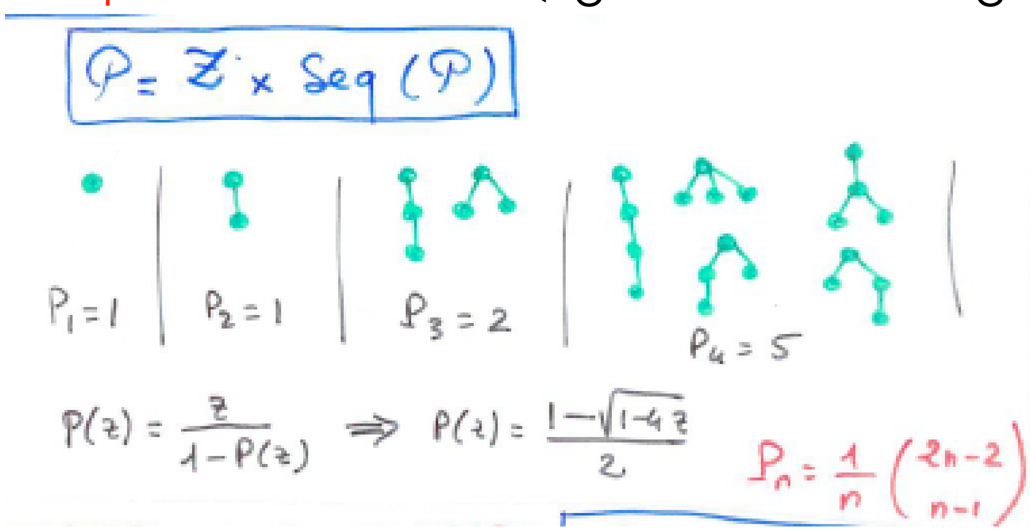
$$\mathcal{W} = \textsc{Seq}(\{a, b\}) \quad \Longrightarrow \quad W(z) = \frac{1}{1-2z}.$$

Get $W_n = 2^n$ (!?). Words starting with b and < 4 consecutive a's:

$$\mathcal{W}^{\bullet} \cong \textsc{Seq}(b \times (1+a+aa+aaa)) \quad \Longrightarrow \quad W^{\bullet}(z) = \frac{1}{1-(z+z^2+z^3+z^4)}.$$

Longest run statistics lead to rational functions (Feller).

Example 2. Plane trees ("general" = all degrees allowed)



Example 3. Nonplane trees (all degrees allowed)

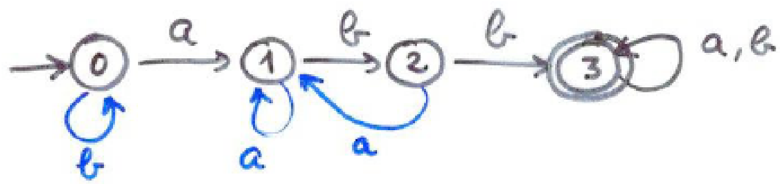
$\mathcal{U} = \mathcal{Z} \times \textbf{MS}\textsc{et}(\mathcal{U})$. $U_1 = 1,\ U_2 = 1,\ U_3 = 2, U_4 = 5$.

$$U(z) = z \exp\left(\frac{1}{1}U(z) + \frac{1}{2}U(z^2) + \frac{1}{3}U(z^3) + \cdots\right).$$

Cayley: recurrences; Pólya: asymptotics of this infinite functional equation.

Exercise: computable in polynomial time ($O(n^2)$).

Example 4. Words containing a pattern (abb)



$\mathcal{L}_j$:= language accepted from state j.

$$\{\mathcal{L}_0 = a\mathcal{L}_1 + b\mathcal{L}_0,\ \mathcal{L}_1 = a\mathcal{L}_1 + b\mathcal{L}_2, \mathcal{L}_2 = a\mathcal{L}_1 + b\mathcal{L}_3, \ldots\}$$

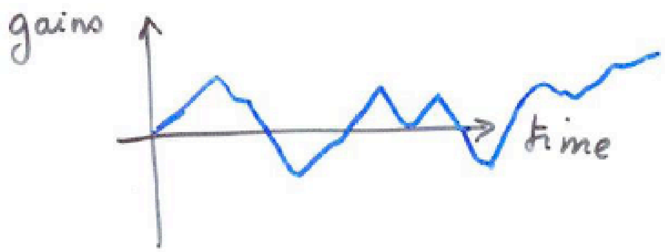
Theorem. Regular language (finite automaton) has rational GF.

$$Reg \quad \mapsto \quad \mathbb{Q}(z).$$

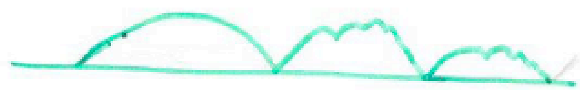
Patterns of all sorts in words. Applications in pattern matching algorithms and computational biology.

Borges' Theorem: *Large enough text contains any finite set of patterns w.h.p.*

Example 5. Walks and excursions.



walks $\mapsto$ $Q(z, \sqrt{1-4z^2})$

Excursion = 

$= \mathrm{Seq}(\nearrow \text{Excursion} \searrow)$

Positive path = Excursion $\times$ $\mathrm{Seq}(\nearrow \text{Excursion})$

Draw game = $\mathrm{Seq}(\searrow \overline{\text{Excursion}} \nearrow + \nearrow \text{Excursion} \searrow)$ etc

Exercise A. Integer compositions. Argue that $C_n = 2^{n-1}$ since

$$\mathcal{C} = \textsc{Seq}(\mathcal{N}),\ \mathcal{N} = \mathcal{Z} \times \textsc{Seq}(\mathcal{Z}) \quad \Longrightarrow C(z) = \frac{1}{1 - \frac{z}{1-z}} = \frac{1-z}{1-2z}.$$

Exercise B. Denumerants. In how many ways can one give change for n cents, given coins of $1, 2, 5, 10$c?

$$D(z) = \frac{1}{(1-z)(1-z^2)(1-z^5)(1-z^{10})}.$$

Exact form of coefficients? Asymptotics?

Exercise C. Unary binary trees. $U = z(1 + U + U^2)$.

Exercise D. Binary trees, general plane trees, excursions, and polygonal triangulations are all enumerated by Catalan numbers $C_n = \frac{1}{n+1}\binom{2n}{n}$. Why?

Simple families of plane trees.

Let $\Omega \subseteq Z_{\geq 0}$ be the set of allowed (out)degrees. Define

$$\phi(y) := \sum_{w \in \Omega} y^{\omega}.$$

Then the simple family $\mathcal{Y}$ has OGF:

$$Y(z) = z\phi(Y(z)).$$

If ϕ is finite, get an algebraic function.

Lagrange Inversion Theorem.

$$[z^n]Y(z) = \frac{1}{n} \operatorname{coeff}[w^n]\phi(w)^n.$$

If ϕ is finite, get multinomial sums.

2 LABELLED STRUCTURES AND EGFS

EGF = exponential generating function

$$(f_n) \qquad \longrightarrow f(z) = \sum_{n \geq 0} f_n \frac{z^n}{n!}.$$

A labelled object has atoms that bear distinct integer labels (canonically numbered on $[1 \,..\, n]$).

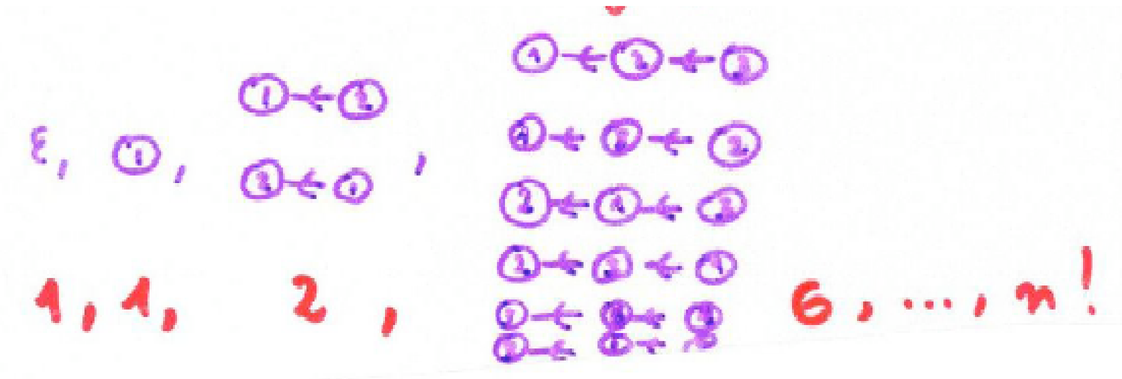
Unlabelled: "anonymous atoms". Labelled: distinguished atoms or colours.

Example. How many (undirected) graphs on n (distinguishable) vertices? $G^n = 2^{n(n-1)/2}$.

Graphs: unlabelled problem is harder (Pólya theory). In general, can get unlabelled by identification of labelled.

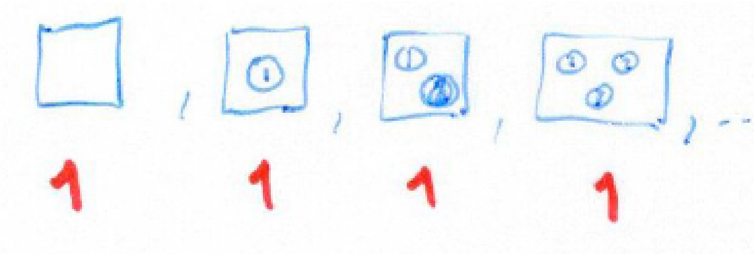
PERMUTATIONS = typical labelled objects: write $\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma_1 & \sigma_2 & \cdots & \sigma_n \end{pmatrix}$

as $\sigma_1 \sigma_2 \cdots \sigma_n$ and view as linear digraph that is labelled:



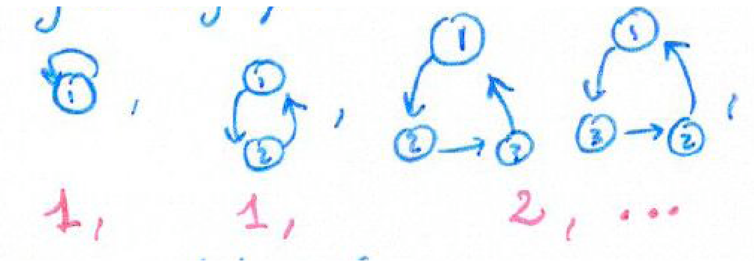
EGF is $\frac{1}{1-z}$ since $P(z) = \sum_n n! \frac{z^n}{n!}$.

DISCONNECTED GRAPHS (labelled) = no edges aka "Urns".



EGF is $U(z) = \exp(z) = e^z$.

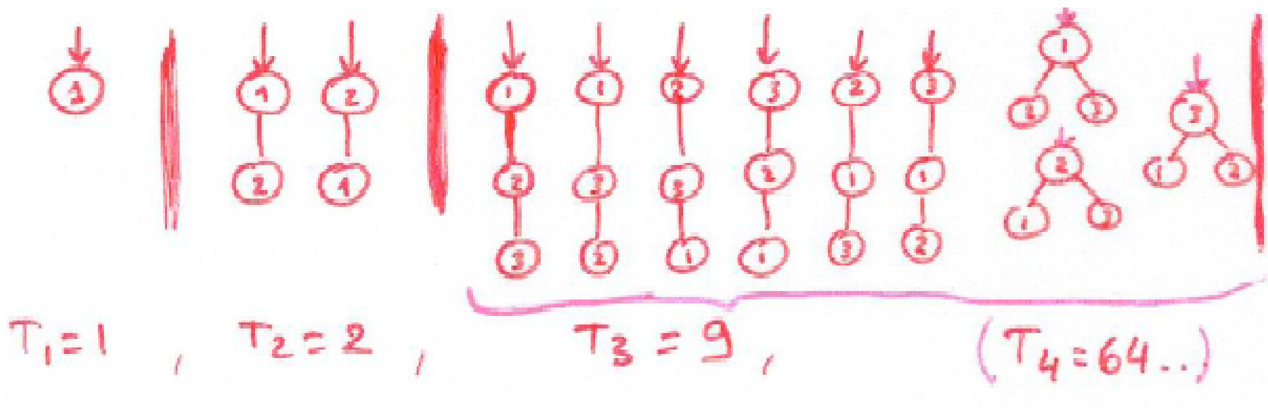
CYCLIC GRAPHS (directed)



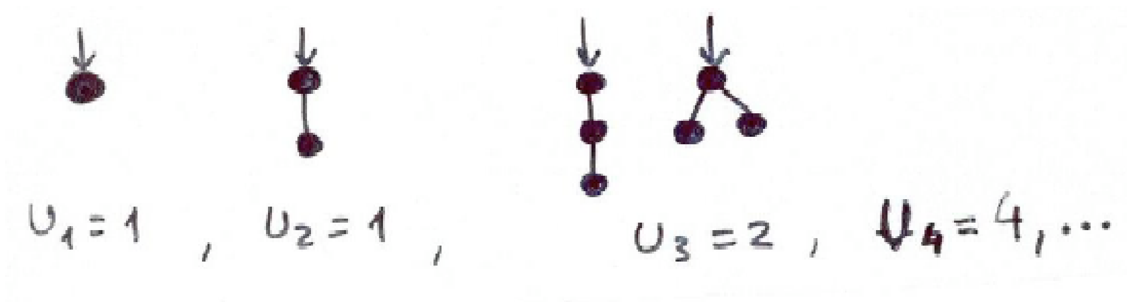
EGF $K(z) = \log \frac{1}{1-z}$.

ROOTED TREES (graphs) nonplane and labelled

$T_n = ??$

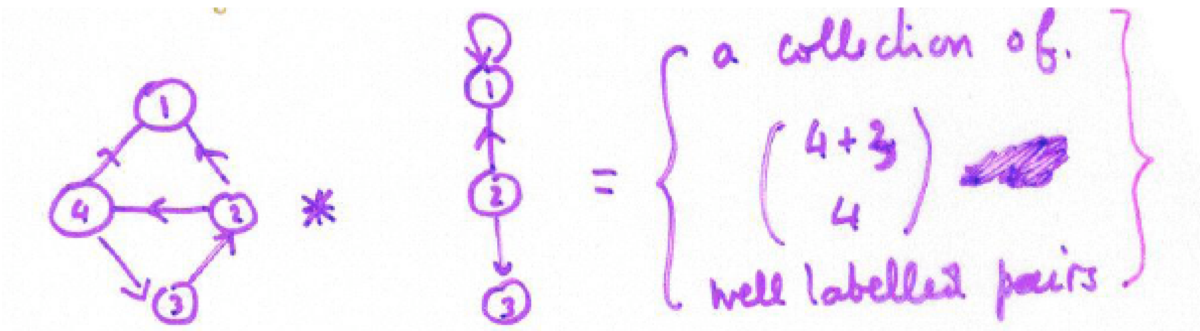


≫ Unlabelled:



Labelled product. Let $\mathcal{A}$ and $\mathcal{B}$ be labelled classes. Then the cartesian product $\mathcal{A} \times \mathcal{B}$ is *not* well-labelled (why?).

Given (β, γ) form all possible *relabellings* that preserve the order structure within β, γ, while giving rise to well-labelled objects.



• Labelled product of two objects.

$$(\alpha \star \beta) := \{\gamma \mid \gamma = (\alpha', \beta')\},$$

where γ is well-labelled and $\alpha' \equiv_{\text{order}} \alpha$ and $\beta' \equiv_{\text{order}} \beta$.

• Labelled product of two classes.

$$\mathcal{C} := \bigcup_{\alpha \in \mathcal{A}, \beta \in \mathcal{B}} (\alpha \star \beta).$$

Example: $\mathcal{U}$ = class of all totally disconnected graphs

1 1 1 1

$\underbrace{\mathcal{U} \star \mathcal{U} \star \cdots \star \mathcal{U}}_{m \text{ times}}$ = Allocations of n elements onto m cells; functions from $[1..n] \to [1..m]$

$\mathcal{V} = \mathcal{U} \setminus \{\varepsilon\}$ 0, 1, 1, 1, ...

$\mathcal{V} \star \mathcal{V} \star \cdots \star \mathcal{V}$ = allocations with no empty cell ≡ surjections from $[1..n]$ to $[1..m]$

GFs; Stirling numbers.

Sequences, Sets, Cycles

- $\mathcal{E}$ (or $\mathbf{1}$): neutral class.
- $\mathcal{Z}$: atomic class $\equiv \boxed{1}$.
- Define $\textsc{Seq}(\mathcal{A})$, $\textsc{Set}(\mathcal{A})$, $\textsc{Cyc}(\mathcal{A})$ by relabellings:

$$\textsc{Seq}(\mathcal{A}) = \mathbf{1} + \mathcal{A} + (\mathcal{A} \star A) + \cdots .$$

Sets: quotient up to perms. Cyc: up to cyclic perms.

— Perms $\mathcal{P} \cong \textsc{Seq}(\mathcal{Z})$
— Urn $\mathcal{U} \cong \textsc{Set}(\mathcal{Z})$
— Circulars graphs $\mathcal{K} \cong \textsc{Cyc}(\mathcal{Z})$
— m–functions: $\mathcal{F}^{[m]} \cong \overbrace{\mathcal{U} \star \cdots \star \mathcal{U}}^{m \text{ times}} \equiv \textsc{Seq}_m(\mathcal{U})$
— m–surjections: $\textsc{Seq}(\mathcal{V}),\ \mathcal{V} = \textsc{Set}_{\geq 1}(\mathcal{Z})$
— Set partitions: $\textsc{Set}(\textsc{Set}_{\geq 1}(\mathcal{Z}))$
— Lab. trees: $\mathcal{T} = \mathcal{Z} \star \textsc{Set}(\mathcal{T})$.

Theorem. *There exists a dictionary:*

Construction	EGF
$\mathcal{C} = \mathcal{A} + \mathcal{B}$	$C(z) = A(z) + B(z)$
$\mathcal{C} = \mathcal{A} \star \mathcal{B}$	$C(z) = A(z) \cdot B(z)$
$\mathcal{C} = \textsc{Seq}(\mathcal{A})$	$C(z) = \dfrac{1}{1 - A(z)}$
$\mathcal{C} = \textsc{Set}(\mathcal{A})$	$C(z) = \exp(A(z))$
$\mathcal{C} = \textsc{Cyc}(\mathcal{A})$	$C(z) = \log \dfrac{1}{1 - A(z)}$

$\mathcal{E}$ or $\mathbf{1}$: "neutral class" formed with element of size $0 \mapsto E(z) = 1$.
$\mathcal{Z}$: "atomic class" formed with element of size $1 \mapsto E(z) = 1$.

Product lemma:

$$\mathcal{C} = \mathcal{A} \times \mathcal{B} \qquad \Longrightarrow \qquad C(z) = A(z) \cdot B(z)$$

$\mathcal{C} = (\mathcal{A} \star \mathcal{B})$ implies $C_n = \sum_{k=0}^{n} \binom{n}{k} A_k B_{n-k}$ (# possibilities $\times$ # relabellings).

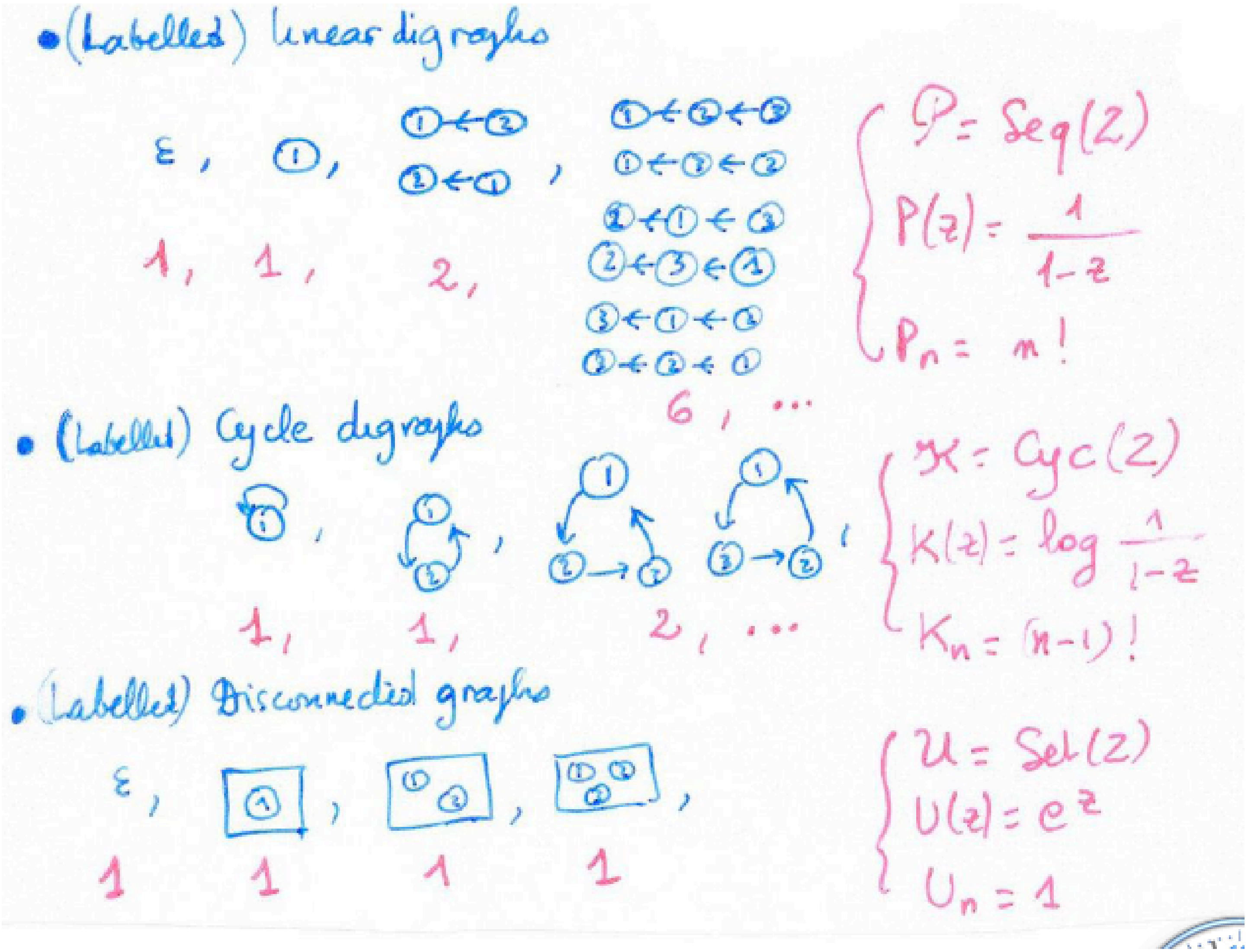
Hence $\frac{C_n}{n!} = \sum_k \frac{A_k}{k!} \cdot \frac{B_{n-k}}{(n-k)!} \rightsquigarrow C(z) = A(z) \cdot B(z)$.

SEQ: $1 + A + A^2 + \cdots = \frac{1}{1-A}$.

SET: $1 + \frac{A}{1!} + \frac{A^2}{2!} + \cdots = \exp(A)$.

CYC: $1 + \frac{A}{1} + \frac{A^2}{2} + \cdots = \log \frac{1}{1-A}$.

Example 0


• (Labelled) linear digraphs
ε , 1 , 1 , 2 , 6 , ...
P = Seq(Z)
P(z) = 1/(1−z)
P_n = n!
• (Labelled) Cycle digraphs
1 , 1 , 2 , ...
K = Cyc(Z)
K(z) = log 1/(1−z)
K_n = (n−1)!
• (Labelled) Disconnected graphs
ε , 1 1 1 1
U = Set(Z)
U(z) = e^z
U_n = 1

Example 1. Permutations and cycles:

$$\mathcal{P} = \textsc{Set}(\textsc{Cyc}(\mathcal{Z})) \quad \Longrightarrow \quad P(z) = \exp\left(\log\frac{1}{1-z}\right) = \frac{1}{1-z}.$$

Derangements (no fixed point)

$$\mathcal{D} = \textsc{Set}(\textsc{Cyc}(\mathcal{Z}) \setminus \mathcal{Z}) \quad \Longrightarrow \quad D(z) = \exp\left(\log\frac{1}{1-z} - z\right) \equiv \frac{e^{-z}}{1-z}.$$

Thus $\boxed{\frac{D_n}{n!} = 1 - \frac{1}{1!} + \frac{2}{2!} - \cdots + \frac{(-1)^n}{n!}} \sim e^{-1}$.

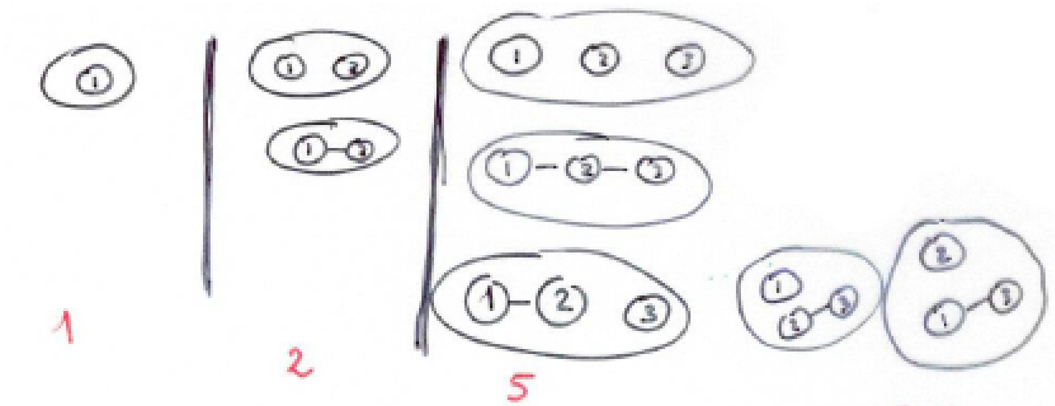
Example 2. Labelled (Cayley) trees:

$$\mathcal{T} = \mathcal{Z} \star \textsc{Set}(\mathcal{T}) \quad \Longrightarrow \quad T(z) = ze^{T(z)}.$$

Thus $\boxed{T_n = n^{n-1}}$ by Lagrange Inversion Th.

Example 3. Set partitions:

$\mathcal{B} = \textsc{Set}(\textsc{Set}_{\geq 1}(\mathcal{Z})) \implies B(z) = e^{e^z - 1}$.



Bell numbers: $\boxed{B_n = e^{-1} \sum_{k \geq 0} \frac{k^n}{k!}}$.

Example 4. Allocations to $[1\,..\,m]$:

— all: $e^{mz} \rightsquigarrow F_n = m^n$.
— surjective: $(e^z - 1)^m \rightsquigarrow$ *Stirling numbers*, $m!\left\{ {m \atop n} \right\} = \sum \binom{m}{k}(-1)^{m-k}k^n$.
— injective: $(1+z)^m \rightsquigarrow \binom{m}{n} n!$ (arrangement #).

Exercise: Birthday Problem and Coupon Collector.

$$\mathbb{E}(B) = \int_0^\infty \left(1 + \frac{t}{m}\right)^m e^{-t}\,dt, \qquad \mathbb{E}(C) = \int_0^\infty \left(e^t - (e^{t/m} - 1)^m\right) e^{-t}\,dt.$$

Multiple birthdays, multiple collections. (Cf Poissonization.)

Example 5. Mappings aka functional graphs = endofunctions of finite set.

$T = ze^T$, $K = \log(1-T)^{-1}$, $M = e^K$: $\boxed{M_n = n^n}$. $\mathbb{P}$(connected)=$O\left(\frac{1}{\sqrt{n}}\right)$.

Exercise: A binary functional graph is such that each x has either 0 or 2 preimages (cf $x^2 + a \mod p$). **Q1.** Construct; **Q2.** enumerate.

Exercise: All graphs $G(z) = 1 + \sum_{n=1}^{\infty} 2^{n(n-1)/2} z^n/n!$. **Q1.** EGF $K(z)$ of connected graphs? **Q2.** Probability of connectedness. **Q3*** Prove not constructible.

3 MULTIVARIATE GFS AND PARAMETERS

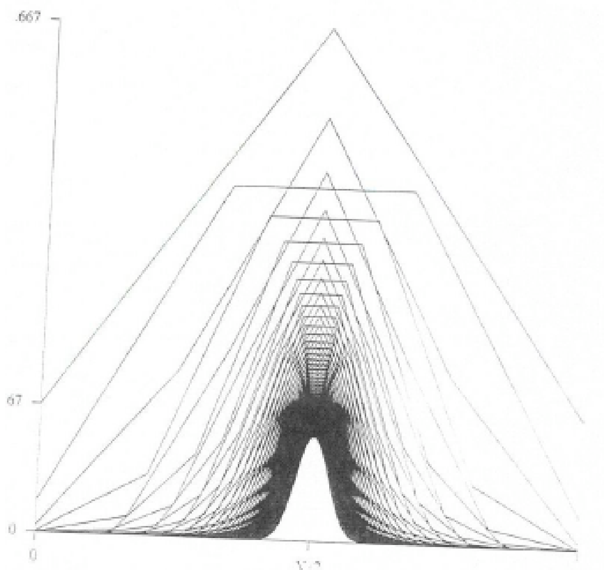
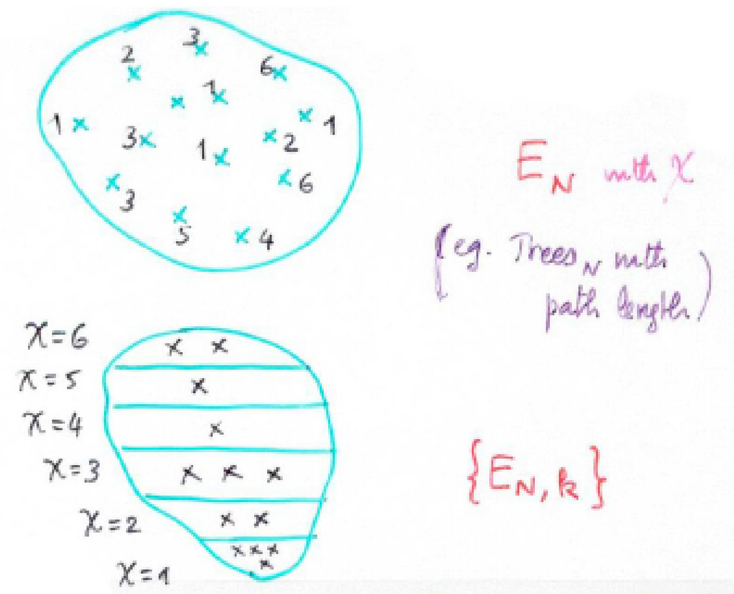


Figure 6.5 Distribution of runs, $3 \le N \le 60$ (k-axes scaled to N) (Eulerian numbers)



Bivariate GF (ordinary) $(E_{n,k}) \quad \leadsto \quad E(z,u) = \sum_{n,k} E_{n,k} u^k z^n.$

Bivariate GF (exponential) $(E_{n,k}) \quad \leadsto \quad E(z,u) = \sum_{n,k} E_{n,k} u^k \frac{z^n}{n!}.$

• BGF encodes exact distributions. hence, moments.

$$\mathbb{E}_{\mathcal{E}_n}[\chi] = \sum_k k \cdot \frac{E_{n,k}}{E_n} = \frac{1}{E_n} \operatorname{coeff}[z^n] \left. \frac{\partial}{\partial u} E(z,u) \right|_{u=1}.$$

Variance & moment of order 2: second derivative, etc.

Chebyshev inequalities: $\sigma_n/\mu_n \to 0$ implies *convergence in probability*.

Bivariate GF (ordinary) $E(z,u) = \sum_{n,k} E_{n,k} u^k z^n \equiv \sum_{\varepsilon \in \mathcal{E}} z^{|\varepsilon|} u^{\chi(\varepsilon)}$.

• BGF is reduction of combinatorial structure. Thus expect **multivariate dictionaries.**

Definition. *Parameter is inherited* if $(\mathbf{i})$ it is compatible with unions; $(\mathbf{ii})$ it is additive over products (also $\textsc{Seq}, \textsc{Set}, \textsc{Cyc}$).

meta-THM *Previous dictionaries (U/L) work verbatim!*

Proof (hint): $\mathcal{C} = \mathcal{A} \times \mathcal{B} \Longrightarrow C(z,u) = \sum_{\gamma} = \sum_{(\alpha,\beta)} = A(z,u) \cdot B(z,u)$.

Same principles as counting, but with size now extended to $\mathbb{N} \times \mathbb{N}$.

Example 1. Permutations, counting # cycles:

$$\mathcal{P} = \textsc{Set}(\textsc{Cyc}(\mathcal{Z})) \quad \Longrightarrow \quad P(z,u) = \exp\left[u\frac{z}{1} + u\frac{z^2}{2} + \cdots\right] = (1-z)^{-u}.$$

Expand and get probability GF: $\frac{1}{n!}u(u+1)\cdots(u+n-1)$; mean is $H_n \sim \log n$; standard dev. is $\sim \sqrt{\log n}$; distribution is concentrated (by Chebyshev).

singleton cycles:

$$P(z,u) = \exp\left[u\frac{z}{1} + \frac{z^2}{2} + \cdots\right] = \frac{e^{z(u-1)}}{1-z}.$$

singleton/doubleton cycles (joint): use u_1, u_2, and so on.

Example 2. Number of summands in compositions.

$$\mathcal{C} = \textsc{Seq}(\mathcal{Z} \times \textsc{Seq}(\mathcal{Z})) \quad \Longrightarrow \quad C(z,u) = \frac{1}{1 - zu/(1-z)}.$$

Example 3. Number of leaves in a general plane tree.

$$\mathcal{G} = \mathcal{Z}u + \mathcal{Z}\,\textsc{Seq}_{\geq 1}(\mathcal{Z}) \quad \Longrightarrow \quad G = z\boxed{u} + z\frac{G}{1-G}.$$

Summary: Place *marker* at appropriate places and translate with usual dictionary.

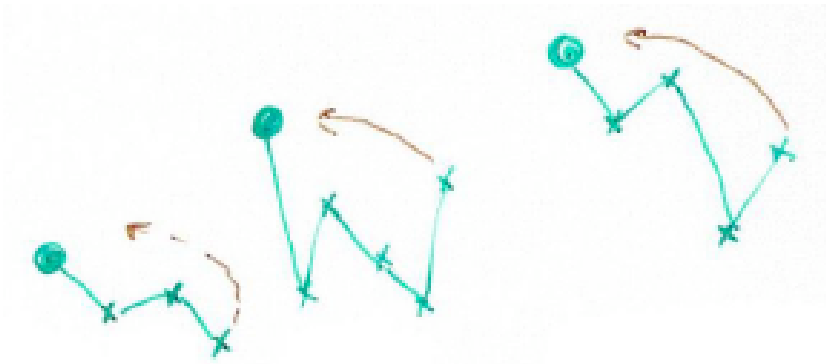
Summary. In order to *enumerate*, it suffices to find a *construction*.
— Get the OGF/EGF automatically;
— Get parameters that are traceable to constructions.

Integer compositions and partitions; words; trees; lattice paths; set partitions; allocations and functions; mappings; permutations and cycles.

Also: associate *families of special functions* to *families of combinatorial classes*.
— Regular languages $\rightsquigarrow$ Rational functions
— Tree grammars & CF languages $\rightsquigarrow$ Algebraic functions
— Simple tree families $\rightsquigarrow$ Implicit functions

Other: Constrained mappings: implicit function $\circ$ modified exp and log functions. Etc.

Exercise A. A record in a permutation is an element σ_j larger than all preceding σ_k. **Q.** Explain why the distribution of # records is the same as # cycles (on $\mathcal{P}_n$).

Hint: 

Exercise B. Throw n balls into m urns. **Q1.** The statistics of empty bins is obtained from $(e^z - 1 + u)^m$. **Q2.** Mean and variance? **Q3.** Same for bins filled with r elements. **Q4.** Relation to *Poisson*?

Santiago de Chile

DEC 2006

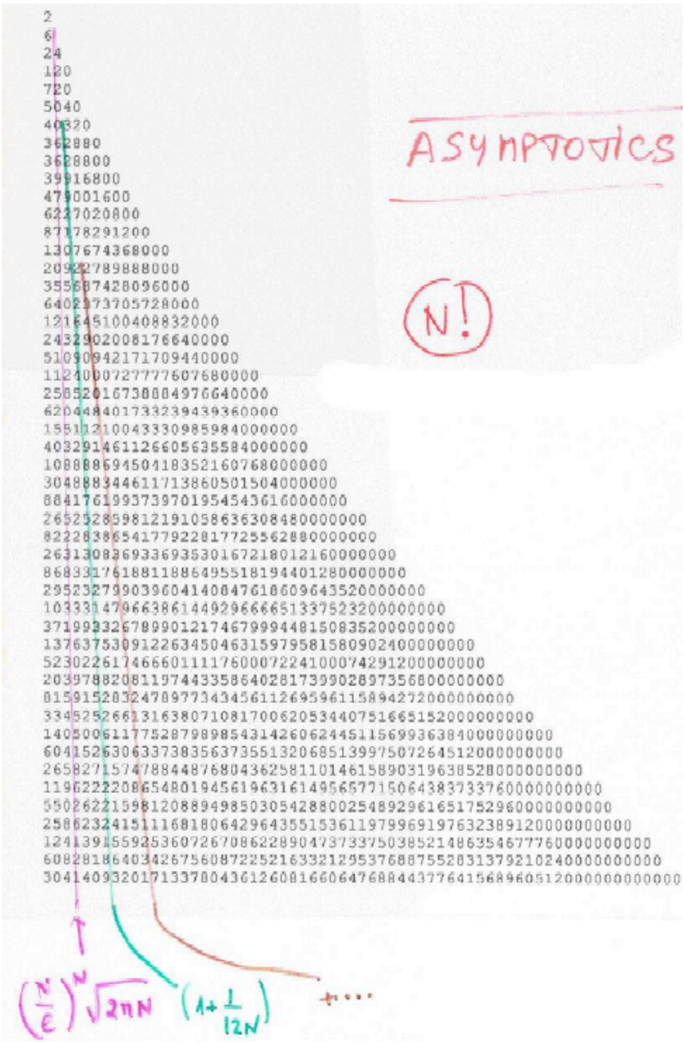
information & randomness 2006 Santiago, Chile • December 11-15

SINGULAR COMBINATORICS
B. Complex Asymptotics

Philippe Flajolet, INRIA, Rocquencourt

`http://algo.inria.fr/flajolet`

Based on *Analytic Combinatorics*, Flajolet & Sedgewick, C.U.P., 2007^{+}.



— Asymptotic analysis is often very precise.

— Can be done from generating functions directly, even if no expression for coefficients is available.

— Works for functional equations

$$U(z) = z \exp\left(U(z) + \frac{1}{2}U(z^2) + \cdots\right).$$

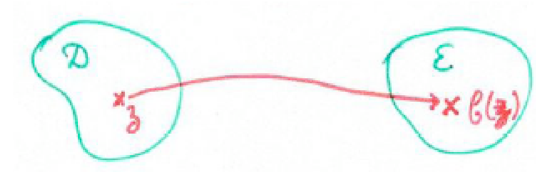
— Makes it possible to discuss universality via schemas.

4 ANALYTIC FUNCTIONS

GFs are (usually) analytic functions near 0.

- Analytic aka holomorphic functions
- Meromorphic functions
- Integrals and residues
- Singularities and exponential growth orders

Let $f(z)$ be defined from D (open connected set) to E:



Definition. • $f(z)$ is analytic at z_0 iff *locally*: $\boxed{f(z) = \sum_{n \geq 0} c_n (z - z_0)^n}$.

• $f(z)$ is *complex differentiable* iff

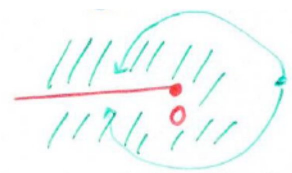
$$\exists \boxed{\lim_{h \to 0,\ h \in \mathbb{C}} \frac{f(z_0 + h) - f(z_0)}{h}} =: f'(z_0) \equiv \frac{d}{dz} f(z) \Big|_{z = z_0}.$$

$\rightsquigarrow f$ analytic/ differentiable in Ω , etc.

Theorem. Equivalence between the two notions!

Combinatorialists love power series; analysts love differentiability!

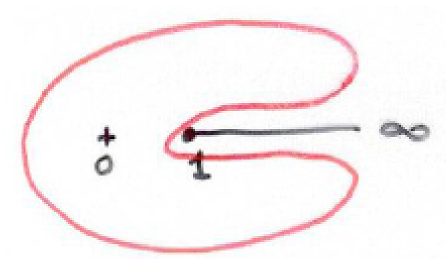
$\dfrac{\Delta f}{\Delta z}$ gives closure under $+, -, \times, \div$, composition, inversion, &c.

Examples. The function $\sqrt{z}$, such that $\sqrt{\rho e^{i\theta}} = \sqrt{\rho} \cdot e^{i\theta/2}$, can only be made continuous in 

.

— Same for $\log z = \log \rho + i\theta$.

— Exponential function $\exp(z)$ is entire.

— $\dfrac{e^z}{\sqrt{1-z}}$ is analytic in 

— Catalan GF $\frac{1-\sqrt{1-4z}}{2z}$ is analytic in slit plane $\mathbb{C} \setminus [\frac{1}{4}, +\infty[$.

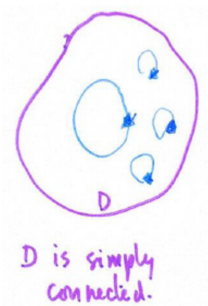
— Rational GF is analytic except at poles.

Integration and residues

Theorem. Let f be analytic in Ω and γ be contractible to a single point in Ω. Then

$$\int_\gamma f(z)\,dz = 0.$$

In particular $\int_A^B f(z)\,dz$ does not depend on path.



Definition. $g(z)$ is *meromorphic* in Ω iff near any z_0, one has $g(z) = \frac{A(z)}{B(z)}$, with A, B analytic.

A point z_0 such that $B(z_0) = 0$ is a *pole*. Its *order* is the multiplicity of z_0 as root of B (assume $A(z_0) \neq 0$).

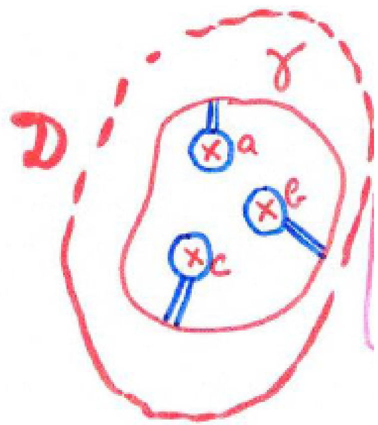
Pole of order m: $g(z) = \dfrac{c_{-m}}{(z - z_0)^m} + \cdots + \dfrac{c_{-1}}{(z - z_0)} + c + 0 + \cdots$.

c_{-1} is called *residue* of $g(z)$ at z_0.

Cauchy's Residue Theorem. If $f(z)$ has poles, then

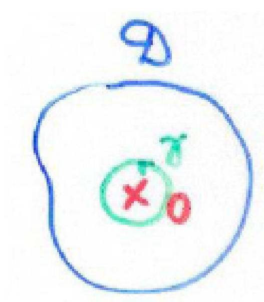
$$\frac{1}{2i\pi}\int_\gamma f(z)\,dz = \sum \text{Residues}\,.$$

Proof: local integration +

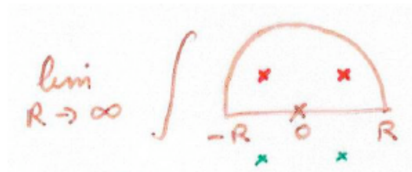


Cauchy's Coefficient Theorem. $$\text{coeff}[z^n]\,f(z) = \frac{1}{2i\pi}\int_\gamma f(z)\,\frac{dz}{z^{n+1}}$$

Proof: by residues:



Residues: local versus global

- Computing integrals: $\int_{-\infty}^{+\infty} \frac{dx}{1+x^4}$ =  $= \frac{\pi}{\sqrt{2}}$

By only considering *local properties* at $\zeta = e^{i\pi/4}, e^{3i\pi/4}$.

- Estimating coefficients: $d_n := \mathbb{P}[\text{derangement}]$ over $\mathcal{P}_n$.

$$d_n = [z^n]\frac{e^{-z}}{1-z} = \frac{1}{2i\pi}\int_{|z|=1/2} \frac{e^{-z}}{1-z}\frac{dz}{z^{n+1}}.$$

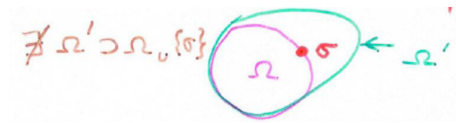
Evaluate instead on $|z| = 2$:

$$\begin{aligned} J_n &= \frac{1}{2i\pi}\int_{|z|=2} \frac{e^{-z}}{1-z}\frac{dz}{z^{n+1}} = O(2^{-n}) \\ &= \text{Res}_{z=0} + \text{Res}_{z=1} = d_n - e^{-1}. \end{aligned}$$

Thus: $\boxed{d_n = e^{-1} + O(2^{-n})}$.

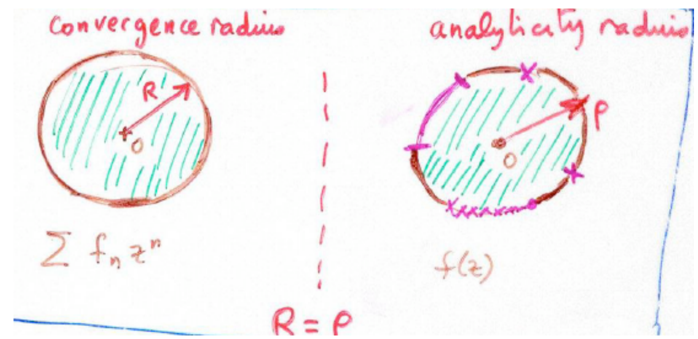
Exercise: Double derangement: $[z^n]e^{-z-z^2/2}/(1-z)$. Generalize!

Singularities.

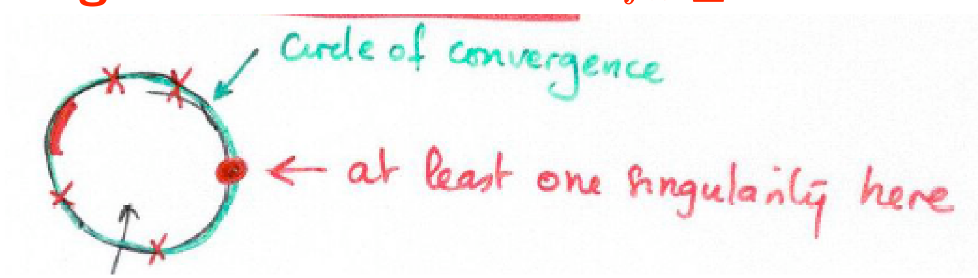
• $f(z)$ has a singularity at border point σ iff 

.

Theorem. A series always has at least one singularity on its circle of convergence (but none inside).

Convergence radius $\equiv$ Analyticity radius: 

Pringsheim's Theorem. If $f_n \geq 0$, one such singulariy is positive.



Exponential growth of coefficents.

If $f(z)$ has radius exactly R, then $\forall \epsilon > 0$:

$$f_n(R-\epsilon)^n \to 0; \qquad f_n(R+\epsilon)^n \text{ is unbounded.}$$

That is $\limsup |f_n|^{1/n} = \frac{1}{R}$, or

$$f_n = R^{-n}\vartheta(n), \quad \text{where } \vartheta(n) \text{ is "subexponental".}$$

Also write $\boxed{f_n \bowtie R^{-n}}$ with R := distance to nearest sing(s).

Find exponential growth by just "looking" at GF!!

Examples (singularities and growth)

- Binary words: $W(z) = \frac{1}{1-2z} \leadsto W_n \bowtie 2^n$.

- Derangements: $D(z) = \frac{e^{-z}}{1-z} \leadsto \frac{D_n}{n!} \bowtie 1^n$.

- General trees: $G(z) = \frac{1}{2}\left(1 - \sqrt{1-4z}\right) \leadsto G_n \bowtie 4^n$. By Stirling: $G_n \sim \frac{4^{n-1}}{\sqrt{\pi n^3}}$.

- Unary-binary trees: $U = z(1+U+U^2), U = \frac{1}{2z}\left(1 - z - \sqrt{1-2z-3z^2}\right)$, so that singularities are at $z = -1, \frac{1}{3}$ and $U_n \bowtie 3^n$.

Exponential order is computable(almost) automatically for GFs given by explicit expressions.
E.g.: $\rho(f+g) = \min(\rho(f), \rho(g))$; $\rho\left(\frac{1}{1-f}\right) = \min(\rho(f), \{z \ / \ f(z) = 1\})$, etc.

5 RATIONAL AND MEROMORHIC FNS

Find subexponential factors in

$$f_n \bowtie R^{-n}, \qquad \text{meaning} \quad f_n = R^{-n}\vartheta(n),$$

where $\vartheta(n)$ is like n^α, $(\log n)^\beta$, $e^{\sqrt{n}}$, etc.

Here: simple case of Rat & Mero.

Coefficients of rational functions

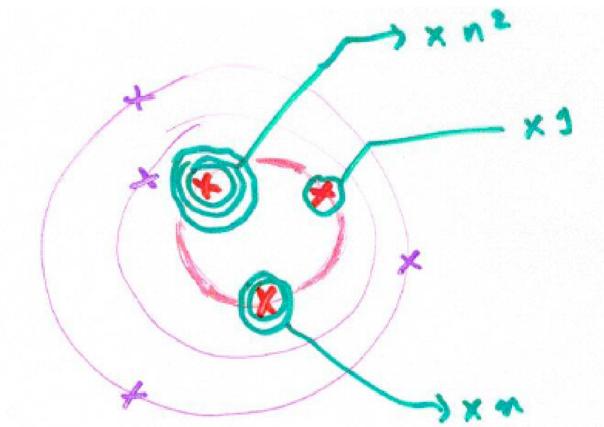
Theorem. Each pole ζ with multiplicity r contributes to coefficients a term

$$\zeta^{-n} P(n),$$

where $P(n)$ is a polynomial of degree $r-1$.

Proof. $[z^n]\dfrac{1}{(z-\zeta)^r} = (-\zeta)^{-r}\dbinom{n+r-1}{r-1}\zeta^{-m}.$

Poles are arranged in order of increasing modulus. Dominant ones matter for exponential growth. Multiplicities give polynomial factors.



Example 1. Denumerants.

- In how many ways can one give change with $1, 2, 5$c coins?

$$D_n = [z^n] \frac{1}{(1-z)(1-z^2)(1-z^5)}.$$

One layer. Poles at $1, \pm 1, \zeta^5 = 1$.
Observe the "transfer" $D(z) \sim \frac{1}{10}(1-z)^{-3}$ implies $D_n \sim n^2/20$.

- General case Ω–denominations, $m = \|\Omega\|$. Then (Schur)

$$D_n \sim \frac{n^{m-1}}{(m-1)!} \prod_{\omega \in \Omega} \frac{1}{\omega}.$$

Example 2. Longest b-runs in strings. (cf Feller)

$$bbb\ \boxed{\mathbf{a}bb}\ \boxed{\mathbf{a}b}\ \boxed{\mathbf{a}}\ \boxed{\mathbf{a}bbbb}$$

$$\textsc{Seq}_{<m}(b) \times \textsc{Seq}(\mathbf{a}\ \textsc{Seq}_{<m}(b))$$
$$\frac{1-z^m}{1-z} \times \frac{1}{1-z\frac{1-z^m}{1-z}} = \frac{1-z}{1-2z+z^{m+1}}.$$

— Dominant pole is near $\frac{1}{2}$: $\rho_m \approx \frac{1}{2}(1+2^{-m-1})$.

— Dominant pole is separated by $|z| = \frac{3}{2}$;error is exp. small.

— Uniform estimates are obtained. Get

$$\mathbb{P}\,(\text{longest } b\text{-run} < m) \approx \left(\frac{1}{2\rho_m}\right)^n \approx e^{-n/2^{m+1}}.$$

Threshold near $\log_2 n$.

Arbitrary patterns: similar with *correlation polynomials* of Guibas–Odlyzko.
Quantitative normality of strings, Borges' Theorem ,etc.

Coefficients of meromorphic functions

Assumption: $g(z)$ is meromorphic in $|z| < R$ and analytic on $|z| = R$.

Theorem. Each pole ζ with multiplicity r contributes to coefficients a term

$$\zeta^{-n} P(n),$$

where $P(n)$ is a polynomial of degree $r-1$. Error term is $O(R^{-n})$.

Proof. (i) Subtracted sngularities. Let $h(z)$ gather contributions of poles. Then $g(z) - h(z)$ is analytic in $|z| \leq R$. Use Cauchy with trivial bounds.

(ii) Estimate $\int g$ by residues.

Example 3. Derangements.

$$\mathcal{D} = \text{SET}(\text{CYC}_{\geq 2}(Z)) \qquad \Longrightarrow \qquad D(z) = e^{-z} 1 - z.$$

Get simple pole at $z = 1$ so that $\frac{1}{n!} D_n = [z^n] \frac{e^{-1}}{1-z} + O(2^{-n}) = e^{-1} + O(2^{-n})$.

Generalized derangement: all cycles of length $> r$:

$$\frac{1}{n!} D_n^{\star} \sim e^{-H_r}, \qquad H_r = 1 + \frac{1}{2} + \cdots + \frac{1}{r}.$$

Example 4. Paths-in-graphs models.

Encapsulates finite automata and finite Markov chains. GFs are rational.

If the graph Γ is *strongly connected* and *aperiodic*, then there is unicity and simplicity of dominant pole ($\ll$ Perron-Frobenius): $f_n \sim c\rho^{-n}$.

Generalized patterns in random strings (F, Nicodème, Régnier, Salvy, Szpankowski, Vallée, &c).

Example 5. Surjections and Supercritical $\textsc{Seq}$ Schema.

Random surjection $\equiv$ ordered partition (pref. arrangement)

$$\mathcal{R} = \textsc{Seq}(\textsc{Set}_{\geq 1}(\mathcal{Z})) \qquad \Longrightarrow R(z) = \frac{1}{2 - e^z}.$$

Pole at $\zeta = \log 2$; subdominant ones at $\zeta = \log 2 \pm 2ik\pi$, etc.

$$\frac{R_n}{n!} \sim c(\log 2)^{-n}.$$

Also, mean number of blocks via $\dfrac{1}{1 - u(e^z - 1)}$ is $O(n)$. There is concentration, etc.

Any supercritical sequence should similarly behave $\leadsto$ schema.

6 SINGULARITY ANALYSIS

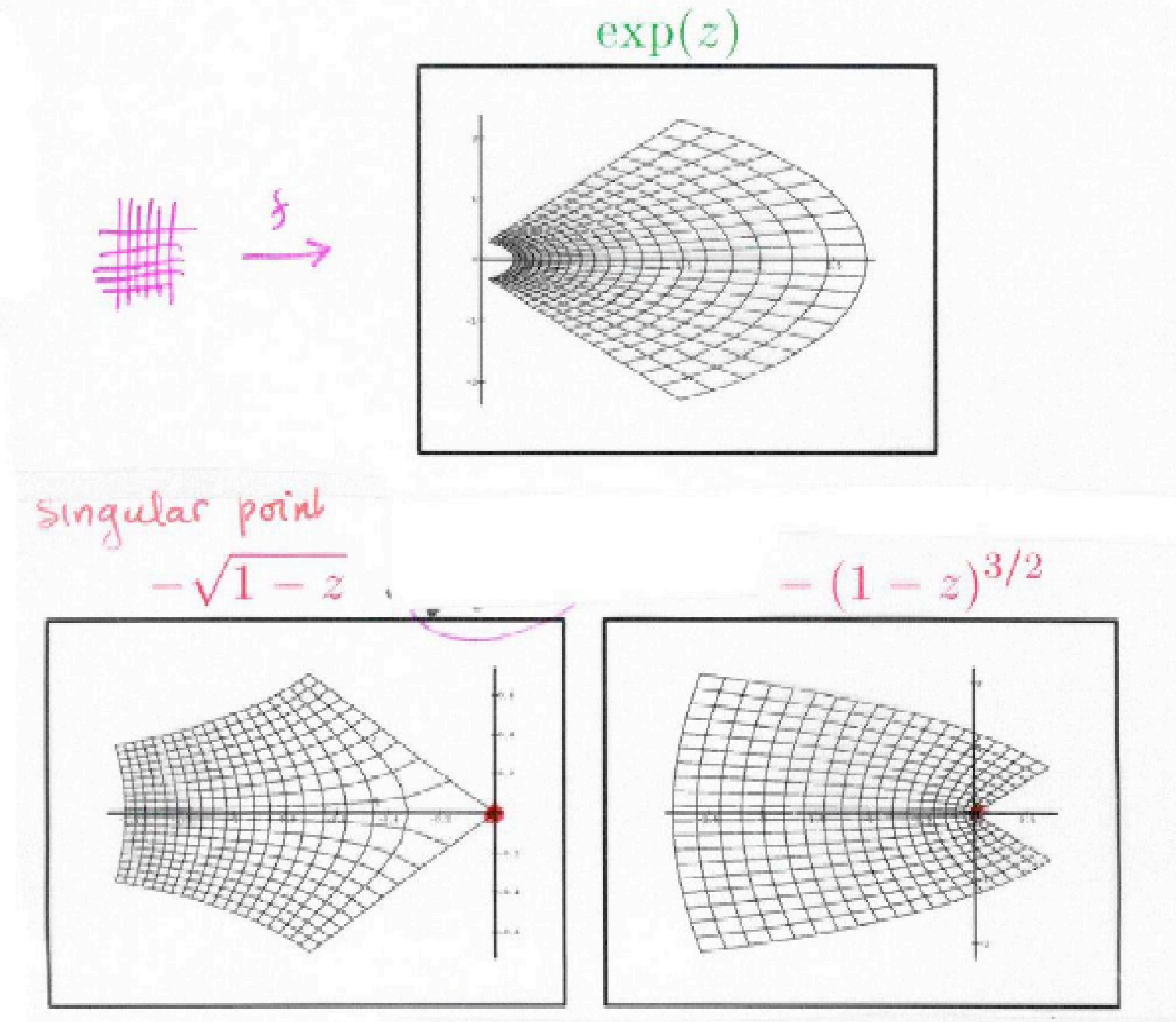
- Singularities more general than poles.
- Subexponential factors more general than polynomials:

$$f_n \sim R^{-n} \vartheta(n),$$

with $\vartheta(n)$ of the form $n^\alpha (\log n)^\beta$.

Note: May assume singularity at 1 by scaling $[z^n]f(\lambda z) = \lambda^n [z^n] f(z)$.

Regular point $f(z) \approx f(z_0) + f'(z_0)(z - z_0)$



Coefficients: $n^{-3/2}$ $n^{-5/2}$

From functions to coefficients:

$\frac{1}{(1-z)^2}$	$\longrightarrow$	$n+1$	$\sim$	n
$\frac{1}{1-z}\log\frac{1}{1-z}$	$\longrightarrow$	$H_n \equiv \frac{1}{1} + \ldots + \frac{1}{n}$	$\sim$	$\log n$
$\frac{1}{1-z}$	$\longrightarrow$	1	$\sim$	1
$\frac{1}{\sqrt{1-z}}$	$\longrightarrow$	$\frac{1}{2^{2n}}\binom{2n}{n}$	$\sim$	$\frac{1}{\sqrt{\pi n}}$

$$\begin{cases} \text{Location of sing's :} & \text{Exponential factor} \quad \rho^{-n} \\ \text{Nature of sing's :} & \text{“Polynomial” factor} \quad \vartheta(n) \end{cases}$$

Principles of Singularity Analysis

Larger functions tend to have larger coefficients.

— Establish this for basic scales $(1-z)^{-\alpha}$. Enrich with $\log$'s, $\log\log$'s, etc.

— Prove transfer theorems. If f "resembles" g via $O(\cdot)$, $o(\cdot)$, then f_n resembles g_n.

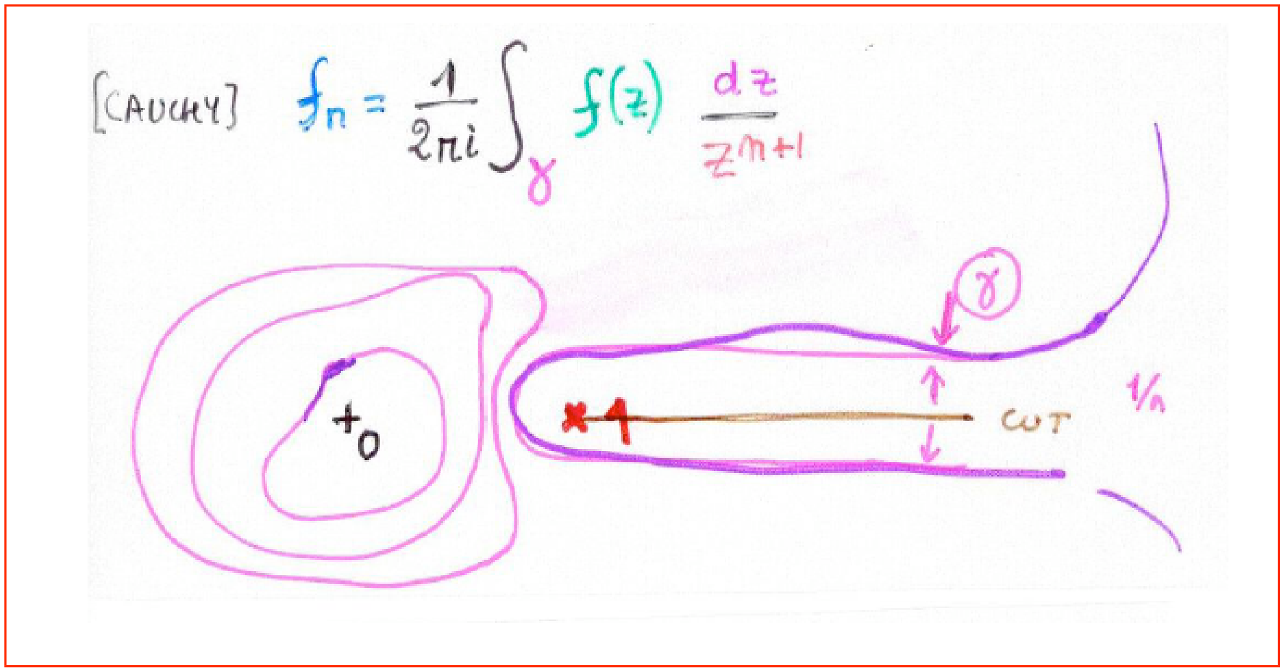
Theorem 1. Coefficients of basic scale:

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

Also: full expansion, log's log-log's, etc.

Gamma function: $\Gamma(s) := \int_0^\infty e^{-t} t^{s-1}\, dt$, with analytic continuation by $\Gamma(s+1) = s\Gamma(s)$.

Idea:



Theorem 1. *Basic scale translates:*

$$\sigma_{\alpha,\beta}(z) := (1-z)^{-\alpha}\left(\tfrac{1}{z}\log\tfrac{1}{1-z}\right)^{\beta}$$
$$\Longrightarrow \quad [z^n]\sigma_{\alpha,\beta} \underset{n\to\infty}{\sim} \frac{n^{\alpha-1}}{\Gamma(\alpha)}(\log n)^{\beta}.$$

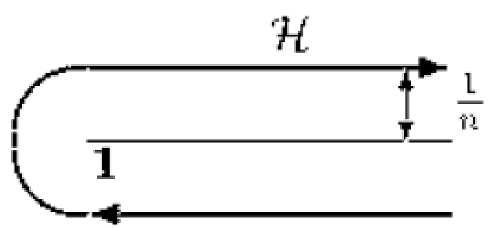
PROOF. Cauchy's coefficient integral, $f(z) = (1-z)^{-\alpha}$

$$[z^n]f(z) = \frac{1}{2i\pi}\int_{\gamma} f(z)\,\frac{dz}{z^{n+1}}$$

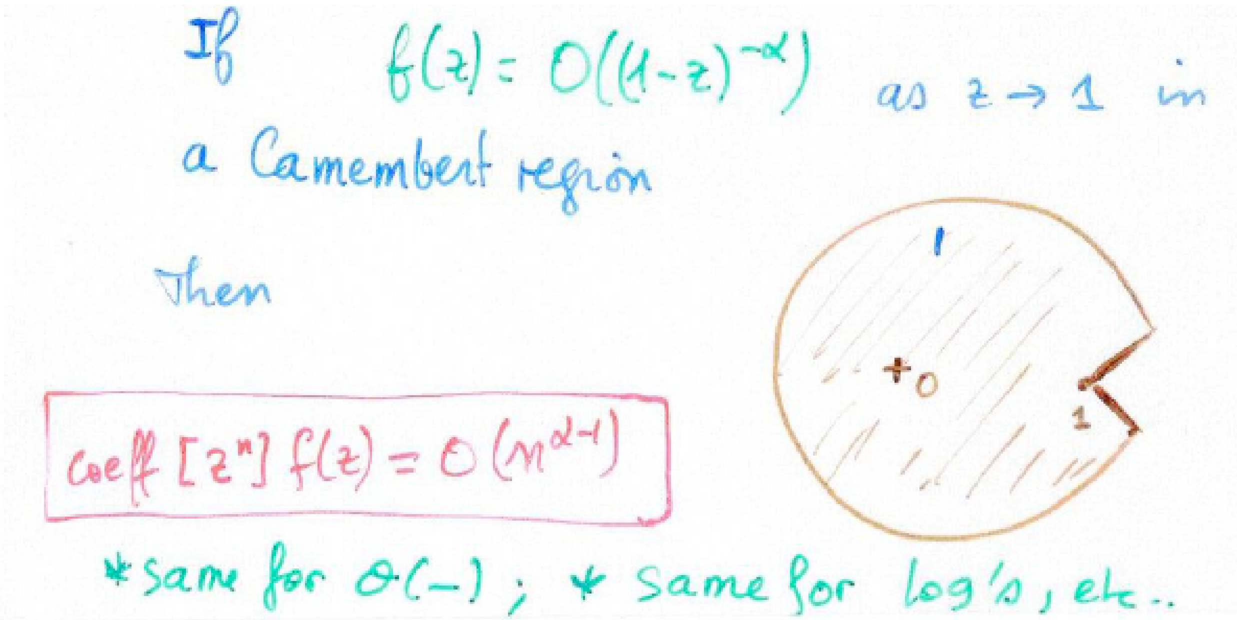
$$\Downarrow \quad (z = 1 + \frac{t}{n}) \quad \Downarrow$$

$$\frac{1}{2i\pi}\int_{\mathcal{H}}\left(-\frac{t}{n}\right)^{-\alpha} e^{-t}\,\frac{dt}{n}$$

$$n^{\alpha-1} \times \tfrac{1}{\Gamma(\alpha)}.$$

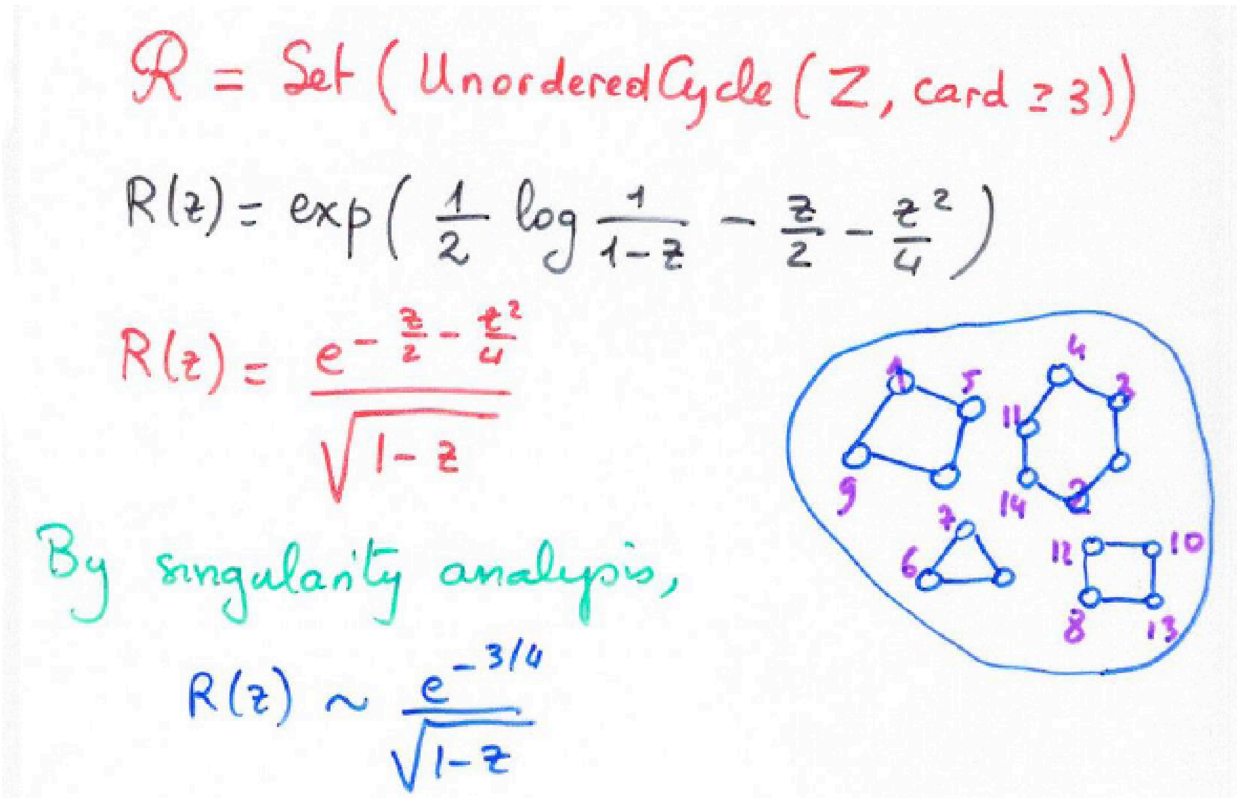


Theorem 2. Transfer of asymptotic properties.



Proof: similarly by Hankel contours.

Example 1. 2–regular graphs.



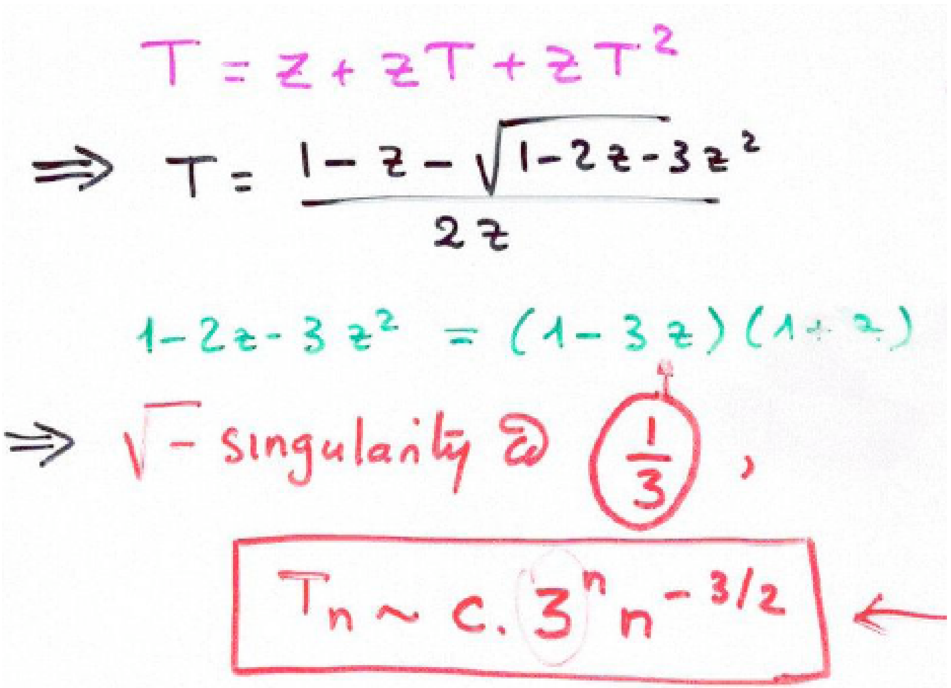
$$\frac{R_n}{n!} \sim e^{-3/4}\sqrt{\pi n}.$$

Comtet's clouds. Also full asymptotics.

Example 2. Some trees.

• Catalan trees have GF $\frac{1}{2}(1-\sqrt{1-4z}) \rightsquigarrow c\frac{4^n}{\sqrt{\pi n^3}}$.

• Unary binary trees.



In fact: *universality* of $n^{-3/2}$ law (later).

Example 3. Cycles in Perms.

Mean number of cycles in a random perm is coeff$[z^n]$ in

$$M(z) = \frac{\partial}{\partial u} \exp\left(u \log \frac{1}{1-z}\right)\Bigg|_{u\to 1} = \frac{1}{1-z} \log \frac{1}{1-z}.$$

Thus $\boxed{[z^n]M(z) \sim \log n}$.

Exercise: Holds for perms with finitely many excluded cycle lengths.

In fact: *universality* for the “exp-log” schema.

Closures

Theorem 3. *Generalized polylogarithms*

$$\mathrm{Li}_{\alpha,k} := \sum (\log n)^k n^{-\alpha} z^n$$

are of S.A.-type.

Theorem 4. *Functions of S.A.-type are closed under integration and differentiation.*

Theorem 5. *Functions of S.A.-type are closed under Hadamard product*

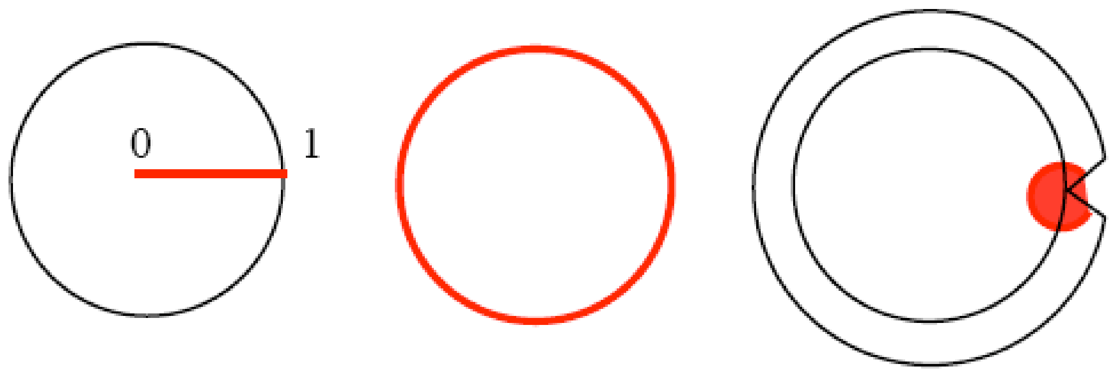
$$f(z) \odot g(z) := \sum_n (f_n g_n) z^n.$$

(F) (Fill-F-Kapur 2005).

Generating Function $\rightsquigarrow$ Coefficients

Solving a "Tauberian" problem

$\mathbb{R}$eal–Tauberian	Darboux-Pólya	Singularity An.



(large $\Longrightarrow$ large)	(smooth $\Longrightarrow$ small)	(Full mappings)

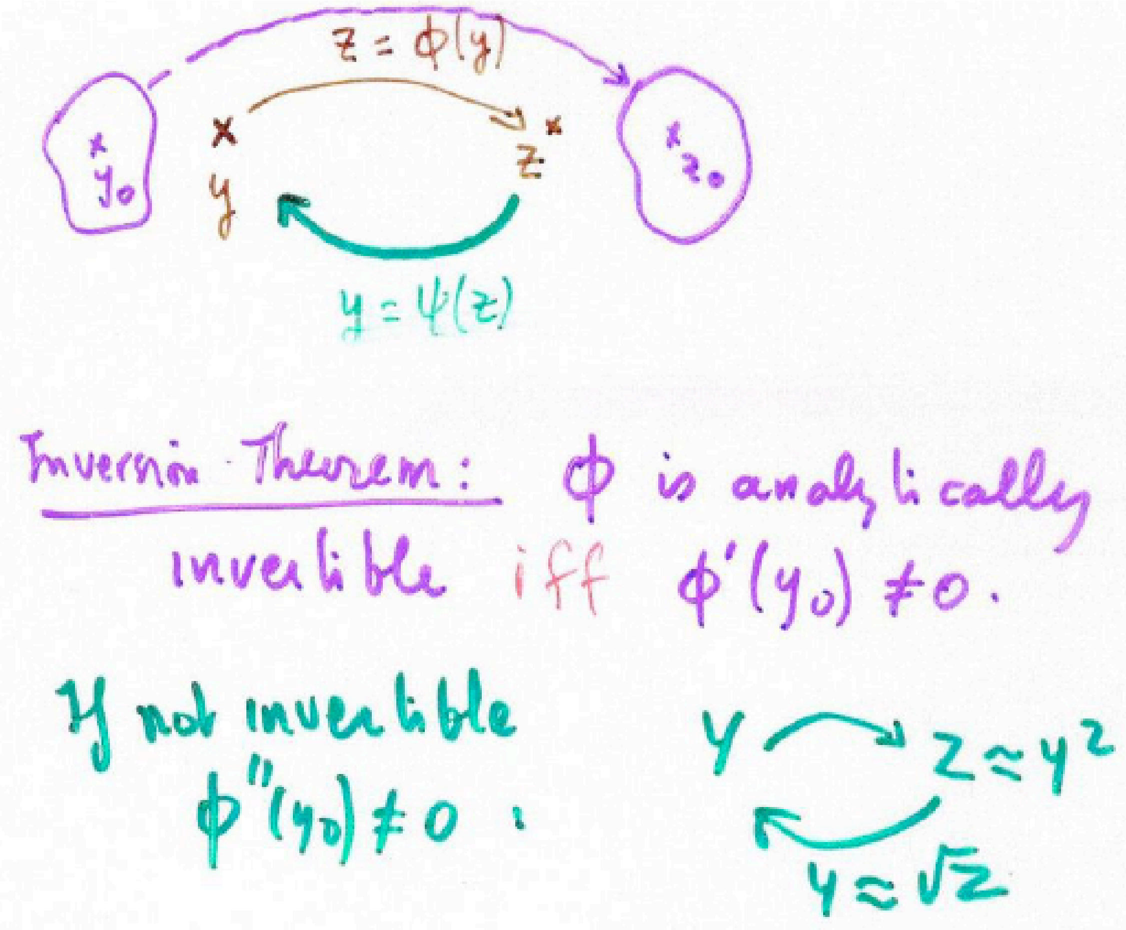
+ Singularity analysis preserves uniformity $\rightsquigarrow$ distributions.

7 APPLICATIONS OF SING. ANA.

Focus on recursive structures including trees, mappings.

- Universality of $\sqrt{\ }$–law for generating functions;
- Universality of $\rho^{-n} n^{-3/2}$–law for counts;
- Universal behaviour for major parameters (e.g., height).

Inversion:



Square-root singularity is expected for inverse functions.

Theorem 1. Let ϕ have nonnegative coeffs and be entire. Then the function that solves

$$Y(z) = z\phi(Y(z))$$

has a square-root singularity, so that

$$[z^n]Y(z) \sim C\rho^{-n}n^{-3/2}.$$

— Characteristic equation (singular value of Y) is $\tau : \frac{d}{dy}\frac{y}{\phi(y)} = 0$, i.e., $\tau\phi'(\tau) - \phi(\tau) = 0$. Then $\rho = \frac{\tau}{\phi(\tau)}$. All is computable.

— $\sqrt{\ }$–singularity propagates via suitable compositions, so that parameters can be analysed.

— Phenomena are robust.

Example 1. Cayley trees. $T = ze^T$ or $z = Te^{-T}$ is not invertible if $\frac{d}{dT}(Te^{-T}) \equiv (1-T)e^{-T} = 0$, that is, $T = 1, z = e^{-1}$. Find:

$$T(z) \underset{z \to e^{-1}}{=} 1 - \sqrt{2}\sqrt{1 - ez} + O((1 - ez)).$$

Implies $[z^n]T(z) \sim \frac{e^n}{\sqrt{2\pi n^3}}$; we rederive Stirling's f. (since $T_n = n^{n-1}$ by Lagrange).

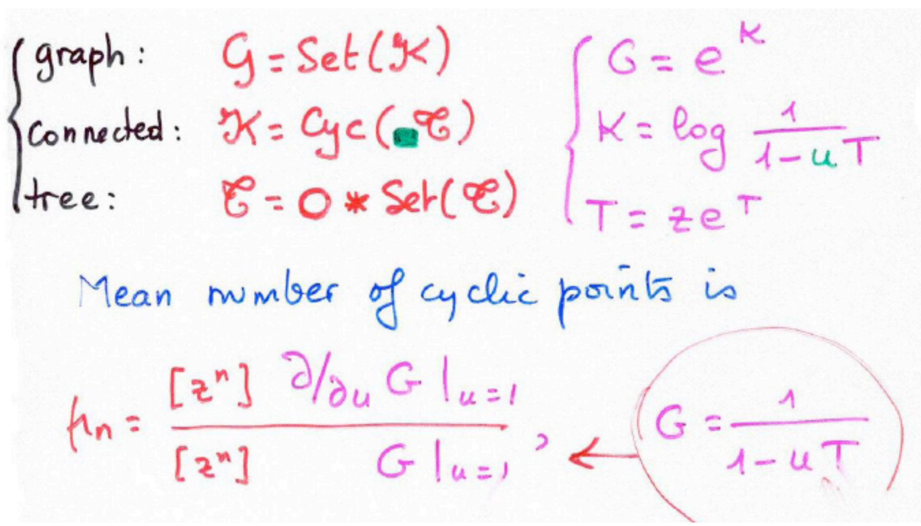
Example 2. Unlabelled trees. Recall

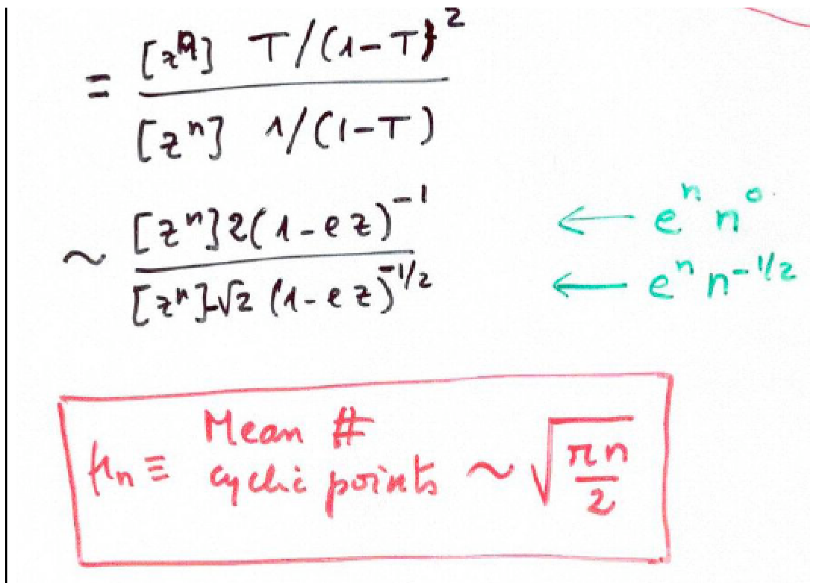
$$U(z) = ze^{U(z) + \frac{1}{2}U(z^2) + \cdots}.$$

Express as T composed with an analytic function and get SQRT sing: $U = \zeta e^U$, where $\zeta := z \exp(\frac{1}{2}U(z^2) + \cdots)$.

Height is universally $O(\sqrt{n})$ wth local and integral limit laws (of theta type). Similarly for width (Marckert et al.). Leaves are universally normally distributed, etc.

Example 3. Mappings (cyclic points).

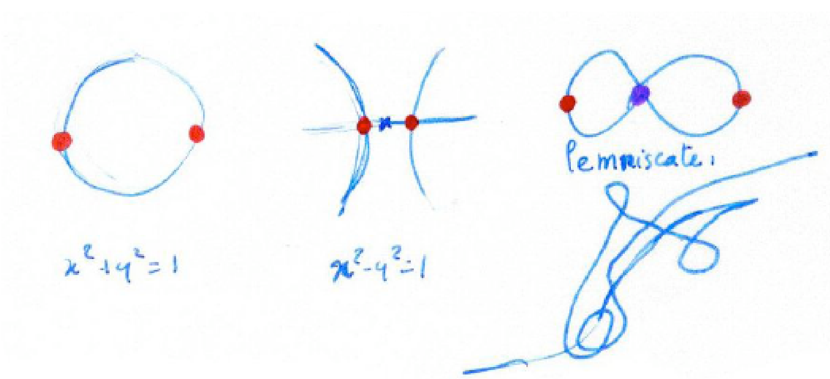


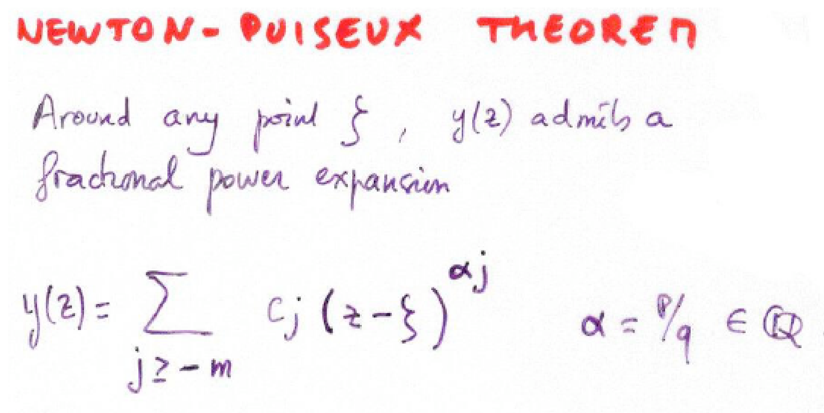


Develop a theory of degree-constrained mappings: (Arney-Bender), (F.-Odlyzko).

Algebraic functions

Singularity analysis applies to any algebraic function





Algebraic function $\Longrightarrow$ Fractional exponents @ singularities.

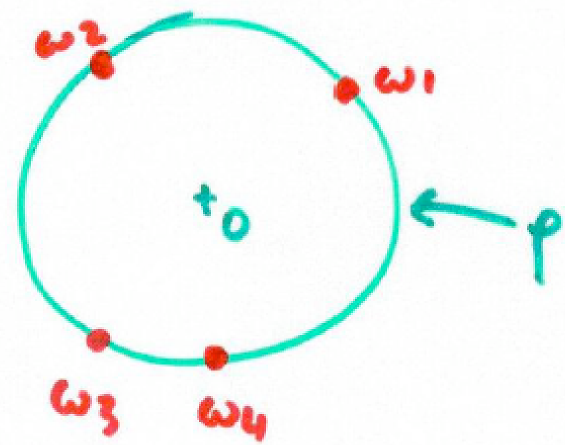
Define an "algebraic" element to be of the form

$$\omega^{-n} \cdot \sum_{j \ge -r} d_j \, n^{-j\beta} \qquad (\beta \in \mathbb{Q})$$

Theorem: If $y(z)$ is an algebraic function, then there exists a finite collection of algebraic elements A_1 (at ω_1), A_2 (at ω_2), ..., A_s (at ω_s) s.t.

$$y_n = A_1 + \cdots + A_s + O(\xi^{-n})$$

$$|\omega_1| = |\omega_2| \cdots = |\omega_s| = \rho \; ; \quad \xi > \rho .$$

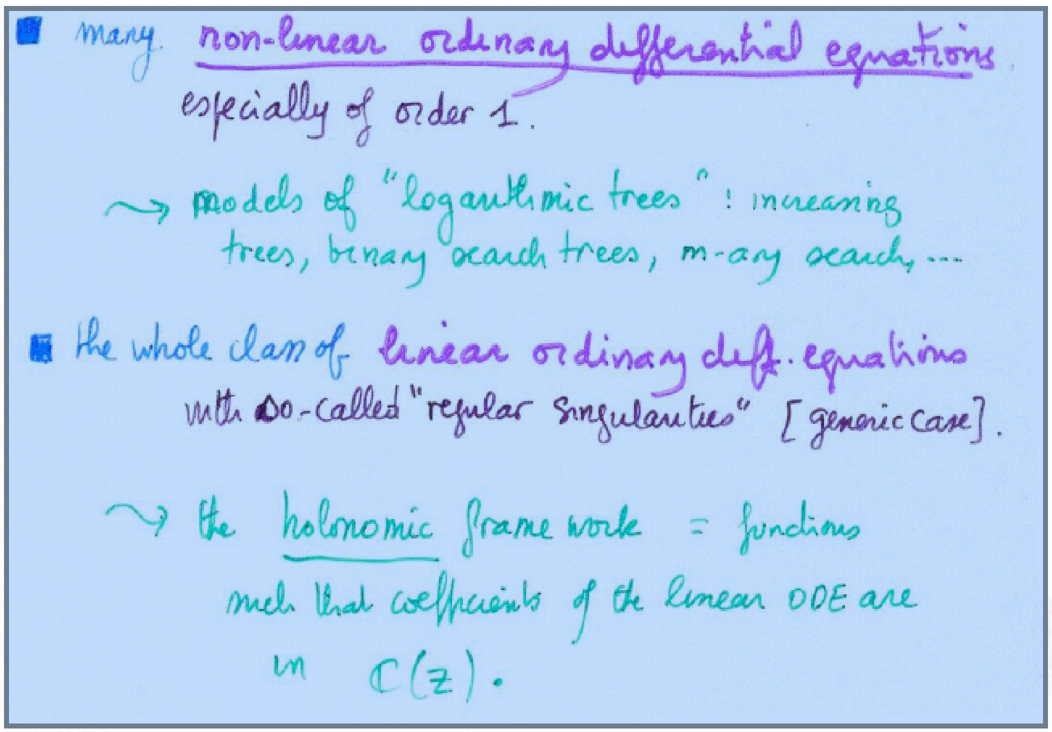


Singularity analysis applies to

- Trees with a finite set of node degrees (we know already $\sqrt{\ }$- singularity, $\rho^{-n} n^{-3/2}$).
- Excursions defined by a discrete set of steps Ω that is finite [Banderier, Fg 2002]
- MAPS = graphs embedded into the plane

→ Gimenez - Noy : counting of planar graphs by gen. function + complex analysis.

Singularity analysis applies to



- many non-linear ordinary differential equations especially of order 1.

  ⤳ models of "logarithmic trees": increasing trees, binary search trees, m-ary search, ...

- the whole class of linear ordinary diff. equations with so-called "regular singularities" [generic case].

  ⤳ the holonomic framework = functions such that coefficients of the linear ODE are in $\mathbb{C}(z)$.

• "Holonomic" functions. Defined as solutions of linear ODE's with coeffs in $\mathbb{C}(z)$ [Zeilberger] $\equiv$ $\mathcal{D}$-finite.

$$\mathcal{L}[f(z)] = 0, \qquad \mathcal{L} \in \mathbb{C}(z)[\partial_z].$$

• Stanley, Zeilberger, Gessel: Young tableaux and permutation statistics; regular graphs, constrained matrices, etc.

Fuchsian case (or "regular" singularity) $(Z^{\beta} \log^k Z)$:

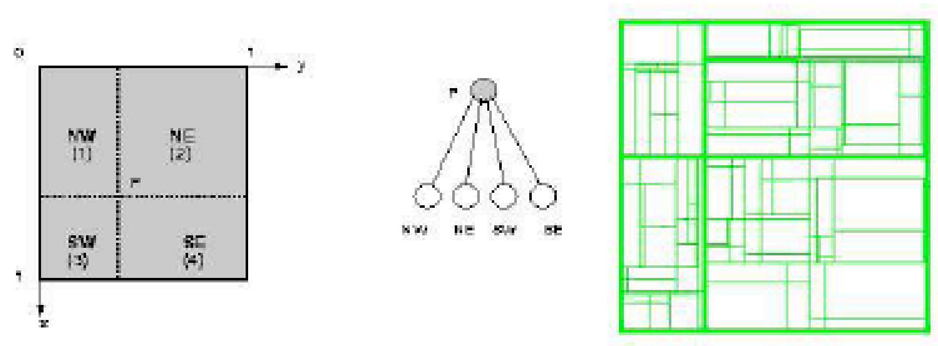
$$[z^n]f(z) \approx \sum \omega^n n^{\beta} (\log n)^k, \qquad \omega, \beta \in \overline{\mathbb{Q}}, \quad k \in \mathbb{Z}_{\geq 0}.$$

S.A. applies automatically to classical classification.

Asymptotics of coeff is decidable

— general case: modulo oracle for connection problem;

— strictly positive case: "usually" OKay.

QTrees: 

<u>Example 6</u>. *Quadtrees—Partial Match* [FGPR'92]

Divide-and-conquer recurrence with coeff. in $\mathbb{Q}(n)$

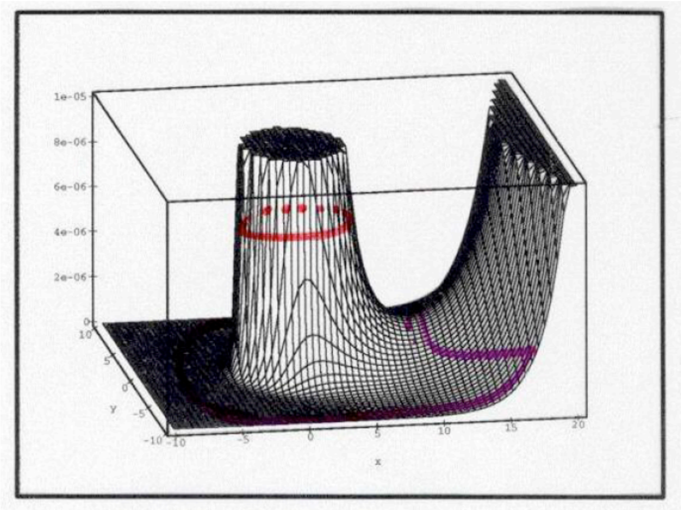
Fuchsian equation of order d (dimension) for GF

$$Q_n^{(d=2)} \asymp n^{(\sqrt{17}-3)/2}.$$

E.g., $d = 2$: Hypergeom ${}_2F_1$ with algebraic arguments.

Extended by Hwang et al. Cf also Hwang's *Cauchy ODE* cases.
Panholzer-Prodinger+Martinez, ...

8 SADDLE POINT METHODS



• For functions with violent growth at singularities, including entire functions.

$$[z^n]f(z) = \frac{1}{2i\pi} \oint f(z) \frac{dz}{z^{n+1}}.$$

Integer partitions, set partitions, involutions, . . .

Santiago de Chile

DEC 2006

information & randomness 2006 Santiago, Chile • December 11-15

SINGULAR COMBINATORICS

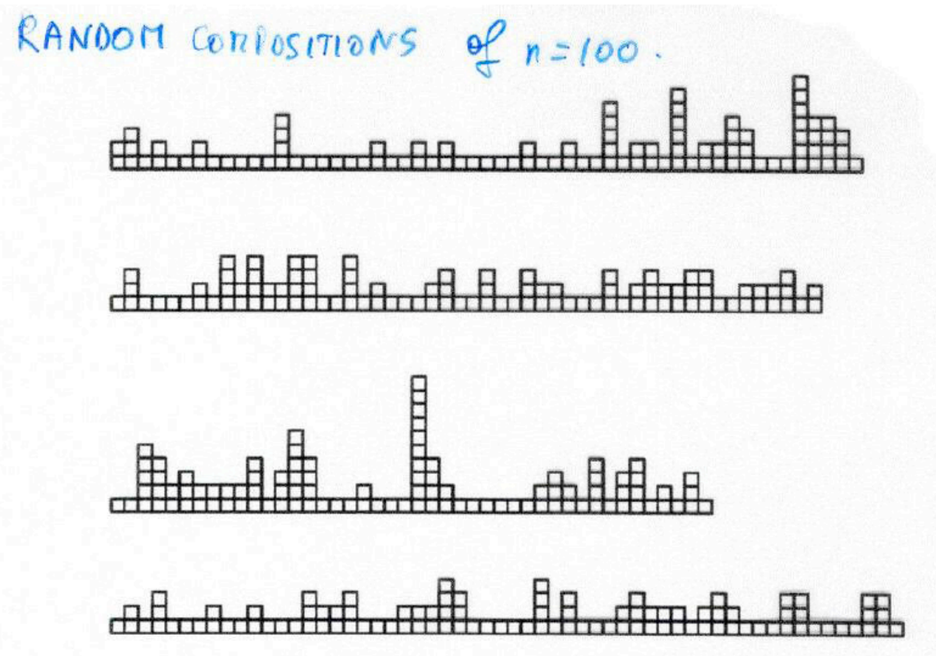
C. Random Structures

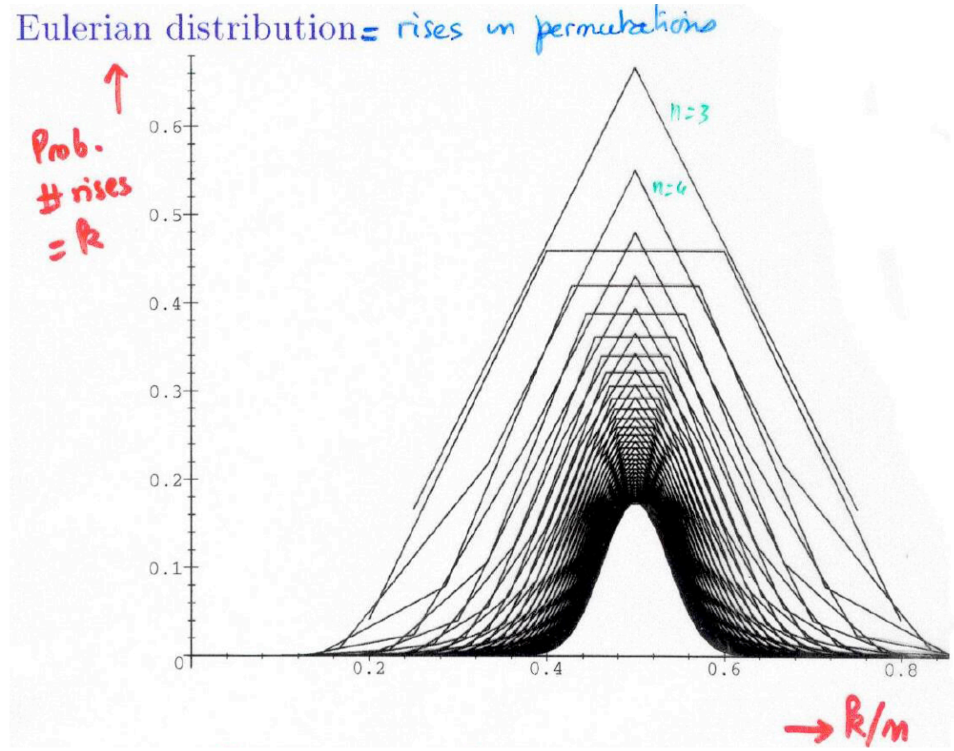
Philippe Flajolet, INRIA, Rocquencourt

`http://algo.inria.fr/flajolet`

Based on *Analytic Combinatorics*, Flajolet & Sedgewick, C.U.P., 2007$^+$.

Large random combinatorial structures exhibit are (often) predictable!





Concentration?

Limit law?

Relation to Bivariate GFs $C(z, u)$ and singularities?

DE MOIVRE LAPLACE GAUSS

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-w^2/2}\, dw$$

Why is the binomial distribution asymptotically normal?

• De Moivre: approximation of $\frac{1}{2^n}\binom{n}{k}$.

• Laplace/Gauss: as sum of many RV's + Lévy: . . . : because of characteristic functions $\rightarrow e^{-t^2/2}$.

• Analytic combinatorics: because of bivariate GF $\frac{1}{1-z(1+u)}$ and *smoothly varying singularity*!

Classical Central Limit Theorem (CLT): $\sum$ RV's *to* $\mathcal{N}$ormal.

Proof: Levy's continuity theorem $\phi_n(t) \to \phi(t)$ implies $F_n(x) \to F(x)$.
+ calculation of PGF $f_n(u) = g(u)^n$ + normalization and $u \mapsto it$.

Quasi-Powers Theorem (HK Hwang, circa 1995).

Assume (X_n) are RV's with probability GF (PGF) $f_n(u) = \mathbb{E}(u^{X_n})$ and for $A(u), B(u)$ *analytic* at 1:

$$\boxed{f_n(u) = A(u)B(u)^{\beta_n}\left(1 + O(\frac{1}{\kappa_n})\right)},$$

for $u \approx 1$, with $\beta_n, \kappa_n \to \infty$, and $\mathbb{V}\mathrm{ar}(B(u)) > 0$. Then

- mean: $\mu_n = \mathbb{E}(X_n) \sim \beta_n B'(1)$; s-dev.: $\sigma_n^2 \sim \beta_n \mathbb{V}\mathrm{ar}(B)$.
- normal limit: $\boxed{\mathbb{P}(X_n \le \mu_n + x\sigma_n) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-w^2/2}\, dw}$
- Speed of convergence is $O(\kappa_n^{-1} + \beta_n^{-1/2})$.

Quasi-Powers Theorem: "If you resemble a power, then your limit law is normal".

Proof. "Analytic expansions are differentiable": this gives moments.
Limit law results from Lévy's continuity theorem.
Speed results from Berry-Esseen.
$\ll$Bender, Richmond^{+}.

Example 1. Supercritical sequence *schema.*

Let $\mathcal{F} = \text{SEQ}(\mathcal{G})$, so that number of components has BGF

$$F(z,u) = \frac{1}{1-uG(z)}.$$

Assume that $G(r) > 1$ where r:=radius of conv. of $G(z)$.

Theorem. *The number of $\mathcal{G}$–components in a random $\mathcal{F}$–structure is asymptotically normal.*

Proof. A. Let $\rho \in (0,r)$ be such that $G(\rho) = 1$. This is r.o.c. of $F(z) \equiv F(z,1)$. There is a simple pole.
B. Equation $1-uG(z) = 0$ has root $\rho(u)$, where $\rho(u)$ depends analytically on u for $u \approx 1$.
C. Function $F(z,u)$, with u parameter, has simple pole at $\rho(u)$ and

$$[z^n]F(z,u) \sim c(u)\rho(u)^{-n}.$$

D. *Uniformity* is granted (by integral representations), so that Quasi-Powers Theorem applies. QED

Example 1. Supercritical sequences (continued)

— Compositions: arbitrary; with Ω–excluded or Ω–forced summands. Compositions into prime summands, $G(z) = z^2 + z^3 + z^5 + \cdots$. Same for twin primes (!!).

— Surjections aka ordered set partitions, $G(z) = e^z - 1$. Same with Ω–constraints.

— k–components in compositions, surjections, etc.

Example 2. Cycles in permutations.

$$F(z,u) = \exp\left(u \log \frac{1}{1-z}\right) = (1-z)^{-u}.$$

A. By *singularity analysis*, get main approximation : $[z^n]F(z,u) \sim \frac{n^{u-1}}{\Gamma(u)}$.

B. Approximation is uniform by nature of singularity analysis process (contour integration).

C. Rewrite as Quasi-Powers approximation:

$$[z^n]F(z,u) \sim \frac{1}{\Gamma(u)} \cdot \left(e^{(u-1)}\right)^{\log n}.$$

Thus, scale is now $\beta_n \sim \log n$.

D. Quasi-Powers Theorem applies. QED

Example 3. Exp-Log schema.

Let $\mathcal{F} = \textsc{Set}(\mathcal{G})$, so that number of components has BGF

$$F(z,u) = e^{uG(z)}.$$

Assume that $G(z)$ is logarithmic: $G(z) \sim \lambda \log \frac{1}{1-z/\rho}$.

Theorem. *The number of $\mathcal{G}$–components in a random $\mathcal{F}$–structure is asymptotically normal, with logarithmic mean and variance.*

Application: Random mappings, etc. $\gg$Arratia-Barbour-Tavaré.

Example 4. Polynomials over finite fields.

```
> Factor(x^7+x+1) mod 29;
   3    2                 2                  2
 (x  + x  + 3 x + 15) (x  + 25 x + 25) (x  + 3 x + 14)
```

- $\mathcal{P}$olynomial is a $\mathfrak{S}$*equence* of coeffs: $\mathcal{P}$ has Polar singularity.
- By unique factorization, $\mathcal{P}$ is also $\mathfrak{M}$*ultiset of* $\mathcal{I}$*rreducibles*:

  $\mathcal{I}$ has log singulariy.

$\Longrightarrow$ Prime Number Theorem for Polynomials $I_n \sim \dfrac{q^n}{n}$.

- Marking number of $\mathcal{I}$–factors is approx uth power:

$$P(z,u) \approx \left(e^{I(z)}\right)^u .$$

Variable Exponent $\Longrightarrow$ $\mathcal{N}$ormality of # of irred. factors.
(cf Erdős-Kac for integers.)

— Useful for analysis of polynomial factorization algorithms.

For a large collection of combinatorial classes & parameters, we have a functional equation

$$\Phi(z, y, u) = 0$$

In the counting case $(u=1)$ get a singular expansion

$$y(z,1) = \cdots (1 - z/\rho)^{\alpha} + \cdots$$

A PERTURBATION of <u>u</u> near 1 will often induce a smooth perturbation of the expansion of $y(z,u)$, e.g.,

movable singularity $\quad y(z,u) = \cdots (1 - z/\rho(u))^{\alpha} + \cdots$

movable exponent $\quad y(z,u) = \cdots (1 - z/\rho)^{\alpha(u)} + \cdots$

with $\rho(u)$ or $\alpha(u)$ analytic at 1

$\Longrightarrow$ Asymptotic normality { by singularity analysis + Quasi Powers

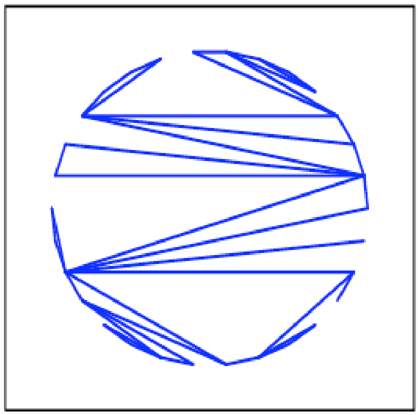
Perturbation of rational functions

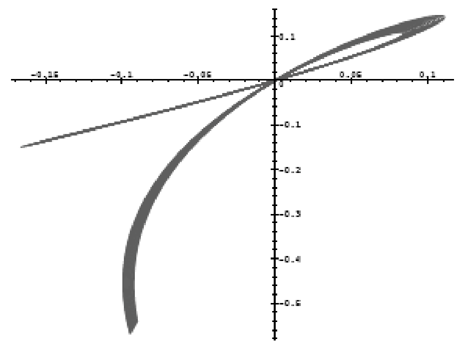
— Regular languages & automata, under irreducibity conditions. *Auxiliary mark u induces a smooth singularity dislacement.* Occurrences of patterns in random texts. Works for sets of patterns.

$\approx$ Extends CLT for finite Markov chains.

Perturbation of algebraic functions: for irreducible systems, the Drmota-Lalley-Woods Theorem implies $\sqrt{\ }$–singularity.

Example 5. Non-crossing graphs (Noy-F.)





= Perturbation of algebraic equation.

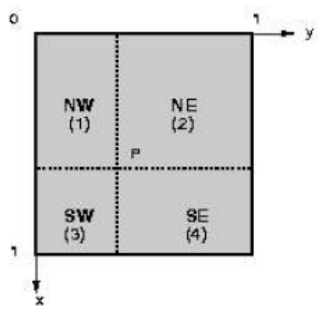
$$G^3 + (2z^2 - 3z - 2)G^2 + (3z + 1)G = 0$$
$$G^3 + (2u^3z^2 - 3u^2z + u - 3)G^2 + (3u^2 - 2u + 3)G + u - 1 = 0$$

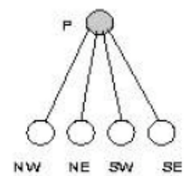
Movable singularity scheme applies: $\mathcal{N}$ormality.

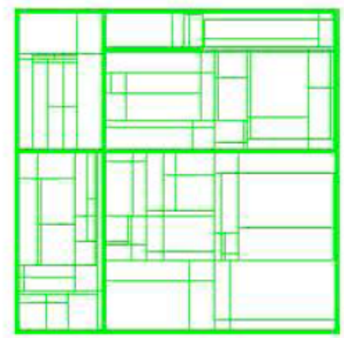
+ Patterns in context-free languages, in combinatorial tree models, in functional graphs: Drmota's version of Drmota-Lalley-Woods.

Perturbation of differential equations.

Example 6. Profile of Quadtrees.



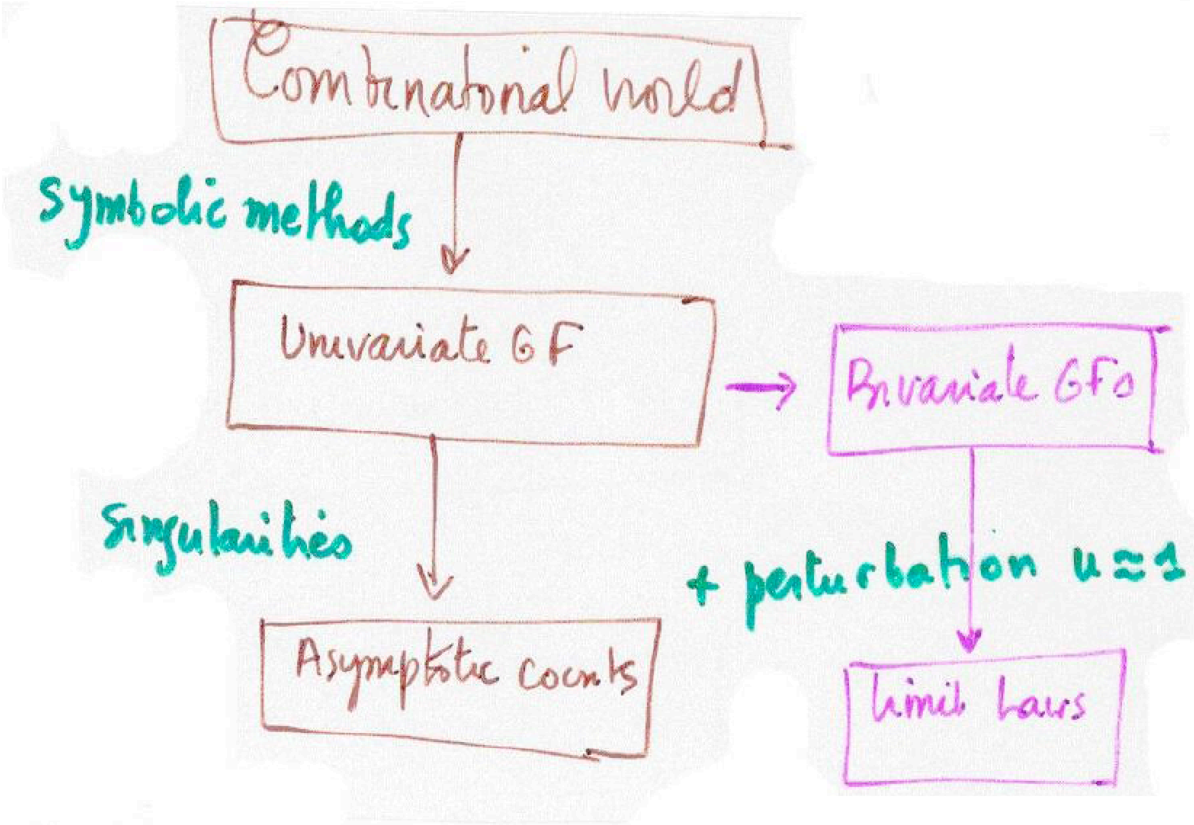




$$F(z,u) = 1 + 2^3 u \int_0^z \frac{dx_1}{x_1(1-x_1)} \int_0^{x_1} \frac{dx_2}{1-x_2} \int_0^{x_2} F(x_3,u) \frac{dx_3}{1-x_3}.$$

Solution is of the form $(1-z)^{-\alpha(u)}$ for algebraic branch $\alpha(u)$;
Variable Exponent $\Longrightarrow$ $\mathcal{N}$ormality of search costs.

Applies to many linear differential models that behave like *cycles-in-perms*.



That's All, Folks!