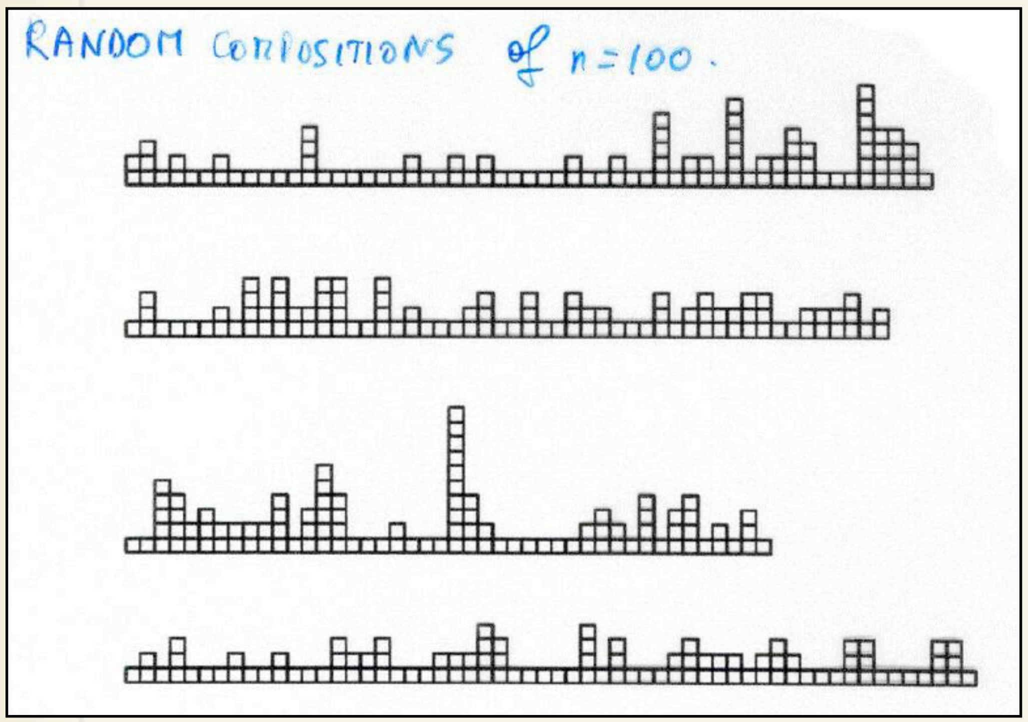
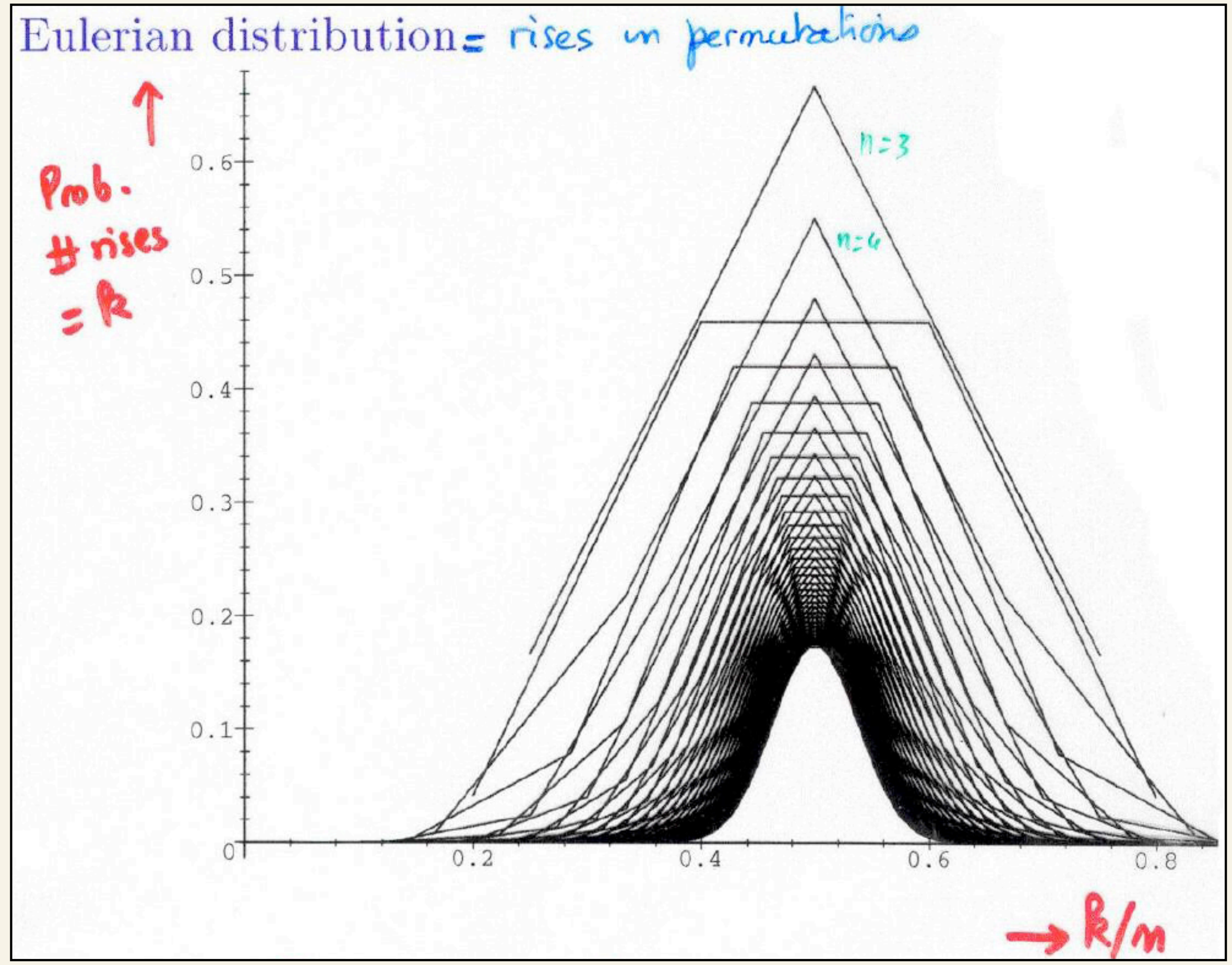
Part C. Random Structures

Chapter 9. Multivariate Generating Functions and Limit distributions

Large random combinatorial structures 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 $\to e^{-t^2/2}$.
- Analytic combinatorics: because of bivariate GF $\frac{1}{1-z(1+u)}$ and *smoothly varying singularity*!

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.

≪Bender, Richmond$^+$.

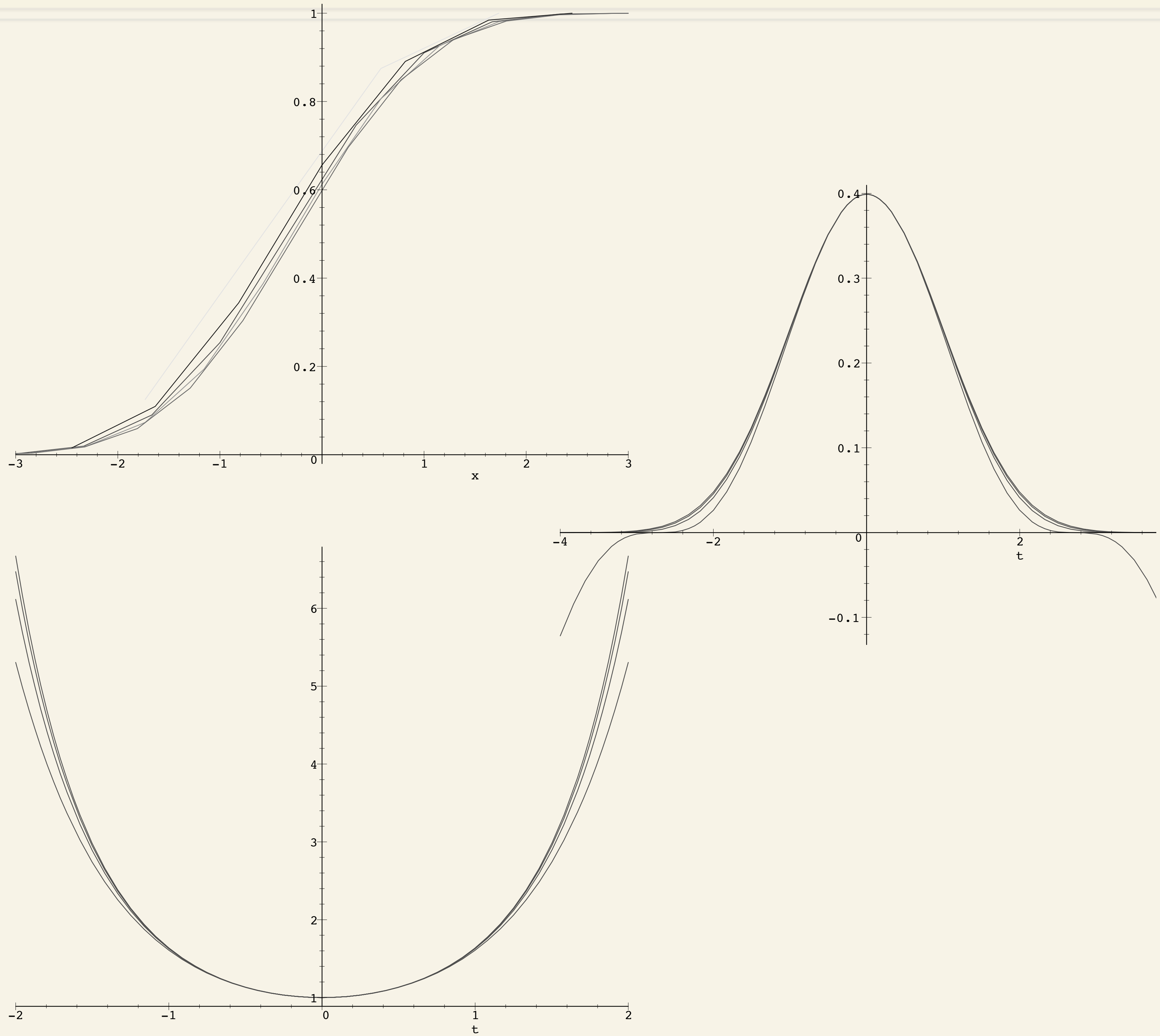
Quasi-Powers Theorem [Bender+Hwang]

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:

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

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: $\mathbb{P}(X_n \leq \mu_n + x\sigma_n) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-w^2/2}\, dw$
- Speed of convergence is $O(\kappa_n^{-1} + \beta_n^{-1/2})$.



Distribution function,

Characteristic function aka Fourier transform,

Moment GF aka Laplace transform

Bender (1973), Bender & Richmond (1983), Hwang (1994)

Quasi-Powers

$(\approx A(u)B(u)^{\beta_n})$

$\downarrow$

Meromorphic[1,2] Alg-log[2] Exp-log[2]

---

[1] Analysis of meromorphic functions

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

$$[z^n]f(z) = \text{Residue} + \text{exponentially small}$$

[2] Singularity analysis

$$f(z) \approx (1-z)^{-\alpha}$$

$$[z^n]f(z) \approx n^{\alpha-1}$$

The methods are *uniform* w.r.t. parameters

- All Gaussian laws eventually based on perturbation of singularities

Supercritical sequences

Example 1.

Let $\mathcal{F} = \textsc{Seq}(\mathcal{G})$, 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.*

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 perms**

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

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

"Assemblies of logarithmic structures"

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 $I_n \sim \frac{q^n}{n}$.

- 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.

Functional equations and limit laws

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 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 $\left\{ \begin{array}{l} \text{by singularity analysis} \\ + \text{Quasi Powers} \end{array} \right.$

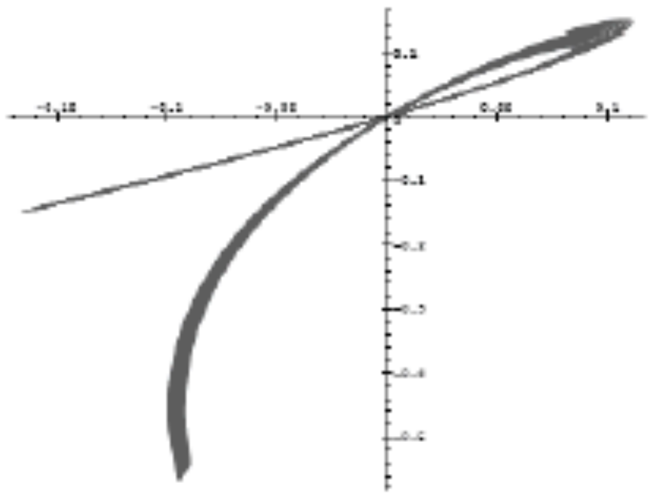
Perturbation of rational functions

— 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.

≈ 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.

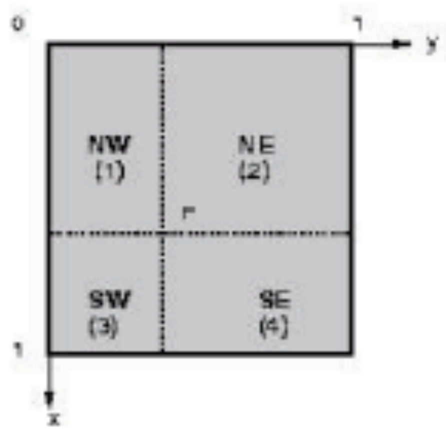
$$G^3 + (2z^2 - 3z - 2)G^2 + (3z+1)G = 0$$
$$G^3 + (2u^3z^2 - 3u^2z + u - 3)G^2 + (3u^2 - 2u + 3)G + u - 1 = 0$$

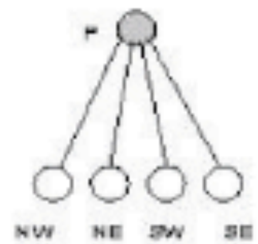
Movable singularity scheme applies: $\mathcal{N}$ormality.

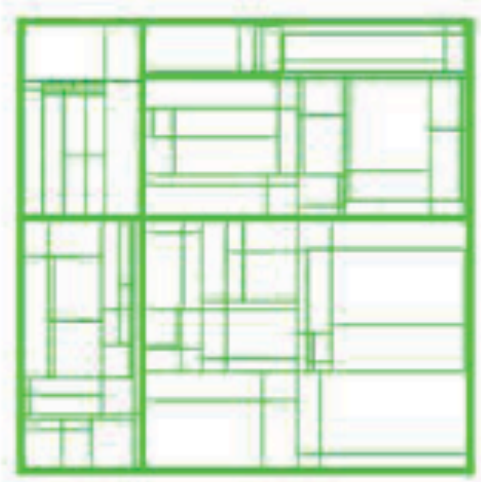
+ Patterns in context-free languages, in combinatorial tree models, in functional graphs: Drmota's version of Drmota-Lalley-Woods.

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*.

Other Properties

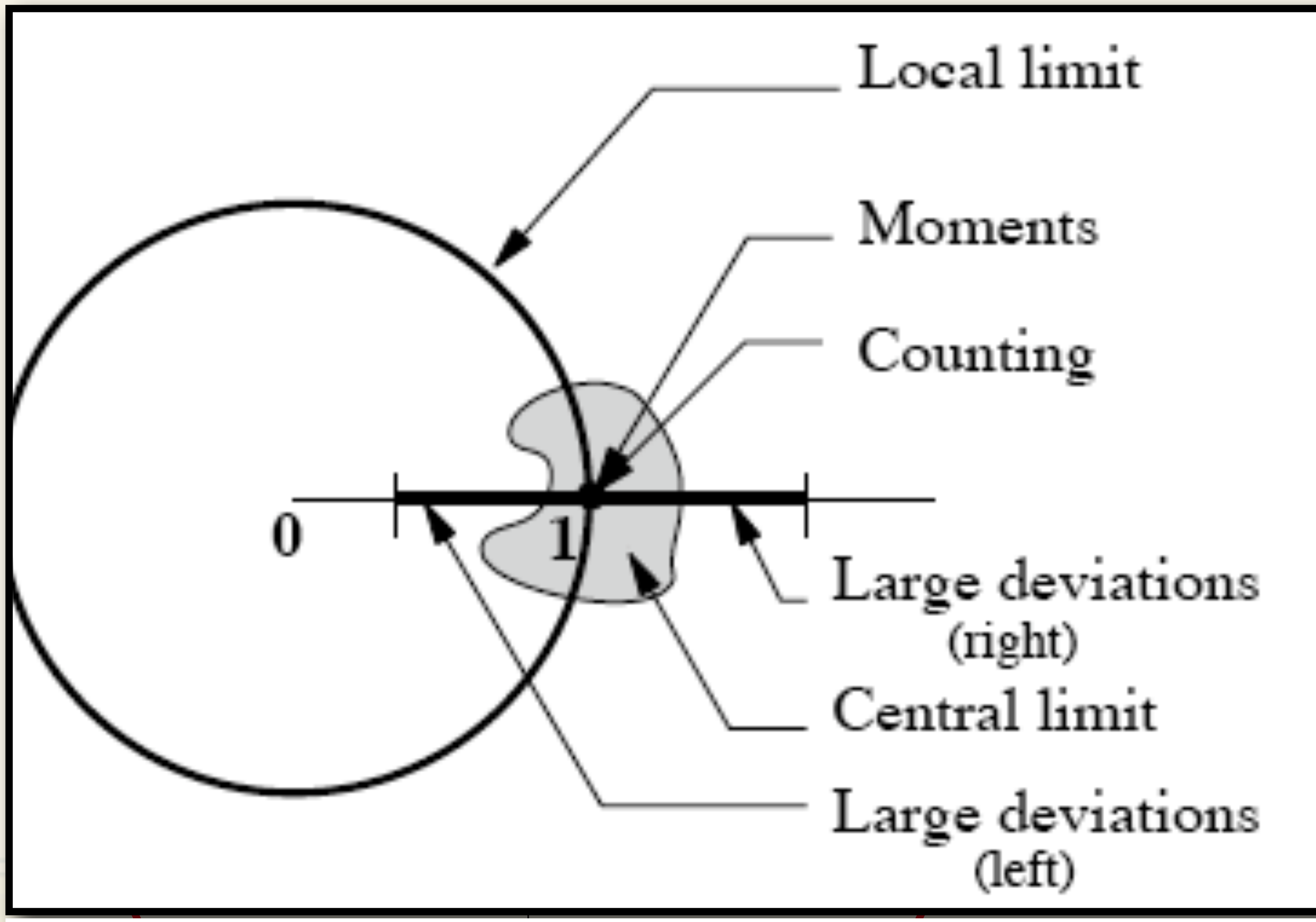
Local Limit Theorem. *If the Quasi-Powers approximation holds on the circle* $|u| = 1$, *then*

$$\sigma_n \Pr\{X_n = \mu_n + x\sigma_n\} \to \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Large deviations. *If the Quasi-Powers approximation holds on an interval containing 1, then*

$$\frac{1}{\beta_n} \log p_{n, x\beta_n} \leq W(x) + O(\beta_n^{-1}).$$

Counting	$u = 1$
Moments	$u = 1 \pm \frac{1}{\infty}$
Large deviations	$u = [1 - \eta, 1 + \eta]$
Central limit	$u = 1 + \Box$
Local limit	$\|u\| = 1$



Rises in permutations

EXAMPLE. Eulerian Numbers = Rises in permutations.

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

$$\rho = 1 \qquad \rho = \frac{\log u}{u-1}[+\frac{2ik\pi}{u-1}]$$

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

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

Uniformly

$\Longrightarrow$ Gaussian Law, $\mu_n \sim \frac{1}{2}n$, $\sigma_n^2 \sim \frac{1}{12}n$.

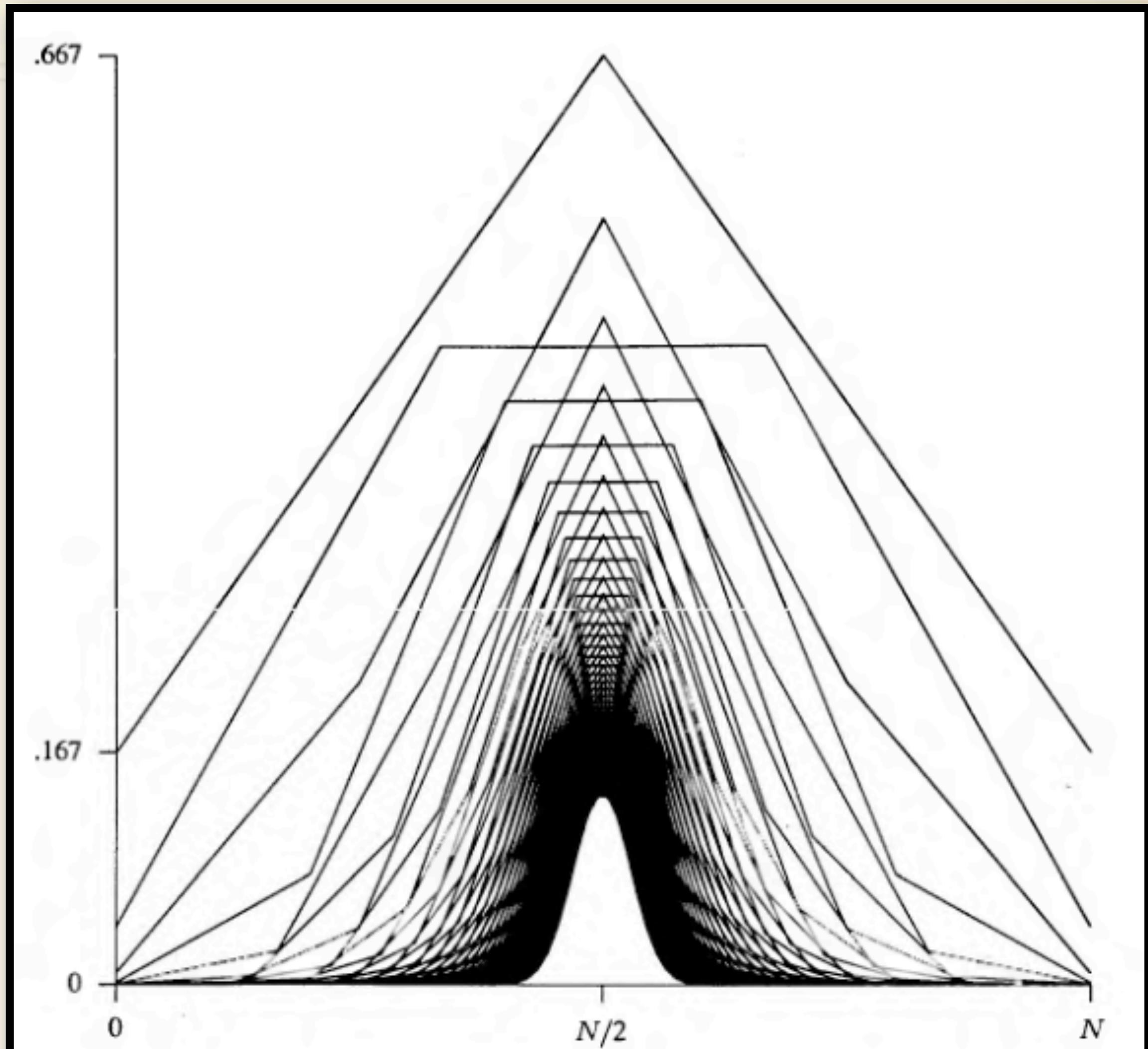
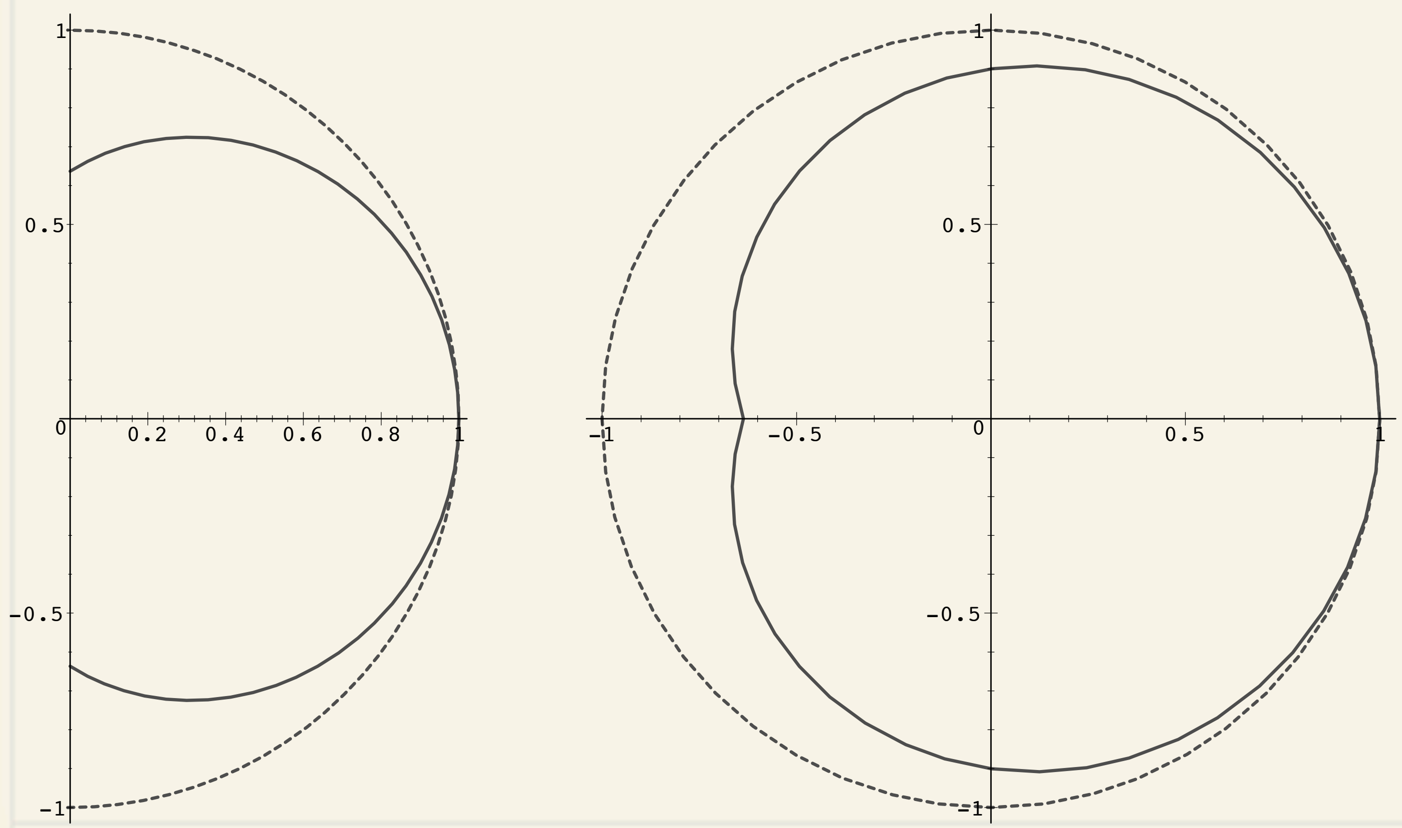


Figure 6.5 Distribution of runs, $3 \leq N \leq 60$ (k-axes scaled to N) (Eulerian numbers)

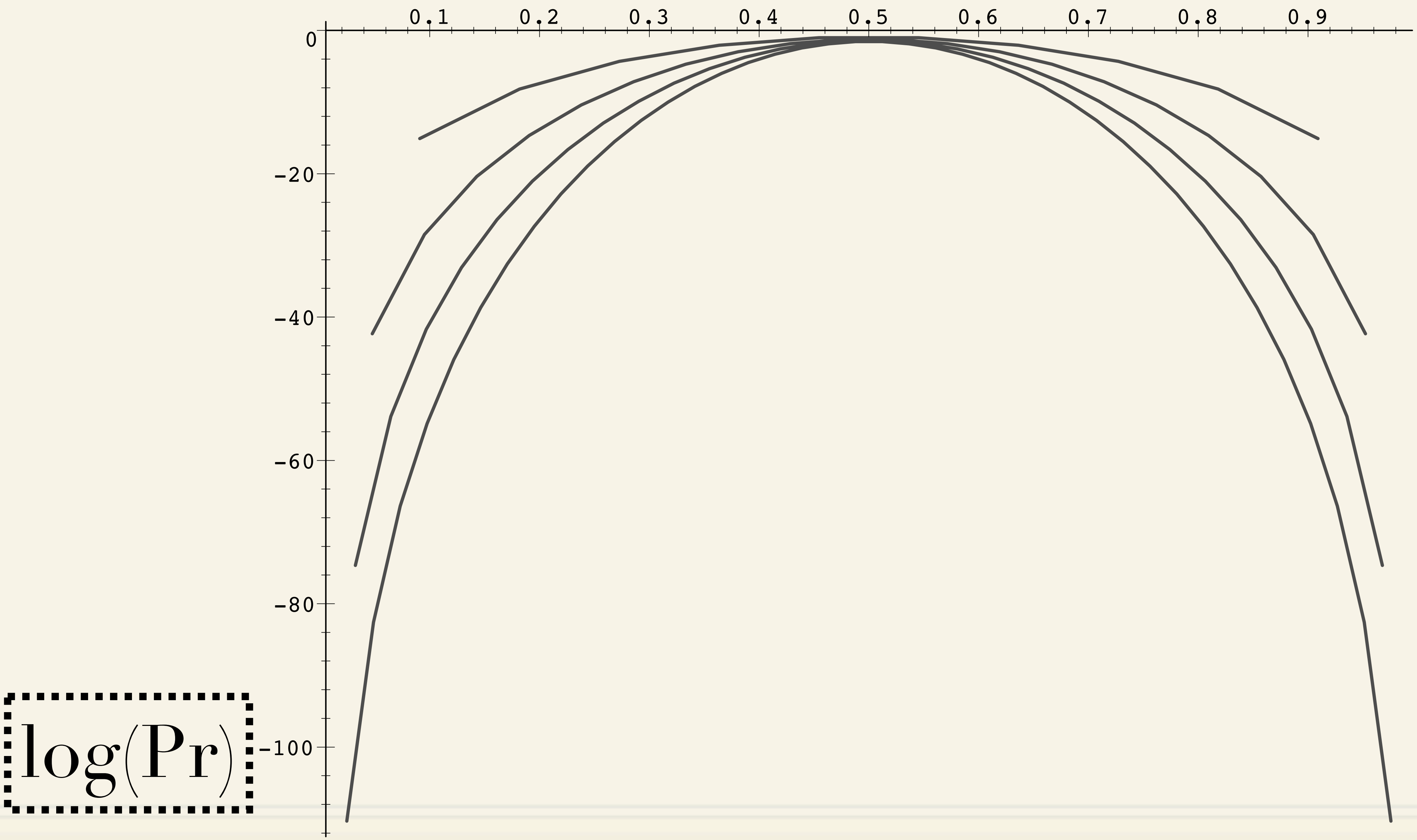
Eulerian distribution,

$$\rho = \frac{\log u}{u-1}\left[+\frac{2ik\pi}{u-1}\right]$$



Runs in perms: mean $\sim n/2$. Proba of deviation by more than 10% from mean is

1,000	10,000	100,000
10^{-6}	10^{-65}	10^{-653}



Non-Gaussian Laws

. Use singularity diagrams

$$F(z,u) = \frac{1}{1-zu} \cdot \frac{1}{1-z}$$

$u = 1 - \epsilon$	$u = 1$	$u = 1 + \epsilon$
$\rho(u) = 1$	$\rho(u) = 1$	$\rho(u) = 1/u$
Z^{-1}	Z^{-2}	Z^{-2}

$(\star)$ $Z = \rho(u) - z$

The double saddle point

Non-Gaussian Laws

$$I(n) = \frac{1}{2i\pi} \int e^{h_n(z)}\, dz$$

double saddle point at $\zeta \equiv \zeta_n \; : \; h_n'(\zeta) = h_n''(\zeta) = 0$

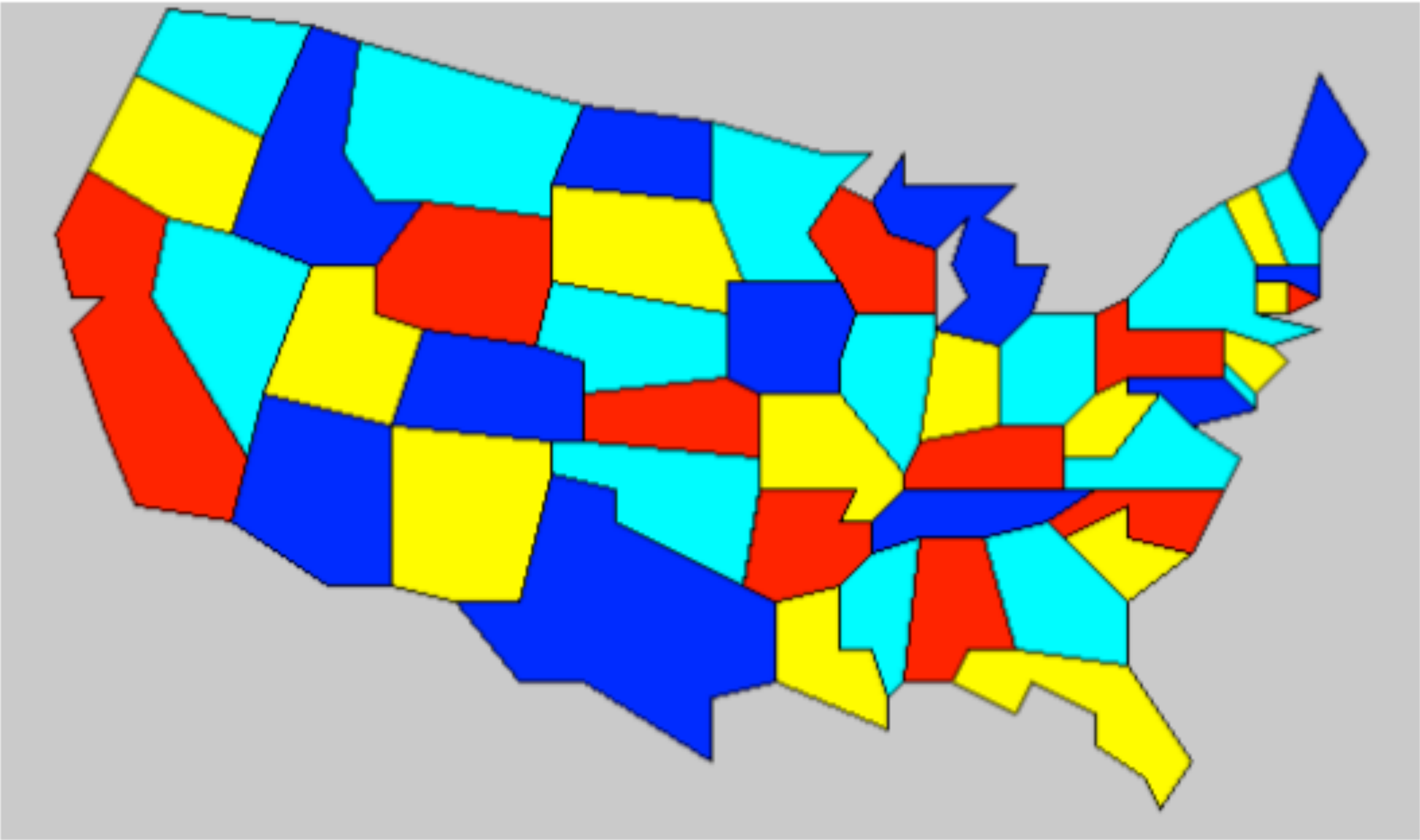
via: $e^{h_n(\zeta)} \int \exp(-\frac{t^3}{3!} |h_n'''(\zeta)|\, dt$

$$I(n) \sim \frac{1}{2\pi} \left((3!)^{1/3}\Gamma(1/3)\right) \frac{1}{\sqrt[3]{h_n'''(\zeta)}} e^{h_n(\zeta)}.$$

Normalization to:

Exp ∘ Cubic; scales $\sqrt[3]{n}, \ldots$

MAPS



Maps are planar graphs with an embedding in the plane (or sphere) plus a *rooted* edge.

Consider:

- Maps $\mathcal{M}$ of sorts;
- Core maps $\mathcal{C} \subset \mathcal{M}$ with higher connectivity;

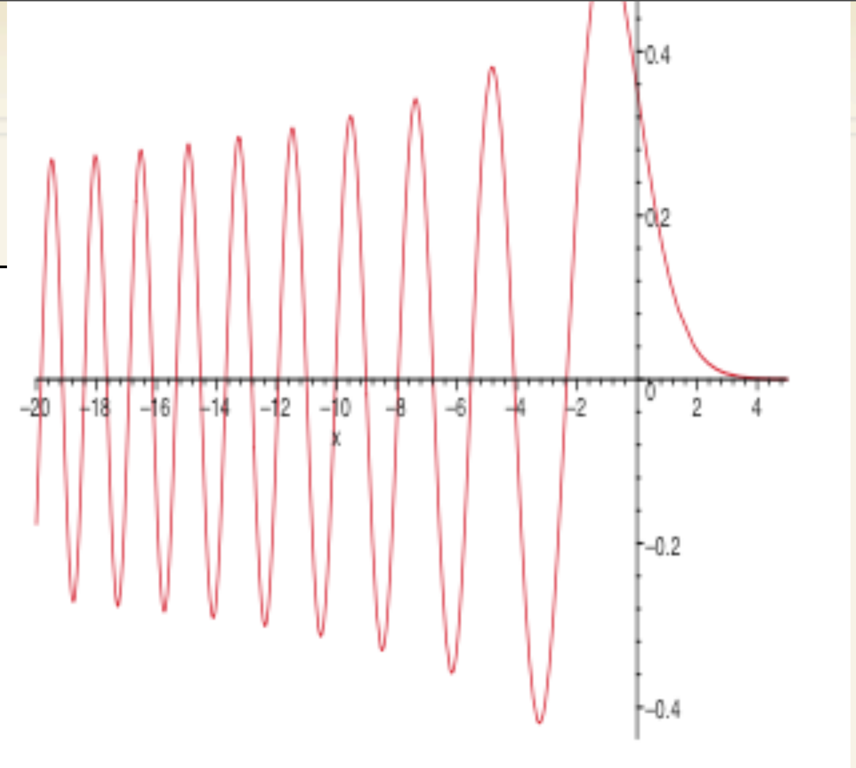
$\mathcal{C}$: class of 2-connected loopless maps.

$$M(z) = C(zM(z)^2)$$

$$\Pr(X_n = k) = C_k \frac{[z^n]M(z)^k}{[z^n]M(z)}.$$

Airy function

A solution of $y'' - zy = 0$



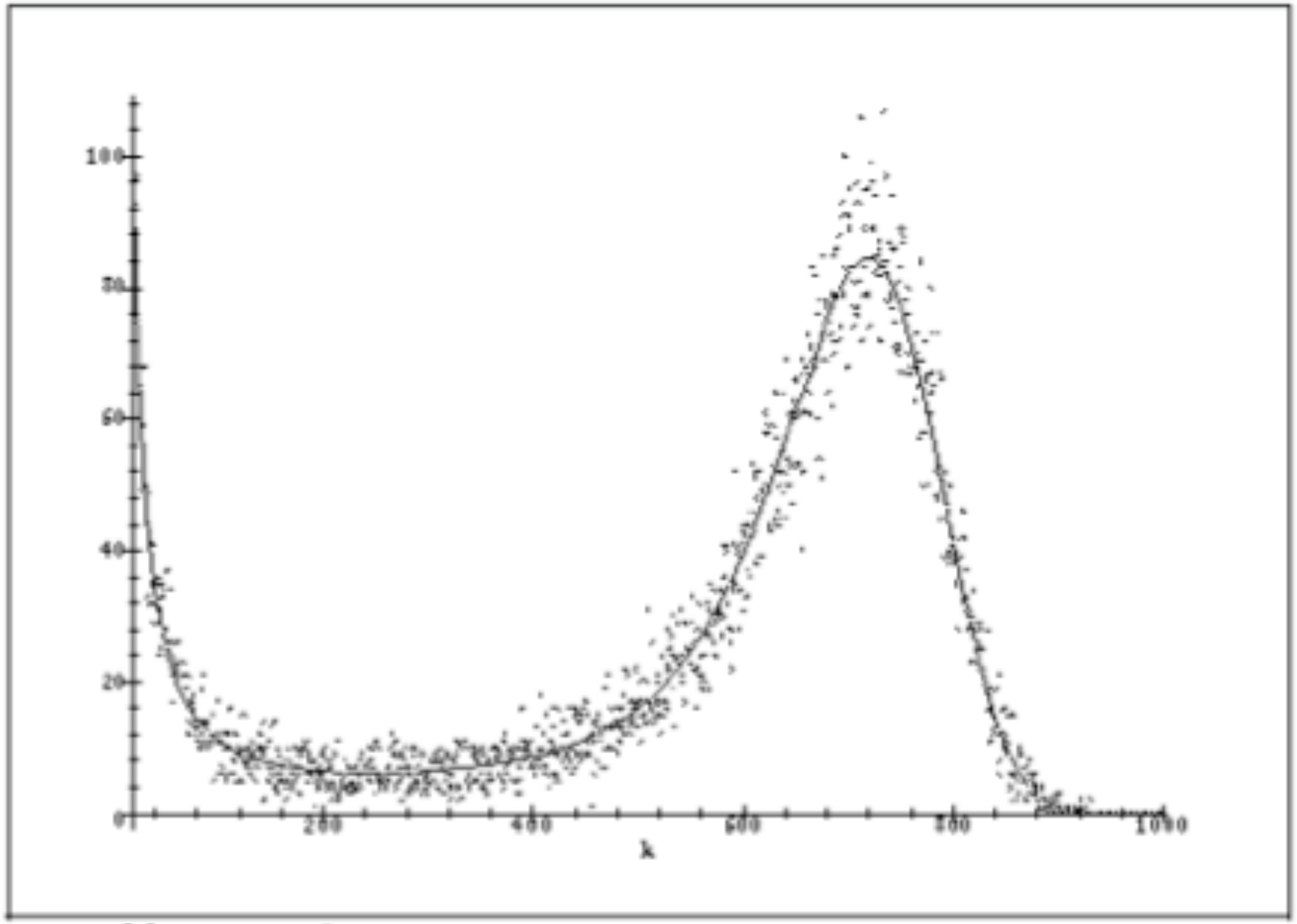
$$\mathrm{Ai}(z) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{i(zt+t^3/3)}\,dt = \frac{1}{\pi 3^{2/3}}\sum_{n=0}^{\infty}\frac{\Gamma((n+1)/3)}{n!}\sin\left(\frac{2(n+1)\pi}{3}\right)\left(3^{1/3}z\right)^n$$

Core size in maps [BaFlScSo]

$$\mathcal{A}(x) = 2\exp\left(-\frac{2}{3}x^3\right)\left(x\mathrm{Ai}(x^2) - \mathrm{Ai}'(x^2)\right).$$

$$\Pr(X_n = k) \sim K n^{-2/3}\mathcal{A}\left(\frac{3^{4/3}}{4}x\right), \qquad k = \frac{n}{3} = xn^{2/3}.$$

The distribution of core size



♡ The Random Graph with n vertices, m edges has phases

Gas ($m = o(n)$); Liquid ($m \sim \frac{n}{2}$);
Solid ($m \sim \frac{1}{2} n \log n$); Hypersolid ($m \gg n \log n$)

At critical stage $m \sim \frac{n}{2}$, RG is "almost" a forest of unrooted trees and unicyclic graphs.

$$W_{-1} = T - T^2/2, \quad W_0 = \frac{1}{2} \log \frac{1}{1-T} - \frac{T}{2} - \frac{T^2}{4}.$$

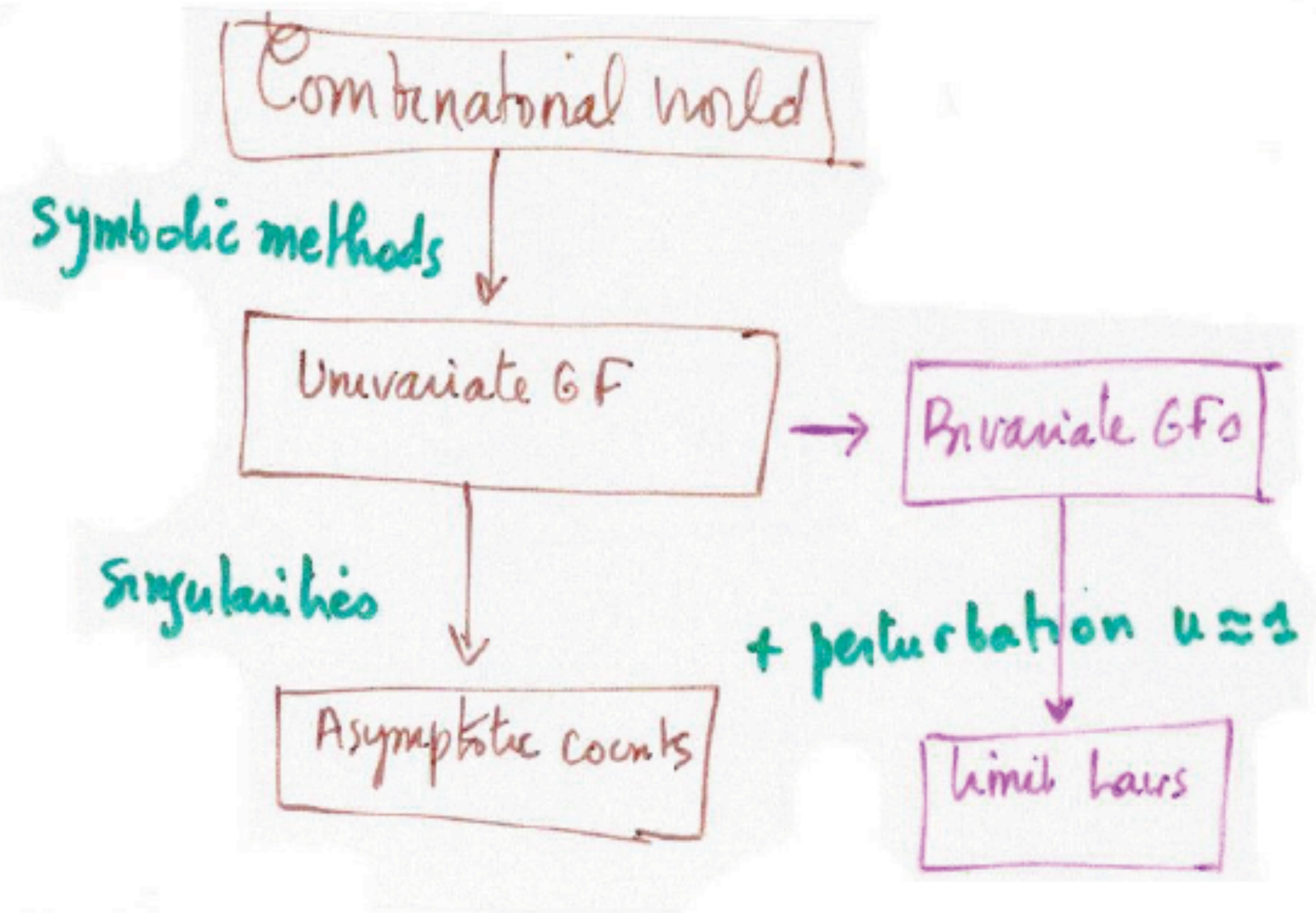
Airy functions have been observed in Fl-Knuth-Pittel and the "Giant paper" by Janson-Knuth-Łuczak-Pittel!

+[F., Salvy, Schaeffer 2006]

♡ There is another class of Airy-related distributions:

Theorem. [Louchard; Takács; Fl-Poblete-Viola] Area below excursions, path length of trees, and displacement in linear probing hashing all converge to a law determined by its moments as (roughly)

$$MGF \approx \frac{\mathrm{Ai}'(z)}{\mathrm{Ai}(z)}$$



That's All, Folks!