Gaussian Laws in Analytic Combinatorics

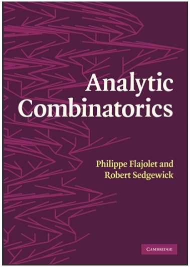
Michèle Soria

LIP6 — UPMC Sorbonne Universités

8 october 2015

IHP – Séminaire Flajolet

Estimate properties of large structured combinatorial objects



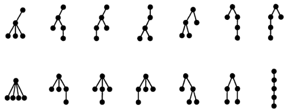
- Symbolic Method
  Combinatorial Specification $\longrightarrow$ Equations on (univariate) Generating Fonctions encoding counting sequences
- Complex Analysis
  GF as Analytic Functions $\longrightarrow$ extract asymptotic information on counting sequences
- Properties of Large Structures
  laws governing parameters in large random objects similar to enumeration
  – deformation (adjunction of auxiliary variable)
  – perturbation (effect of variations of auxiliary variable)

Gaussian Laws with Singularity Analysis + Combinatorial applications

Univariate Asymptotics

Combinatorial Specification → Univariate Generating Functions → **Asymptotic Enumeration**

Enumeration of general Catalan trees



$G(z) = \sum g_n z^n$, $g_n = \#$ trees of size n

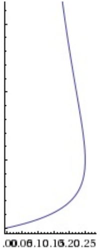
$$\mathcal{G} = \bullet \times \textsc{Seq}(\mathcal{G}) \rightarrow G(z) = \frac{z}{1 - G(z)}$$

$$G(z) = \frac{1}{2}\left(1 - \sqrt{1-4z}\right)$$

$$G(z) = \frac{1}{2}\left(1 - \sqrt{1-4z}\right)$$

$$\Longrightarrow g_{n+1} = \frac{1}{n+1}\binom{2n}{n} \sim \frac{4^n n^{-3/2}}{\sqrt{\pi}}$$

stirling approx. $n! = \sqrt{2\pi n}(n/e)^n(1 + 0(1/n))$



$G(z)$ has a singularity of square-root type at $\rho = \frac{1}{4}$

$\Longrightarrow g_n \sim c\, 4^n n^{-3/2}$

Singularity Analysis

Bivariate Asymptotics

Combinatorial Specification → Bivariate Generating Functions → **Asymptotic Behavior of Parameters**

Number of leaves

$G(z,u) = \sum g_{n,k} z^n u^k$, $g_{n,k} = \#$ trees of size n with k leaves

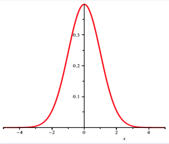
$$\mathcal{G}^\circ = \bullet^\circ + \bullet \times \textsc{Seq}_{\geq 1}(\mathcal{G}^\circ) \;\rightarrow\; G(z,u) = zu + \frac{zG(z,u)}{1 - G(z,u)}$$

$$G(z,u) = \tfrac{1}{2}\left(1 + (u-1)z - \sqrt{1 - 2(u+1)z + (u-1)^2 z^2}\right)$$

$$\Longrightarrow g_{n,k} = \frac{1}{n}\binom{n}{k}\binom{n-2}{k-1}$$

Mean value $\mu_n = n/2$; standard deviation $\sigma_n = O(\sqrt{n})$

for $k = \frac{n}{2} + x\sqrt{n}$, $\quad \frac{g_{n,k}}{g_n} \rightarrow_{n\to\infty} \frac{1}{\sigma_n\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_n^2}}$



Bivariate Asymptotics

Combinatorial Specification → Generating Functions Bivariate → **Asymptotic Behavior of Parameters**

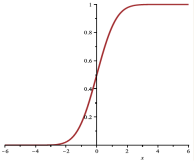
$$G(z,u) = \frac{1}{2}(1 + (u-1)z - \sqrt{1 - 2(u+1)z + (u-1)^2 z^2})$$
$$= g(z,u) - h(z,u)\sqrt{1 - \frac{z}{\rho(u)}}, \qquad \rho(u) = (1+\sqrt{u})^{-2}$$

Probability generating function $p_n(u) = \sum_k \frac{g_{n,k}}{g_n} u^k = \frac{[z^n]G(z,u)}{[z^n]G(z,1)}$

$$\boxed{p_n(u) = \left(\frac{\rho(u)}{\rho(1)}\right)^{-n} (1 + o(1))}$$

Perturbation of Singularity Analysis
+ Quasi-Powers approximation

→ Central Limit Theorem



TOOLS

- Generating Functions and Limit Laws
  - Bivariate Generating Functions
  - Continuous Limit Laws
  - Continuity Theorem for Characteristic functions
- Quasi-Powers Theorem
  - Central Limit Theorem
  - Quasi-Powers
- Singularity Analysis
  - Asymptotic enumeration
  - Perturbation of Singularity Analysis

Bivariate Generating Functions and Limit Laws

Combinatorial class $\mathcal{A}$

Size $|.| : \mathcal{A} \to \mathbb{N}$
Counting Generating Function
$A(z) = \sum_{\alpha \in \mathcal{A}} z^{|\alpha|} = \sum_n a_n z^n$

Parameter $\chi : \mathcal{A} \to \mathbb{N}$
Bivariate Generating Function
$$A(z,u) = \sum_{\alpha \in \mathcal{A}} u^{\chi(\alpha)} z^{|\alpha|} = \sum_{n,k} a_{n,k} u^k z^n$$

Uniform Model

Random Variables $(X_n)_{n \geq 0}$
$\mathbb{P}(X_n = k) = \frac{a_{n,k}}{a_n}$

Probability Generating Function
$$p_n(u) \equiv \sum_k \mathbb{P}(X_n = k) u^k = \mathbb{E}(u^X) = \frac{[z^n]A(z,u)}{[z^n]A(z)}$$

Mean value $\mu_n \equiv \mathbb{E}(X_n) = p_n'(1)$
Variance $\sigma_n^2 = p_n''(1) - \mu_n^2 + \mu_n$

Question : Asymptotic behavior of X_n ?

moments, limit law (density, cumulative), tails of distribution

Answer by Analytic Combinatorics : Evaluate $[z^n]A(z,u)$

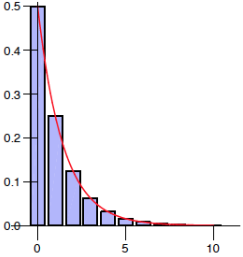
in different u-domains

Examples of discrete Limit Law

- $\mu_n = O(1)$ and $\sigma_n^2 = O(1)$ discrete limit law :

$$\forall k, \mathbb{P}(X_n = k) \to_{n\to\infty} \mathbb{P}(X = k), \; X \text{ discrete RV}$$

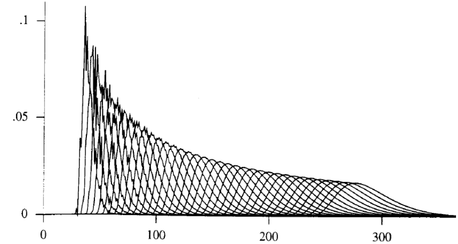
Ex. root degree distribution in Catalan trees : Geometric

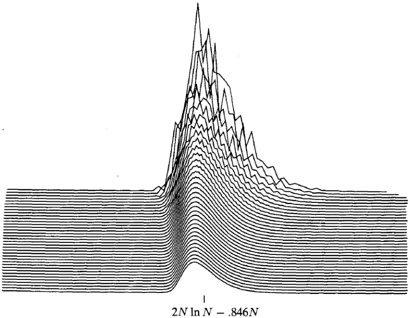


Examples of continuous Limit Law

- μ_n and $\sigma_n^2 \to \infty$

Quicksort : number of comparisons $10 < N \leq 50$





Normalized variable $X_n^* = \dfrac{X_n - \mu_n}{\sigma_n}$

continuous limit law

Continuous Limit Laws

Y Continuous Random Variable

Distribution Function
$F_Y(x) := \mathbb{P}(Y \leq x)$

Characteristic function
(Fourier Transform)
$\varphi_Y(t) = \mathbb{E}(e^{itY})$

$F_Y(x)$ differentiable :
density $f_Y(x) := F'_Y(x)$

(X_n^*) sequence of (norm. combinatorial) Random Variables $X_n^* = \frac{X_n - \mu_n}{\sigma_n}$

Weak Convergence $X_n^* \Rightarrow Y$

$\forall x, \lim_{n\to\infty} F_{X_n^*}(x) = F_Y(x)$

Local Limit Law $X_n^* \to Y$

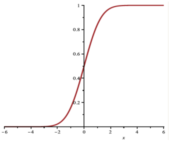
$\mathbb{P}(X_n = \lfloor \mu_n + x\sigma_n \rfloor) = \frac{1}{\sigma_n} f(x)$

Continuity Theorem (Lévy 1922)

$X_n^* \Rightarrow Y \iff \varphi_{X_n^*} \to \varphi_Y$

$$\varphi_{X_n^*}(t) = \mathbb{E}(e^{it\frac{X_n - \mu_n}{\sigma_n}}) = e^{-it\frac{\mu_n}{\sigma_n}} \mathbb{E}(e^{it\frac{X_n}{\sigma_n}}) = e^{-it\frac{\mu_n}{\sigma_n}} p_n(e^{\frac{it}{\sigma_n}})$$

Gaussian Limit Laws $\mathcal{N}(0,1)$

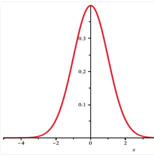


$\mathcal{N}(0,1)$

Distribution Function $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$

Probability density $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

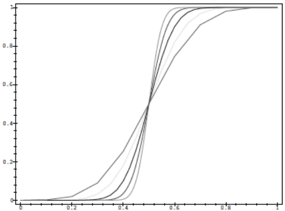
Characteristic Function $\varphi(t) = e^{-\frac{t^2}{2}}$

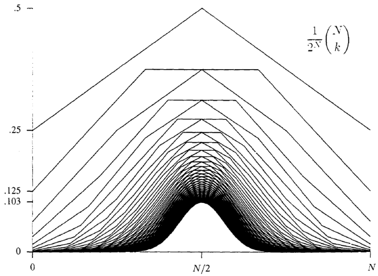


Number of a's in words on $\{a,b\}^*$ of length n

Bivariate GF $F(z,u) = \dfrac{1}{1 - z - zu}$

Binomial Distribution $\mathbb{P}(X_n = k) = \frac{1}{2^n}\binom{n}{k}$





Central Limit Theorem (Demoivre, Laplace, Gauss)

Let $T_1, \ldots, T_n$ be i.i.d RV, with finite mean μ and variance σ^2, and let $S_n = \sum T_i$. The normalized variable S_n^* weekly converges to a normalized Gaussian distribution :

$$S_n^* \equiv \frac{S_n - n\mu}{\sigma\sqrt{n}} \Rightarrow \mathcal{N}(0,1)$$

Proof by Levy's theorem : characteristic function $\mathbb{E}(e^{itS_n}) = (\phi_{T_1}(t))^n$, with $\phi_{T_1}(t) = 1 + it\mu - \frac{t^2}{2}(\mu^2 + \sigma^2) + o(t^2)$, $t \to 0$

$$\phi_{S_n^*}(t) = \mathbb{E}(e^{it\frac{S_n - n\mu}{\sigma\sqrt{n}}}) = e^{-it\frac{n\mu}{\sigma\sqrt{n}}}\,\mathbb{E}(e^{it\frac{X_n}{\sigma\sqrt{n}}}) = e^{-it\frac{n\mu}{\sigma\sqrt{n}}}\,(\phi_{T_1}(\tfrac{t}{\sigma\sqrt{n}}))^n$$

$$\log \phi_{S_n^*}(t) = -it\frac{n\mu}{\sigma\sqrt{n}} + n\log\left(1 + \frac{it\mu}{\sigma\sqrt{n}} - \frac{t^2}{2\sigma^2 n}(\mu^2+\sigma^2) + o(\tfrac{t^2}{n})\right)$$

$$= -\frac{t^2}{2} + O(\tfrac{1}{\sqrt{n}})$$

Thus $\phi_{S_n^*}(t) \to e^{-\frac{t^2}{2}}$

$$\boxed{\phi_{S_n^*}(t) = p_n(u) = p^n(u)}$$

$$\boxed{u = e^{\frac{it}{\sigma n}} : u \approx 1 \text{ when } n \to \infty}$$

Quasi-Powers Theorem – Example

Cycles in Permutations

$$\mathcal{P} = \textsc{Set}(\circ\,\textsc{Cycle}(\mathcal{Z})) \quad \rightarrow \qquad F(z,u) = \exp\left(u \log \frac{1}{1-z}\right)$$

$F(z,u) = \sum_{n,k} f_{n,k} u^k \frac{z^n}{n!} = \sum_n p_n(u) z^n$ with $p_n(u) = \sum \mathbb{P}(X_n = k) u^k$

$F(z,u) = (1-z)^{-u} = \sum \binom{n+u-1}{n} z^n$

$$p_n(u) = \frac{u(u+1)\dots(u+n-1)}{n!} = \frac{\Gamma(u+n)}{\Gamma(u)\Gamma(n+1)}$$

$$p_n(u) = \frac{1}{\Gamma(u)} (e^{u-1})^{\log n} \left(1 + O\left(\frac{1}{n}\right)\right), \quad \text{uniform for} \quad u \approx 1$$

Theorem (Goncharov 1944)

The number of cycles in a random permutation is asymptotically Gaussian, with $\mu_n = H_n$ and $\sigma_n = \sqrt{\log n} + o(1)$.

Quasi-Powers Theorem

Quasi-powers Theorem (H-K Hwang 1994)

X_n non-negative discrete random variables with PGF $p_n(u)$.
Assume that, uniformly for $u \approx 1$ (u in a fixed neighbourhood of 1)

$$p_n(u) = A(u)B(u)^{\lambda_n}\left(1 + O\left(\frac{1}{\kappa_n}\right)\right), \text{ with } \lambda_n, \kappa_n \to \infty, \; n \to \infty$$

with $A(u), B(u)$ analytic at $u = 1$, $A(1) = B(1) = 1$ and "variability conditions" on $B(u)$. ***Then***

Mean $\mathbb{E}(X_n) = \lambda_n B'(1) + A'(1) + O(\frac{1}{\kappa_n})$,
Variance $Var(X_n) = \lambda_n(B''(1) + B'(1) - B'^2(1)) + ...$
Distribution asymptotically Gaussian :

$$\mathbb{P}(X_n < \mu_n + x\sigma_n) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} \, dt$$

Speed of Convergence $O\left(\frac{1}{\kappa_n} + \frac{1}{\sqrt{\beta_n}}\right)$

Proof : analyticity + uniformity $\to$ error terms transfer to differentiation.
Like in CLT only local properties near $u = 1$ are needed since $\lambda_n \to \infty$.
Normalization, Characteristic function + Levy Theorem

Other Properties

Local Limit Theorem (H-K Hwang 1994)

If the Quasi-Powers approximation holds on the circle $|u| = 1$ then

$$\sigma_n \mathbb{P}(X_n = \lfloor \mu_n + x\sigma_n \rfloor) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Large Deviations Theorem (H-K Hwang 1994)

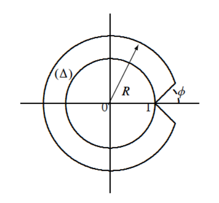
If the Quasi-Powers approximation holds on an interval containing 1 then

$$\log \mathbb{P}(X_n < x\lambda_n) \leq \lambda_n W(x) + O(1)$$

where $W(x) = \min \log(B(u)/u^x)$

Theorem : Singularity Analysis (Flajolet-Odlyzko 90)

$F(z) = \sum f_n z^n$ analytic function at 0, unique dominant sigularity at $z = \rho$, and $F(z)$ analytic in some Δ-domain (Δ-analytic)



and locally,

$$F(z) \sim \frac{1}{(1 - z/\rho)^{\alpha}} \quad \alpha \in \mathbb{R} - \mathbb{Z}^{-}$$

then

$$f_n \sim \frac{\rho^{-n} n^{\alpha-1}}{\Gamma(\alpha)}$$

Cauchy coefficient integral : $[z^n]\, F(z) = \frac{1}{2i\pi} \int_{\gamma} F(z) \frac{dz}{z^{n+1}}$

F alg-log singularity *Hankel Contour*

$$F(z) \sim \frac{1}{(1-z/\rho)^{\alpha}} \log^{\beta} \frac{1}{(1-z/\rho)} \to f_n \sim \frac{\rho^{-n} n^{\alpha-1}}{\Gamma(\alpha)} \log^{\beta} n$$

Perturbation of Singularity Analysis

Robust : contour integrals $\rightarrow$ uniform error terms for parameters

$$p_n(u) = \frac{[z^n]F(z,u)}{[z^n]F(z,1)}$$

u parameter
Sing. of $F(z,1) \rightarrow$ Sing. of $F(z,u)$ for $u \approx 1$

Movable singularities

$$F(z,u) \approx \frac{1}{(1 - z/\rho(u))^{\alpha}} \longrightarrow p_n(u) \sim \left(\frac{\rho(u)}{\rho(1)}\right)^{-n}, \text{ uniform for } u \approx 1.$$

$\Longrightarrow$ Gaussian limit law, with a mean and variance of order n.

Variable exponent

$$F(z,u) \approx \frac{1}{(1 - z/\rho)^{\alpha(u)}} \longrightarrow p_n(u) \sim n^{\alpha(u)-\alpha(1)}, \text{ uniform for } u \approx 1$$

$\Longrightarrow$ Gaussian limit law, with a mean and variance of order $\log n$.

Analytico-Combinatorial Schemes

- Perturbation of Meromorphic Asymptotics
  - Meromorphic functions
  - Linear systems
- Perturbation of Singularity Analysis
  - Alg-Log Scheme : movable singularities
  - Algebraic Systems
  - Exp-Log Scheme : variable exponent
  - Differential equations

Meromorphic Schemes

Example : Eulerian numbers

Rises in permutations $2 \nearrow 6 \;\; 4 \;\; 1 \nearrow 3 \nearrow 5 \nearrow 8 \;\; 7$

$$F(z,u) = \frac{u(1-u)}{e^{(u-1)z} - u} \qquad\qquad F(z,1) = \frac{1}{1-z}$$

$$\rho(u) = \frac{\log u \; + \; 2ik\pi}{u-1} \quad u \approx 1 \qquad\qquad \rho = 1$$

$$[z^n]F(z,u) \;=\; \frac{1}{2i\pi}\int_{|z|=1/2} F(z,u)\frac{dz}{z^{n+1}}$$

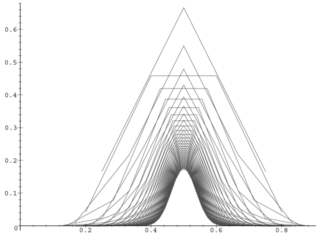
$$\textit{uniformly} \;=\; (\rho(u))^{-n-1} + O(2^{-n})$$

Quasi-Powers Theorem

$\Longrightarrow$ Number of rises in permutations

Gaussian limit law

$\mu_n \sim \frac{n}{2}, \sigma_n^2 \sim \frac{n}{12}$



Meromorphic Scheme : Theorem

Meromorphic Theorem (Bender 93)

- *$F(z,u)$ analytic at $(0,0)$, non negative coefficients.*
- *$F(z,1)$ single dominant simple pole at $z=\rho$.*
- $F(z,u) = \dfrac{B(z,u)}{C(z,u)}$, *for* $|z| < \rho + \varepsilon$ *and* $|u-1| < \eta$

  $B(z,u), C(z,u)$ analytic for $|u-1| < \eta$, $|z| < \rho + \varepsilon$

  (ρ simple zero of $C(z,1)$ and $\rho(u)$ analytic $u \approx 1$, $C(u, \rho(u) = 0)$)
- *Non degeneracy + variability conditions*

Then

- *Gaussian limit law, mean and variance asymptotically linear : $\mathcal{N}(c_1 n, c_2 n)$*
- *Speed of Convergence $O(n^{-1/2})$*

Proof : $f_n(u) =$ Residue + exponentially small uniform for $u \approx 1$

Applications : Supercritical Sequences

Supercritical Composition

Substitution $\mathcal{F} = \mathcal{G}(\mathcal{H}) \to f(z) = g(h(z)), \;\; h(0) = 0$

$\tau_h = \lim_{z \to \rho_h^-} h(z)$. The composition is supercritical if $\tau_h > \rho_g$

Supercritical Sequences $\mathcal{F} = \textsc{Seq}(\mathcal{H}) \to f(z) = \frac{1}{1-h(z)}$, $\tau_h > 1$

The number of components in a random supercritical sequence of size n is asymptotically Gaussian, with $\mu_n = \frac{n}{\rho h'(\rho)}$, where ρ is the positive root of $h(\rho) = 1$, and $\sigma^2(n) = c_\rho n$

- blocks in a surjection $F(z,u) = \frac{1}{1-u(e^z-1)}$
- parts in an integer composition $F(z,u) = \frac{1}{1-uh(z)}$, $h(z) = \sum_{i \in \mathcal{I}} z^i$
- cycles in an alignment $F(z,u) = \frac{1}{1-u\log(1-z)^{-1}}$

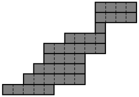
Other applications

- Occurrences of a fixed pattern of length k in words

$$F(z,u) = \frac{1-(u-1)(c(z)-1)}{1-2z-(u-1)(z^k+(1-2z)(c(z)-1))}$$

- Parallelogram polyominoes of area n

  - perimeter
  - width
  - height



$$F(z,u) = \frac{J_1(z,u)}{J_0(z,u)}$$

Bessel Functions

- GCD of polynomials over finite field

  $\mathcal{P}$ monic polynomials of $d^o n$ in $\mathbb{F}_p(z)$: $P(z) = \frac{1}{1-pz}$

steps in Euclid's algorithm

$$\begin{cases} u_0 = q_1 u_1 + u_2 \\ \cdots \\ u_{h-2} = q_{h-1} u_{h-1} + u_h \\ u_{h-1} = q_h u_h + 0 \end{cases}$$

$$F(z,u) = \frac{1}{1-uG(z)} P(z) = \frac{1}{1-u\frac{p(p-1)z}{1-p(z)}} \frac{1}{1-pz}$$

Gaussian limit law with mean and variance asymptotically linear

Theorem Gaussian Limiting Distribution Bender 93

$$\mathbf{Y}(z,u) = \mathbf{V}(z,u) + \mathbf{T}(z,u).\mathbf{Y}(z,u)$$

- *each V_i and $T_{i,j}$ polynomial in z, u with non negative coefficients.*
- *Technical conditions : irreducibility, aperiodicity, unicity of dominant pole of $Y(z,1)$.*

Then
Limiting distribution of the additional parameter u : $\mathcal{N}(c_1 n, c_2 n)$

Applications

- Irreducible and aperiodic finite Markov chain, after *n* transitions, number of times a certain state is reached asymptotically Gaussian.
- Set of patterns in words : number of occurrences asymptotically Gaussian

ALG-LOG SCHEMES

Example : leaves in Catalan Trees

$$G(z,u) = zu + \frac{zG(z,u)}{1 - G(z,u)}$$

$F(z,u) = G(z,u^2)$ satisfies

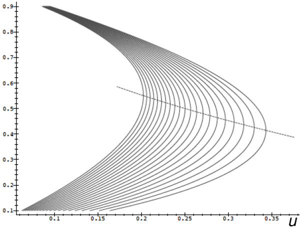
$$F(z,u) = \boxed{\frac{1}{2} + \frac{1}{2}z(u^2-1)} - \boxed{\frac{1}{2}\sqrt{1 - z(1-u)^2}} \times \boxed{\frac{1}{2}\sqrt{1 - z(1+u)^2}}$$

$F(z,u) = A(z,u) - B(z,u)(1 - \frac{z}{\rho(u)})^{1/2}$ with $\rho(u) = (1+u)^{-2}$

$$[z^n]F(z,u) = b_u \frac{\rho(u)^{-n} n^{-3/2}}{2\sqrt{\pi}} \left(1 + O\left(\frac{1}{n}\right)\right)$$

Singularity Analysis $\Longrightarrow$ uniformity for $u \approx 1$

$p_n(u) \sim \left(\frac{\rho(u)}{\rho(1)}\right)^{-n}$ Quasi-Powers Theorem



Gaussian limit law mean and variance $\sim$ linear

Alg-Log Scheme : Theorem

Algebraic Singularity Theorem (Flajolet-S. 93, Drmota 97)

- $F(z,u)$ *analytic at* $(0,0)$*, non negative coefficients.*
- $F(z,1)$ *single dominant alg-log singularity at* $z = \rho$
- *For* $|z| < \rho + \varepsilon$ *and* $|u - 1| < \eta$
  $F(z,u) = A(z,u) + B(z,u)C(z,u)^{-\alpha}, \alpha \in \mathbb{R} - \mathbb{Z}^-$
  $A(z,u), B(z,u), C(z,u)$ *analytic for* $|u - 1| < \eta$*,* $|z| < \rho + \varepsilon$
- *Non degeneracy* $+$ *variability conditions.*

Then

- *Gaussian limit law, mean and variance asymptotically linear :* $\mathcal{N}(c_1 n, c_2 n)$
- *Speed of Convergence* $O(n^{-1/2})$

Proof : Uniform lifting of univariate SA for $u \approx 1$
Implicit Function Theorem :$\rho(u)$; $C(u, \rho(u) = 0)$ analytic $u \approx 1$,
+ Quasi-Powers Theorem

Applications

Functional equation (Bender, Canfield, Meir-Moon)

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

$\Phi(z, y)$ has a power series expansion at $(0, 0)$ with non-negative coefficients, non linear ($\Phi_{yy}(z, y) \neq 0$), and well-defined ($\Phi_z(z, y) \neq 0$).
Let $\rho > 0, \tau > 0$ such that $\tau = \Phi(\rho, \tau)$ and $1 = \Phi_y(\rho, \tau)$.
Then $\exists$ $g(z)$, $h(z)$ analytic functions such that locally

$$y(z) = g(z) - h(z)\sqrt{1 - z/\rho} \quad \text{with} \quad g(\rho) = \tau \quad \text{and} \quad h(\rho) \neq 0.$$

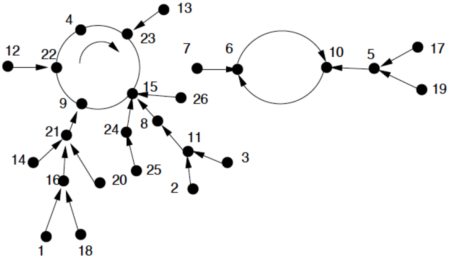
Simple family of trees $Y(z) = z\Phi(z, Y(z))$

Catalan, Cayley , Motzkin, ...

- number of leaves
- number of occurrences of a fixed pattern
- number of nodes of any fixed degree

Gaussian limiting distribution, mean and variance $\sim$ linear.

Random Mappings



$$\begin{cases} \mathcal{M} = \text{SET}(\text{CYCLE}(\mathcal{T})) \\ \mathcal{T} = \bullet \times \text{SET}(\mathcal{T}) \end{cases}$$

$$M(z) = \exp\left(\log \frac{1}{1-T(z)}\right),$$
$$T(z) = ze^{T(z)}$$

$$T(z) = 1 - \sqrt{2(1-ez)} + sot.$$

- number of leaves
- number of nodes with a fixed number of predecessors

Gaussian limiting distribution, with mean and variance asymptotically linear.

- Extends to Pólya operators :

  Otter Trees : $B(z) = z + \frac{1}{2}B^2(z) + \frac{1}{2}B(z^2)$,

  General non plane trees : $H(z) = z \exp(\sum \frac{H(z^k)}{k})$

  number of leaves, number of nodes with a fixed number of predecessors asymptotically Gaussian

- Characteristics of random walks in the discrete plane : number of steps of any fixed kind, number of occurrences of any fixed pattern asymptotically Gaussian

- Planar maps : number of occurrences of any fixed submap asymptotically Gaussian

Algebraic Systems

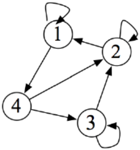
Theorem Gaussian Limiting Distribution Drmota, Lalley – 95-97

$\mathbf{Y}(z) = \mathbf{P}(z, \mathbf{Y}(z), u)$ *positive* and *well defined polynomial* (entire) ***system of equations*** , *in* $z, \mathbf{Y}, u$ *which has a solution* $\mathbf{F}(z, u)$.
Strongly connected dependency graph (non linear case), locally

$$f_j(z,u) = g_j(z,u) - h_j(z,u)\sqrt{1 - z/\rho(u)}$$

$g_j(z,u)$, $h_j(z,u)$, $\rho(u)$ *analytic non zero functions.*

Limiting distribution of the additional parameter u : $\mathcal{N}(c_1 n, c_2 n)$



$$\begin{aligned} y_1(z) &= p_1(z, y_1, y_2) \\ y_2(z) &= p_2(z, y_2, y_3, y_4) \\ y_3(z) &= p_3(z, y_3, y_4) \\ y_4(z) &= p_4(z, y_1) \end{aligned}$$

General dependancy graph Drmota, Banderier 2014

Gaussian Limiting Distribution when all strongly connected components have different radius of convergence.

Algebraic Systems : Applications

- number of independent sets in a random tree
- number of patterns in a context-free language

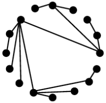
Gaussian limiting distribution, mean and variance asymptotically linear.

Non-crossing familiy : trees, forests, connect. or general graphs

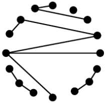
- number of connected components
- number of edges

Forests

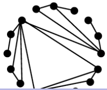
$$F^3(z) + (z^2 - z - 3)F^2(z) + (z+3)F - 1 = 0$$

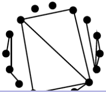


(tree)



(forest)

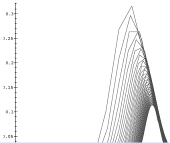




irreducible aperiodic system

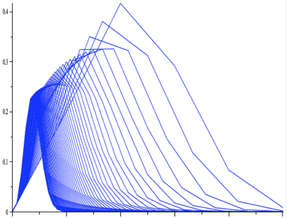
$$F = 1 + zFU$$
$$U = 1 + UV$$
$$V = zFU^2$$

$\Longrightarrow$ square-root type singularity



EXP-LOG SCHEMES

Example : cycles in permutations



Stirling numbers of 1st kind

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

$$p_n(u) = [z^n]F(z,u) = \frac{n^{u-1}}{\Gamma(u)}\left(1 + O(\tfrac{1}{n})\right)$$

Singularity Analysis $\rightarrow$ uniformity for $u \approx 1$

Quasi-Powers approximation with $\beta_n = \log n$

$$p_n(u) \sim \frac{1}{\Gamma(u)}(\exp(u-1))^{\log n}$$

Number of cycles in a random permutation asymptotically Gaussian, with μ_n and σ_n^2 equivalent to $\log n$

Exp-log Scheme : Theorem

Variable exponent Theorem (Flajolet-S. 93)

- *$F(z, u)$ analytic at $(0, 0)$, non negative coefficients.*
- *$F(z, 1)$ single dominant alg-log singularity at $z = \rho$*
- *For $|z| < \rho + \varepsilon$ and $|u - 1| < \eta$*
  $F(z, u) = A(z, u) + B(z, u)C(z)^{-\alpha(u)}$, $\alpha(1) \in \mathbb{R} - \mathbb{Z}^-$
  $A(z, u), B(z, u)$ analytic for $|u - 1| < \eta$, $|z| < \rho + \varepsilon$;
  $\alpha(u)$ analytic for $|u - 1| < \eta$;
  $C(z)$ analytic for $|z| < \rho + \varepsilon$ and unique root $C(\rho) = 0$.
- *Non degeneracy + variability conditions.*

Then

- *Gaussian limit law, mean and variance asymptotically logaritmic :* $\mathcal{N}(\alpha'(1) \log n, (\alpha'(1) + \alpha''(1)) \log n)$
- *Speed of Convergence* $O((\log n)^{-1/2})$

Proof : Singularity Analysis + Quasi-Powers Theorem

Number of components in a set

$\mathcal{G}$ **of *logarithmic type*** : $G(z)$ singular at ρ, and locally

$$G(z) = k \log \frac{1}{1 - z/\rho} + \lambda + O(\log^{-2}(1 - z/\rho))$$

Labelled exp-log schema $\mathcal{F} = \textsc{Set}(\circ\mathcal{G})$

$$F(z,u) = \exp(uG(z)) \to F(z,u) \sim \frac{1}{(1 - z/\rho)^{uk}}$$

Number of components $\to \mathcal{N}(k \log n, k \log n)$

Unlabelled exp-log schem $\mathcal{F} = \textsc{MSet}(\circ\mathcal{G})$ **or** $\mathcal{F} = \textsc{PSet}(\circ\mathcal{G})$

$$F(z,u) = \exp\left(\sum \frac{u^k}{k} G(z^k)\right) \quad \text{or} \quad F(z,u) = \exp\left(\sum \frac{(-u)^k}{k} G(z^k)\right)$$

$$\rho < 1 \Longrightarrow F(z,u) = H(z,u)\, exp(uG(z))$$

$H(z,u)$ analytic $|u-1| < \eta$, $|z| < \rho + \varepsilon$

Number of components $\to \mathcal{N}(k \log n, k \log n)$

Connected components in Random Mappings

Random Mappings (labelled)

$$\begin{cases} \mathcal{M} = \text{SET}(\circ\text{CYCLE}(\mathcal{T})) \\ \mathcal{T} = \bullet \times \text{SET}(\mathcal{T}) \end{cases}$$

$$\begin{cases} M(z) = \exp\left(\log \frac{1}{1-T(z)}\right) \\ T(z) = ze^{T(z)} \end{cases}$$

$$T(z) = 1 - \sqrt{2(1-ez)} + sot.$$

Connected components

$$M(z,u) = (1 - T(z))^{-u}$$
$$\sim (2(1-ez))^{-\frac{u}{2}}$$

Gaussian limiting law

mean and variance $\sim \frac{1}{2} \log n$.

Random Mapping Patterns (unlabelled)

$$\begin{cases} \mathcal{MP} = \text{MSET}(\circ\mathcal{L}) \\ \mathcal{L} = \text{CYCLE}(\mathcal{H}) \\ \mathcal{H} = \bullet \times \text{MSET}(\mathcal{H}) \end{cases}$$

$$\begin{cases} MP(z) = \exp\left(\sum \frac{L(z^k)}{k}\right) \\ L(z) = \sum \frac{\phi(k)}{k} \log \frac{1}{1-H(z^k)} \\ H(z) = z \exp(\sum \frac{H(z^k)}{k}) \end{cases}$$

$$H(z) = 1 - \gamma\sqrt{(1 - z/\eta)} + sot., \ \eta < 1$$

Connected components

$$MP(z,u) \sim c_u \, (1 - z/\eta)^{-\frac{u}{2}}$$

Gaussian limiting law

mean and variance $\sim \frac{1}{2} \log n$.

Irreducible factors in a polynomial $d^o n$ in $\mathbb{F}_p$

- $\mathcal{P}$ monic polynomials of $d^o n$ in $\mathbb{F}_p(z)$: Polar singularity $1/p$

$$P_n = p^n, \qquad P(z) = \frac{1}{1 - pz}$$

- Unique factorization property :

$$\mathcal{P} = \text{MSET}(\mathcal{I}) \to P(z) = exp(\sum \frac{I(z^k)}{k})$$

- Mobius inversion : $I(z) = \sum_{k \geq 1} \frac{\mu(k)}{k} \log(P(z^k))$
- Logarithmic type : $I(z) = \log \frac{1}{1 - pz} + \sum_{k \geq 2} \frac{\mu(k)}{k} \log(P(z^k))$
- Thus number of irreducible factors asymptotically Gaussian with mean and variance $\sim \log n$.

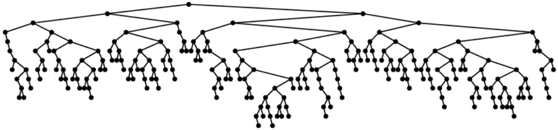
(Erdös-Kac Gaussian Law for the number of prime divisors of natural numbers.)

Linear Differential Equations – node level in BST

Binary Search Trees (increasing trees)

$$\mathcal{F} = 1 + \mathcal{Z}^{\square} \times \mathcal{F} \times \mathcal{F} \qquad \rightarrow \qquad F(z) = 1 + \int_0^z F^2(t)dt$$

Internal nodes



$$F(z,u) = 1 + 2u \int_0^z F(t,u) \frac{dt}{1-t}$$

$F'_z(z,u) = \dfrac{2u}{1-z} F(z,u)$, with $F(0,u) = 1 \rightarrow F(z,u) = \frac{1}{(1-z)^{2u}}$

Singularity analysis

$[z^n]F(z,u) = \frac{n^{2u-1}}{\Gamma(2u)} \left(1 + O(\frac{1}{n})\right)$, uniform for $u \approx 1$

Distribution of depth of a random node in a random Binary Search Trees asymptotically Gaussian, with mean and variance $\sim 2 \log n$

Linear Differential Equations

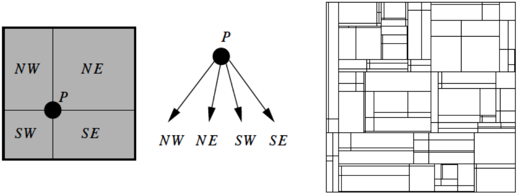
Linear differential equations (Flajolet-Lafforgue 94)

- *$F(z,u)$ analytic at $(0,0)$, non negative coefficients.*
- *$F(z,1)$ single dominant alg-log singularity at ρ : $f_n \sim C\rho^{-n} n^{\sigma-1}$*
- *$F(z,u)$ satisfies a linear differential equation*
  $a_0(z,u)F^{(r)}(z,u) + \frac{a_1(z,u)}{(\rho-z)}F^{(r-1)}(z,u) + \ldots + \frac{a_r(z,u)}{(\rho-z)^r}F(z,u) = 0$
  $a_i(z,u)$ analytic at $|z| = \rho$
- *Non degeneracy + variability conditions.*
- *Indicial polynomial*
  $J(\theta) = a_0(\rho,1)(\theta)_r + a_1(\rho,1)(\theta)_{r-1} + \ldots + a_1(\rho,1)$
  unique simple root $\sigma > 0$. $(\theta)_r = \theta(\theta-1)\ldots(\theta-r+1)$

Then

- *Gaussian limit law, mean and variance asymptotically logarithmic : $\mathcal{N}(c_1 \log n, c_2 \log n)$*

Node level in quadtrees



$$F(z,u) = 1 + 4u\,\mathbf{I}\,.\,\mathbf{J}F(z,u)$$

$$\mathbf{I}[g](z) = \int_0^z g(t)\frac{dt}{1-t} \quad \mathbf{J}[g](z) = \int_0^z g(t)\frac{dt}{t(1-t)}$$

Indicial polynomial $J(\theta,u) = \theta^2 - 4u \to$ root $\sigma(u) = 2\sqrt{u}$

Depth of a random external node in a random quadtree asymptotically Gaussian, mean and variance $\sim$ *logn*

Non Linear Differential Equations ?

Binary Search Trees, *parameter* χ ,

$$F(z,u) = \sum_T \lambda(T) u^{\chi(T)} z^{|T|}, \text{ where } \lambda(T) = \prod_{V \prec T} \frac{1}{|V|}$$

- Paging of BST : index + pages of size $\leq b$

$$F_z'(z,u) = uF(z,u)^2 + (1-u)\frac{d}{dz}\left(\frac{1-z^{b+1}}{1-z}\right), \quad F(0,u) = 1$$

- Occurences of a pattern P in BST

$$F_z'(z,u) = F(z,u)^2 + |P|\lambda(P)(u-1)z^{|P|-1}, \quad F(0,u) = 1$$

Riccati equations :

$$Y' = aY^2 + bY + c \rightarrow W'' = AW' + BW \text{ with } Y = -\frac{W'}{aW}$$

Poles and movable singularity $\rho(u)$ analytic for $u \approx 1$ $\rightarrow$

Gaussian limit with mean and variance $\sim cn$

Varieties of increasing trees $\mathcal{F} = \mathcal{Z}^{\square} \times \phi(\mathcal{F})$ $F_z'(z) = \phi(F(z))$, $F(0) = 0$

Nodes fixed degree : Gaussian limit law with mean and variance $\sim cn$

Conclusion

- Analytic Combinatorics methods :

Combinatorial Decomposability + Strong Analycity + Smooth Singulatity Perturbation	$\Longrightarrow$	Gaussian Laws + Local limits + Large deviations

- Also
  - Gaussian limit laws with analytic perturbation of Saddle-point method and Sachkov Quasi-Powers
  - Discontinuity of singularity for $u \approx 1$ (confluence, ...) $\rightarrow$ non Gaussian continuous limit laws : Rayleigh, Airy, ...

- Beyond the scope of Analytic Combinatorics : functional limit theorems (Probabilistic approach)