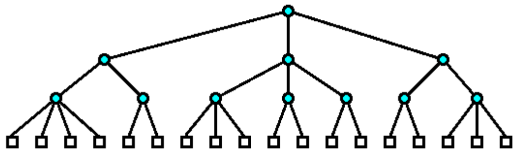
THE UBIQUITOUS GAUSSIAN LAW

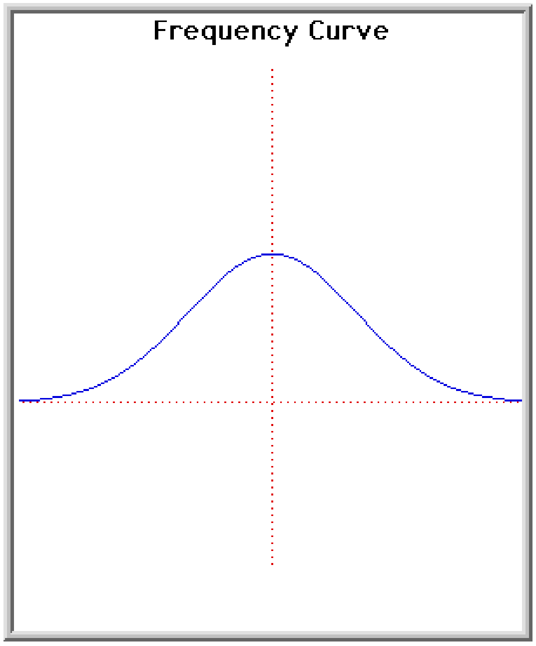
IN ANALYTIC COMBINATORICS

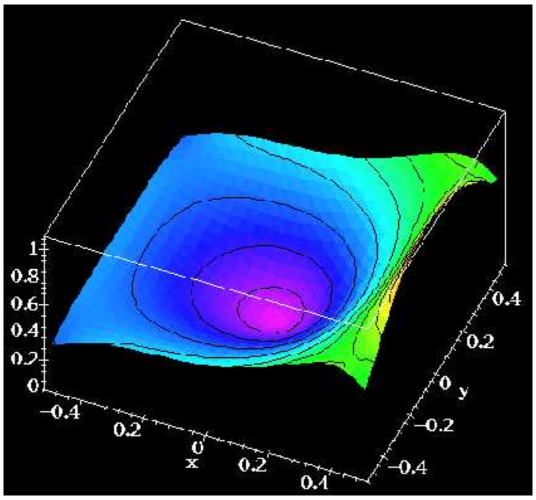
Philippe Flajolet

INRIA, Rocquencourt

with Hsien Kuei HWANG, Michèle SORIA
+ Bob SEDGEWICK





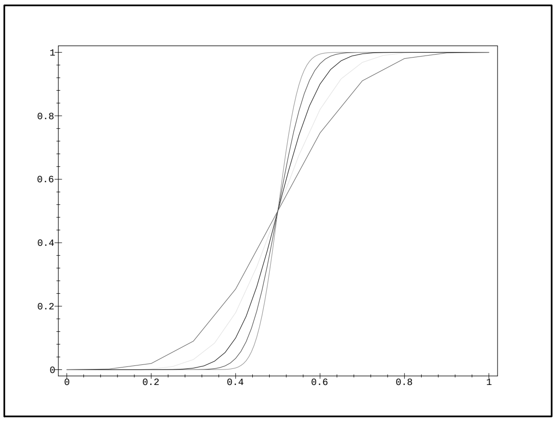


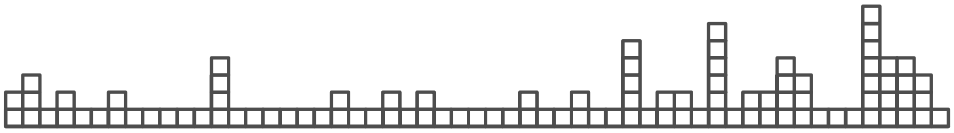
Random errors are inevitable in instrumental observations, measurements are equally prone to err in one direction or the other,
and the majority of observations tend to center around mean values
— Galileo Galilei [1564–1642]

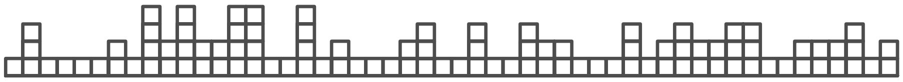
Convergence in probability.

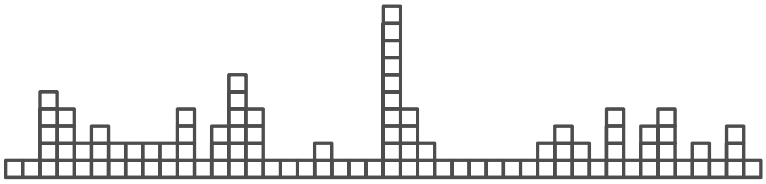
$$\Pr\left\{\left|\frac{X}{\mu}-1\right|>\epsilon\right\}\to 0.$$

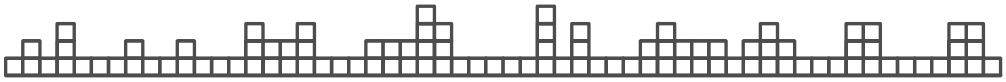
$$F(z,u)=\cfrac{1}{1-u\cfrac{z}{1-z}}$$

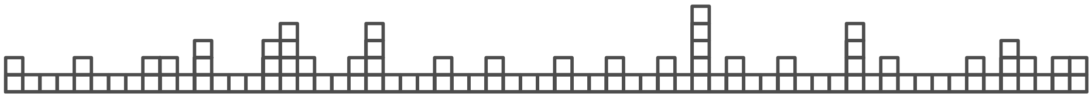


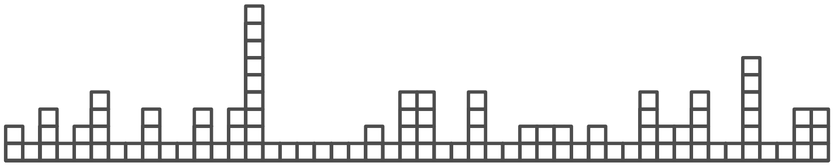


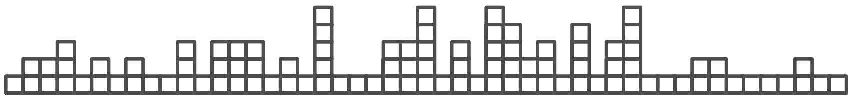


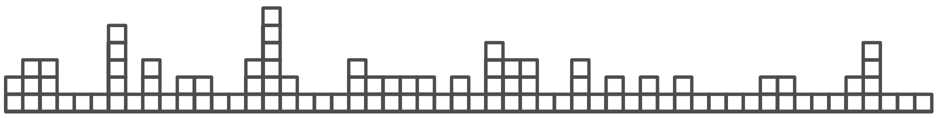


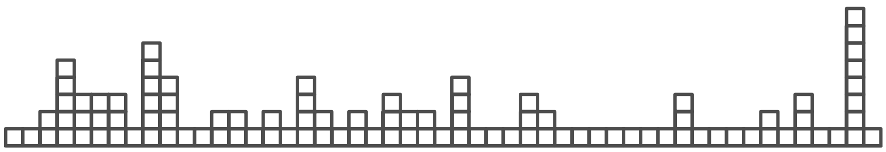












De Moivre, Laplace, Gauß

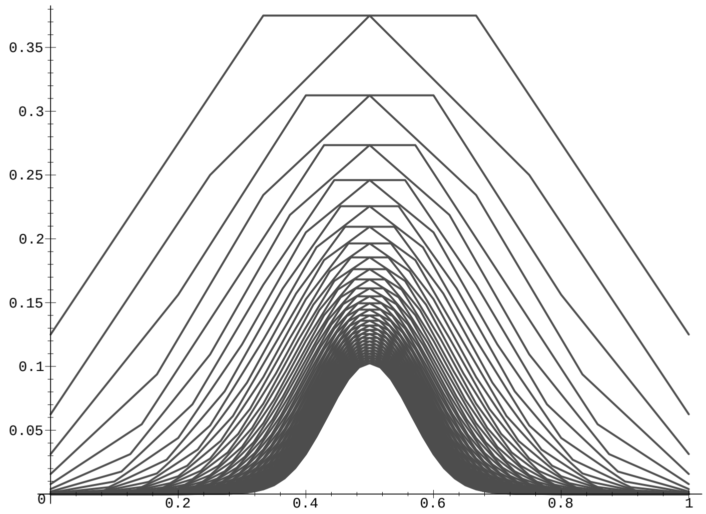
- Central Limit Theorem

Let $S_n = X_1 + X_2 + \cdots + X_n$ *with* $E\{X\} = \mu$, $Var\{X\} = \sigma^2$, *then*

$$\Pr\left\{\frac{X_n - n\mu}{\sqrt{n}\sigma} \le x\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-w^2/2}\, dw$$

- Local Limit Theorem.

The probability density of S_n *tends to*



$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

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

Distribution functions (cumulative)

$$F_X(x) := \Pr\{X \le x\}$$

Characteristic functions (Fourier transform)

$$\phi_X(t) := E\{e^{itX}\} \equiv \int_{-\infty}^{\infty} e^{itx}\, dF_X(x)$$

Continuity Theorem. [Lévy]

$$F_n \Rightarrow F \qquad \Longleftrightarrow \qquad \phi_n \to \phi$$

E. g., Central Limit Theorem

X has char. fn. $\phi(t)$; assume 0 mean and unit variance.

$S_n = X_1 + \cdots + X_n$ has char. fn. $\phi(t)^n$

$S_n^\star = \frac{1}{\sqrt{n}} S_n$ (normalized) has char. fn.

$$\begin{aligned} \phi(\frac{t}{\sqrt{n}})^n \quad &= \quad \left(1 - \frac{t^2}{2n} + \cdots\right)^n, \\ &\sim \quad e^{-t^2/2} \end{aligned}$$

ROBUSTNESS "Quasi-Powers"

Univariate Asymptotics

$$\begin{array}{c}\text{Combinatorial objects}\\ \downarrow\\ \text{Univariate generating Function}\\ \downarrow\\ \text{Singularity Asymptotics}\end{array}$$

Surjections from $[1\,..\,n]$ to $[1\,..\,k]$ (some k)

$$F(z) = \frac{1}{1-(e^z-1)} \equiv \frac{1}{2-e^z}$$

♡ Singularity analysis

$$[z^n]\,F(z) \approx \rho^{-n}, \qquad \rho = \log 2$$

♡ PHILOSOPHY

Combinatorial Constructions
$\mapsto$ Functionals
$\mapsto$ Singularity Transformers

Dictionary I. (Symbolic methods) + Pólya Op's

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

Dictionary II. (Singularity analysis)

$$f(z) \sim C \cdot (1-z)^{-\alpha} \log^k \frac{1}{1-z}$$

$$\Longrightarrow \quad [z^n] f(z) \sim C \cdot \frac{n^{\alpha-1}}{\Gamma(\alpha)} \log^k n$$

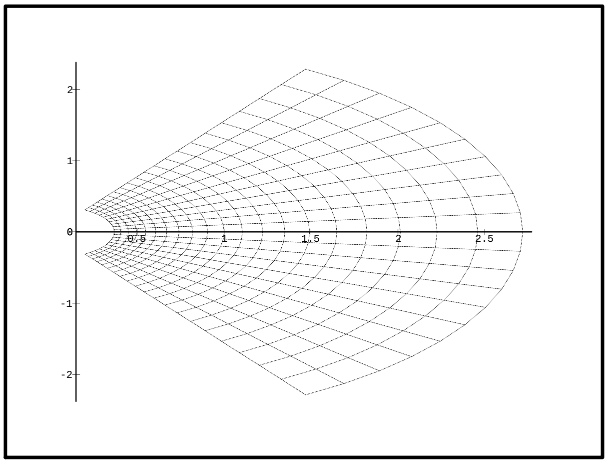
COMPLEX ASYMPTOTICS

Dictionary (II)

Point of regularity.

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

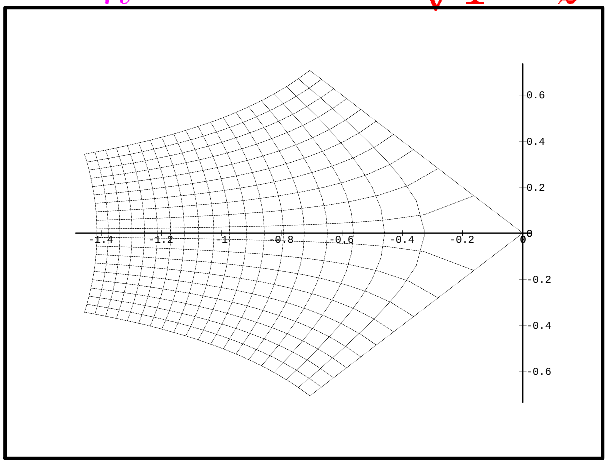
$\exp(z)$

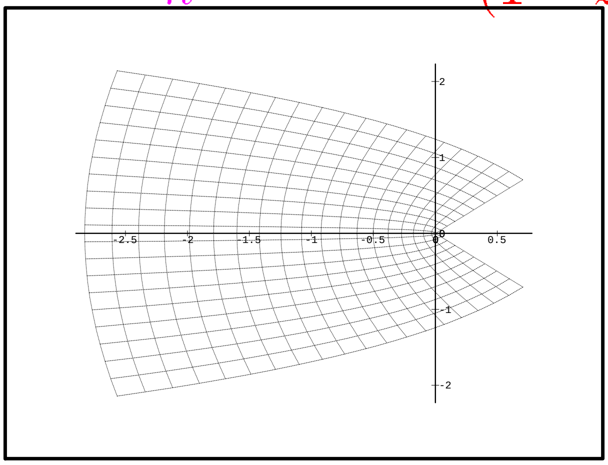


Point of singularity.

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

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





Multivariate Asymptotics

$$\begin{array}{c}\text{Combinatorial objects}\\ \downarrow \\ \text{Bivariate generating Function}\\ \downarrow \\ \text{Singularity perturbation Asymptotics}\end{array}$$

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

♡ Singularity analysis

$$[z^n]\,F(z,u) \approx \rho(u)^{-n}, \qquad \rho(u) = \log(1+\frac{1}{u})$$

♣ Fourier inversion (continuity theorem for characteristic f.)

$$[z^n u^k]\,F(z,u) \approx \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{w^2/2}\,dw$$

1 The Quasi-Powers Framework

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

The Quasi-Powers Theorem. *The PGF of a discrete random variable X_n satisfies*

$$g_n(u) \equiv \frac{f_n(u)}{f_n(1)} = A(u)B(u)^{\beta_n}\left(1 + \emptyset(\frac{1}{\kappa_n})\right)$$

uniformly in $u \approx 1$, with analytic $A(u), B(u)$.

Then

$$E\{X_n\} = \mu_n = \beta_n B'(1) + A'(1) + \emptyset(\frac{1}{\kappa_n})$$

$$Var\{X_n\} = \sigma_n^2 = \beta_n(B''(1) + B'(1) - B'(1)^2) + \cdots$$

$$\Pr\{X_n < \mu_n + x\sigma_n)\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-w^2/2}\, dw.$$

Speed of convergence is

$$\emptyset(\kappa_n^{-1} + \beta_n^{-1/2}).$$

Proof.

1) The moments are obtained by differentiation. Permissible for analytic functions, given uniformity assumption.

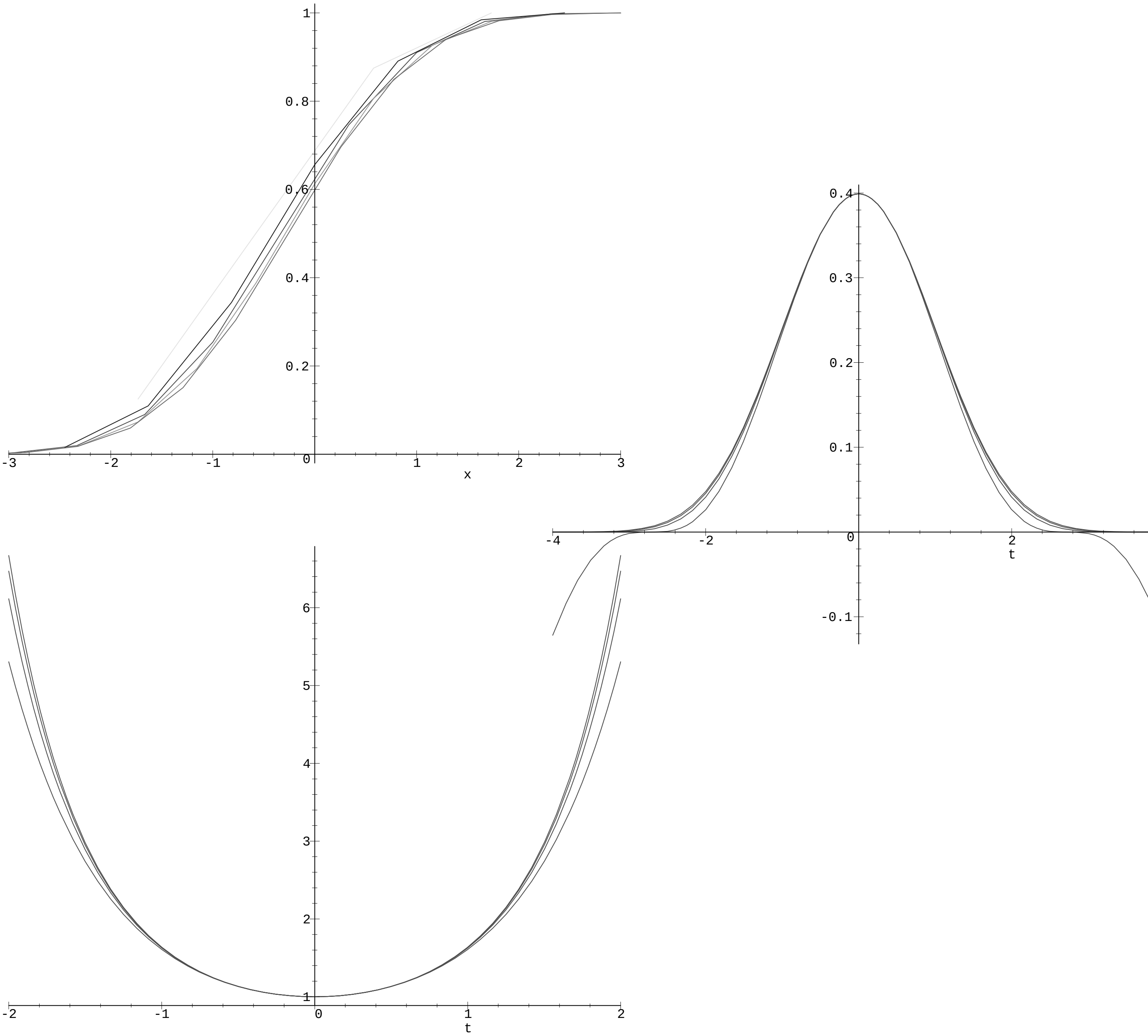
2) The characteristic functions obtains by normalization and simple algebra.

3) Like in Central Limit Theorem, only local properties near $u = 1$ are needed, since $\beta_n \to +\infty$.

4) Simple computation a la CLT yield limit characteristic function, hence Gaussian law by continuity.

5) Berry-Esseen inequalities provide "constructive" version of continuity theorem, hence speed of convergence.

$$||F - G||_\infty \le c_1 \int_{-T}^{+T} \left| \frac{\phi(t) - \gamma(t)}{t} \right| dt + c_2 \frac{||G'||_\infty}{T}.$$



Distribution function,

Characteristic function aka Fourier transform,

Moment GF aka Laplace transform

Other Properties

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

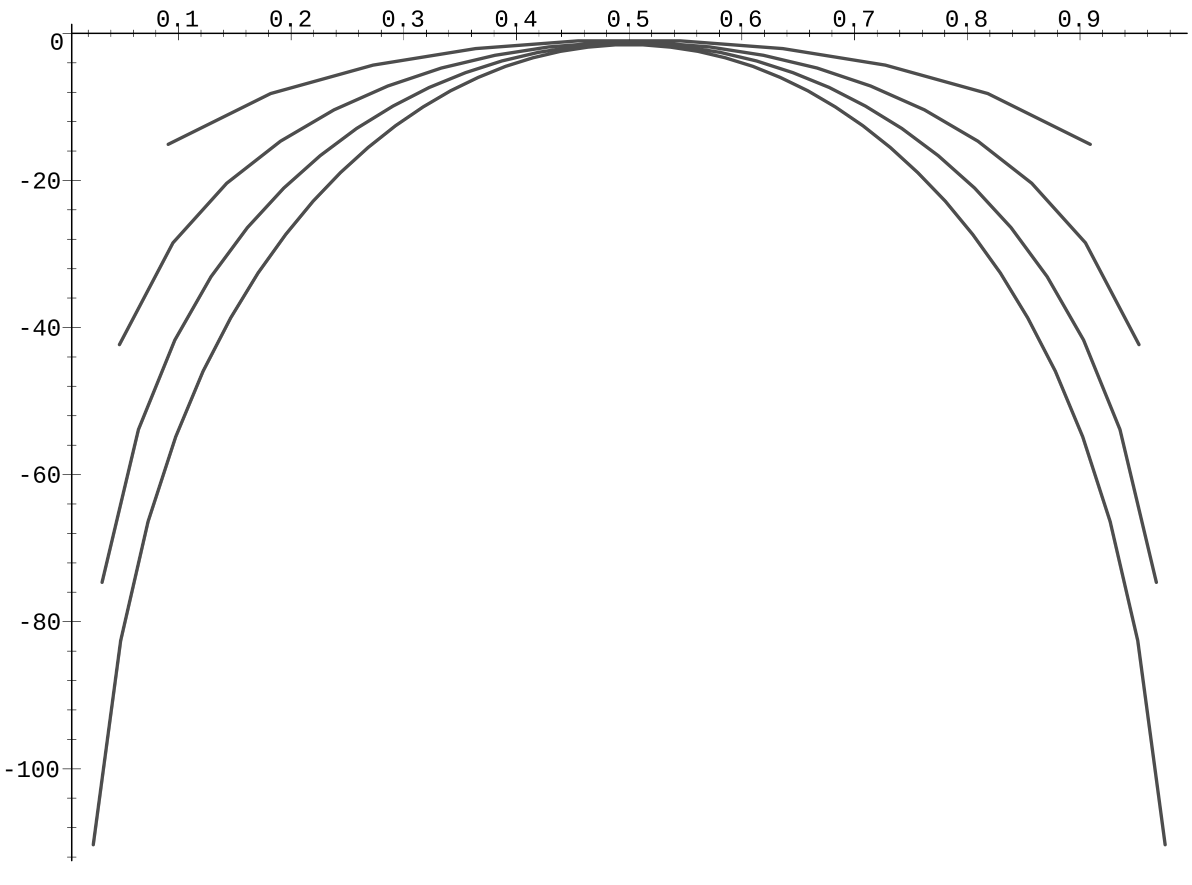
$$\sigma_n \Pr\{X_n = \mu_n + x\sigma_n\} \to \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

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

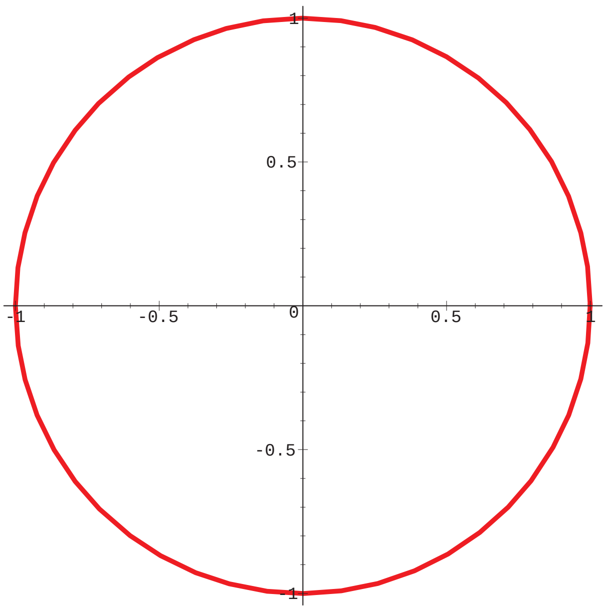
$$\frac{1}{\beta_n} \log p_{n,x\beta_n} \leq W(x) + \emptyset(\beta_n^{-1}).$$

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}



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



2 Three Bivariate Analytic Schemes

Bivariate Analyticity

$\downarrow$

Quasi-Powers

$\downarrow$

Gaussian Laws

PRINCIPLE. A certain ***type of singularity*** of $F(z,1)$ usually ***persists*** for $F(z,u)$, when $u \approx 1$.

Implies a Quasi-Powers approximation.

EXAMPLE. Eulerian Numbers = Rises in permutations.

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

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

$$\begin{aligned} [z^n]F(z,u) &= \frac{1}{2i\pi}\int_{|z|=1/2} F(z,u)\,\frac{dz}{z^{n+1}} \\ &= \rho(u)^{-n-1} + O(2^{-n}) \end{aligned}$$

Uniformly

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

Meromorphic Scheme Theorem.

$$F(z,u) = \frac{B(z,u)}{C(z,u)}, \qquad u \in \mathcal{V}(1),\ z \in \mathcal{V}(0).$$

under the conditions

(i) $C(\rho,1)=0$, $B(\rho,1)=1$, ρ s *simple & single.*

(ii) B, C *are analytic in* $u \in \mathcal{V}(1)$, $|z| < \rho + \epsilon$.

then a NORMAL LAW holds under variance condition,

$$\left.\frac{d^2}{ds^2}\rho(e^s)\right|_{s=0} \neq 0.$$

Bender 1973, F+Soria 1993, Hwang 1994 (D.S.), B. Vallée 1995 (Operators).

Also: multiple poles, logarithmic terms, etc.

Meromorphic Scheme

1) De Moivre + weak CLT: $F(z,u) = \dfrac{1}{1 - z(1+u)/2}$ or

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

2) Critical ***Sequences***. $\dfrac{1}{1 - uf(z)}$, $f(\rho) > 1$.

   – Compositions $F(z,u) = \dfrac{1}{1 - u\sum_{s\in S} z^s}$.

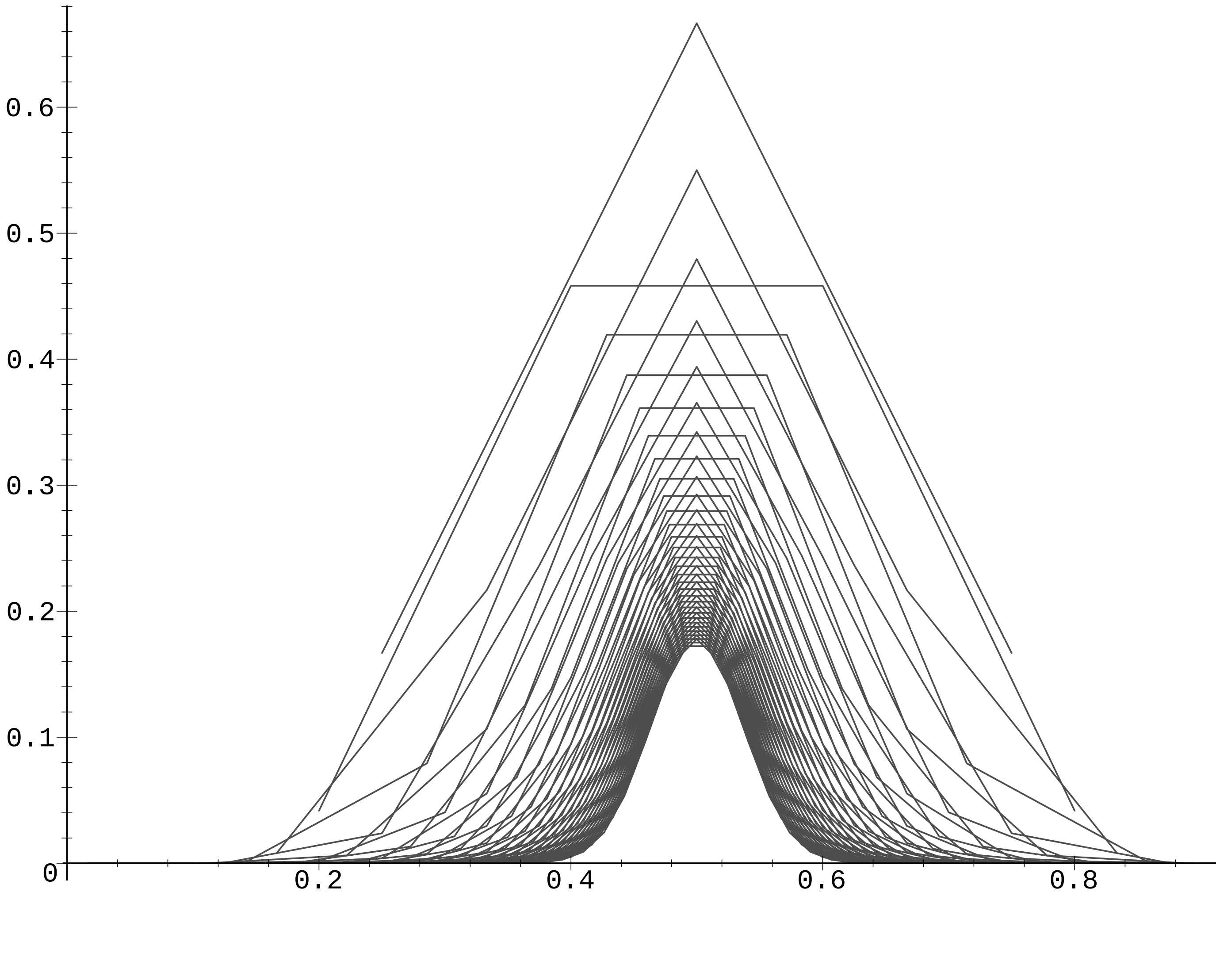
   – Preferentials arrangements, Alignments, etc.

3) Runs in permutations.

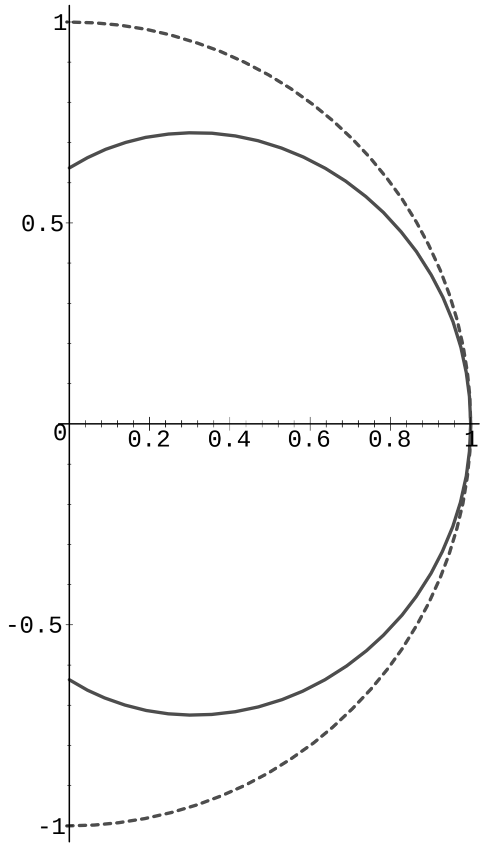
4) Patterns in strings $F(z,u) = \mathrm{Rat}(z,u,c(z))$.

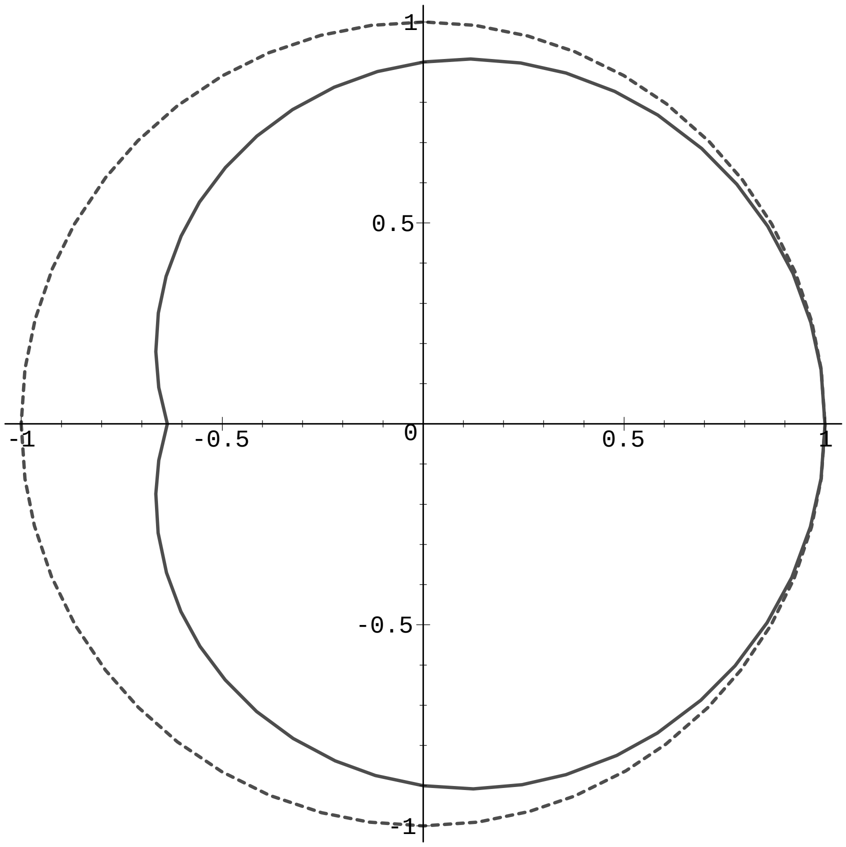
5) Stack polyominos, coin fountains, etc.

Eulerian distribution.

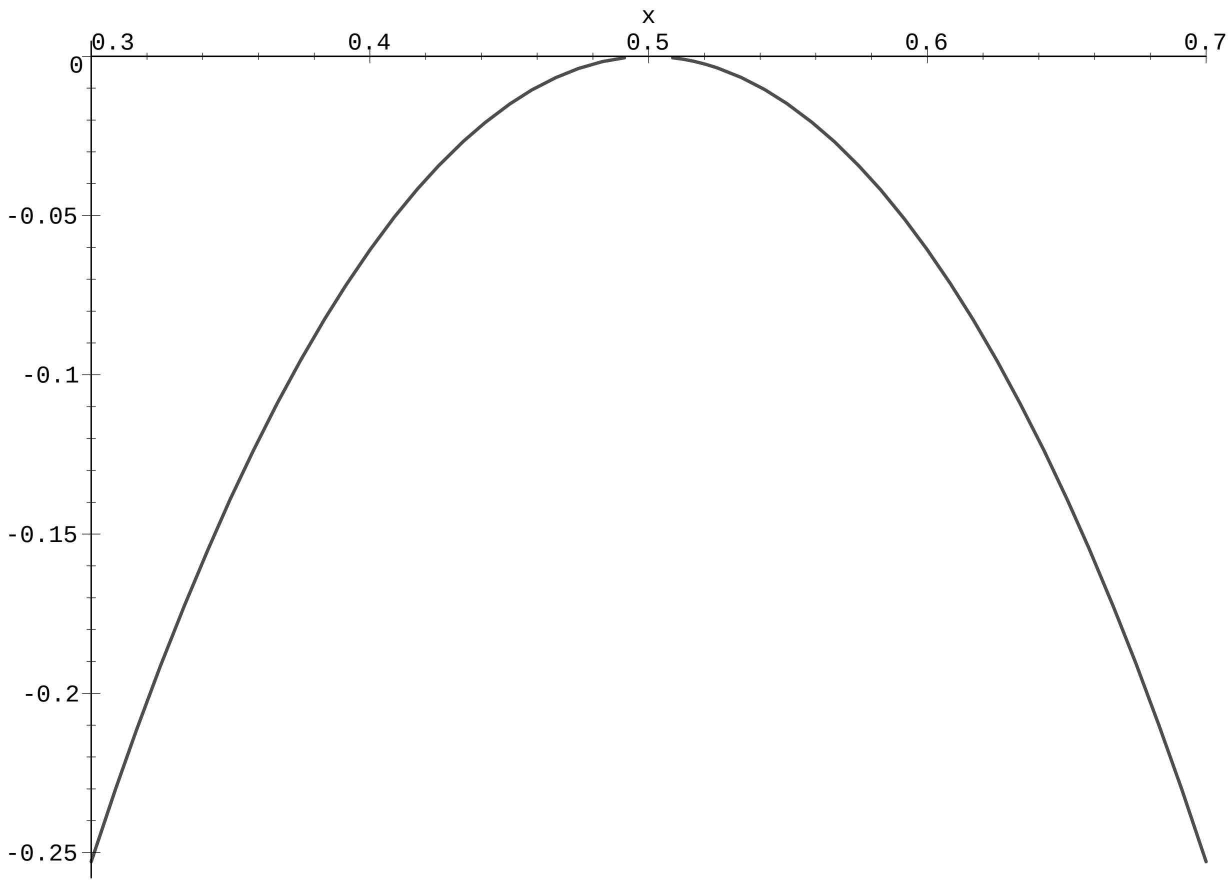


Eulerian distribution, $\rho(u)$.



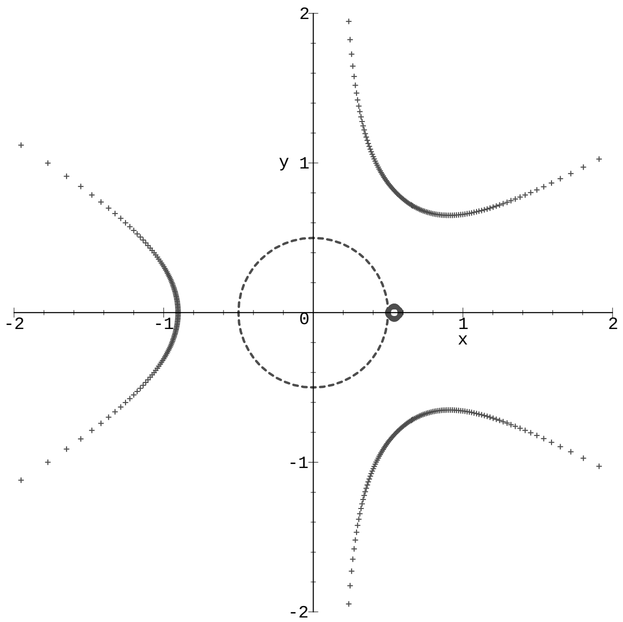


Eulerian distribution, Large Deviations



PATTERNS

Singularities of the BGF $F(z, u)$ for $|u| = 1$



Alg-Log Scheme

EXAMPLE. Leaves in Catalan trees

$$F = zu + \frac{zF}{1-F}$$

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

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

$$\times \boxed{\tfrac{1}{2}\sqrt{1-z(1+u)^2}}$$

$$\rho(u) = \frac{1}{(1+u)^2}, \qquad g_n(u) \sim A(u)(1+u)^{2n} + s.o.t$$

Based on uniformity of singularity analysis methods.

Alg-Log Scheme Theorem

$$F(z,u) = A(z,u) + B(z,u)C(z,u)^{-\alpha}$$

$\alpha(u)$ regular at $u = 1$.

$F(z,1)$ *has a single dominant alg-log singularity*:
$A(z,u), B(z,u), C(z,u)$ *are analytic in* $u \in \mathcal{V}(1)$, $|z| < \rho + \epsilon$.

$\Longrightarrow$ A Gaussian Law.

Proof. Quasi-Powers approximation, plus Weierstrass Preparation Theorem $\rho(u)$.

F+Soria 1993, Drmota.

Applications

1) Leaves in Catalan trees.

2) Patterns in binary Catalan trees [FSS], independent sets (Drmota)

3) Characteristics of random walks or gambler ruins.

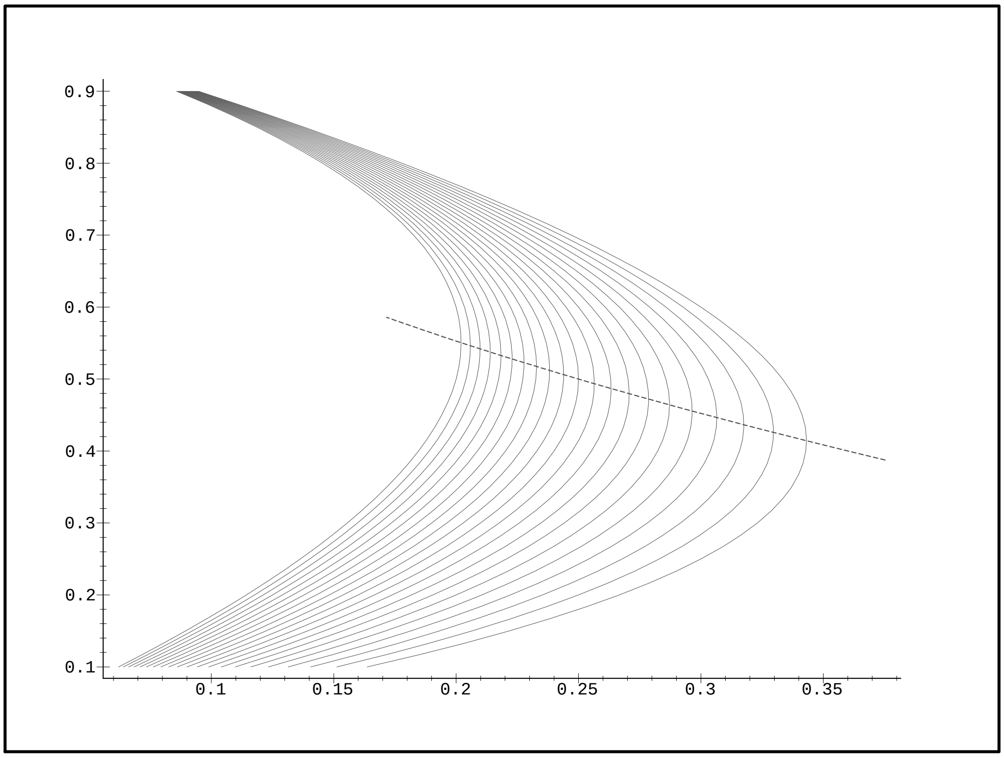
4) Extends to Cayley trees $T(z) = ze^{T(z)}$.

5) Extends to Pólya operators, e.g, Otter trees.

6) Extends further to many "Context-free" examples and recursive structures (later!)

LEAVES IN CATALAN TREES

Singularities of the BGF $F(z,u)$ for $|u| = 1$



The Exp-Log Scheme

EXAMPLE.

Goncharov 1994 on cycles in permutations.

$$Perm = \text{Set}Cycles$$

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

By singularity analysis,

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

pointwise for each u but also $\boxed{\text{uniformly}}$

$$[z^n]F(z,u) = \frac{1}{\Gamma(u)}\left(e^{u-1}\right)^{\log n} + \cdots$$

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

Limit Gaussian Law

Exp-Log Scheme Theorem [F-Soria'93]

$$F(z,u) = A(z,u) + B(z,u)C(z)^{-\alpha(u)}$$

$F(z,1)$ *has a single dominant exp-log singularity*.

$C(\rho) = 0$.

$A(z,u), B(z,u), C(z)$ *are analytic in* $u \in \mathcal{V}(1)$, $|z| < \rho + \epsilon$.

$\Longrightarrow$ A Gaussian Law.

Proof. Singularity Analysis +QPT.

Applications

1) Cycles in permutations.

2) ***Sets*** of *logarithmic* generators, e.g., derangements

$$\exp\left(-u\left(\sum_{d\in D}\frac{z^d}{d}+\log\frac{1}{1-z}\right)\right).$$

3) Mappings.

$$\mathcal{F}=\mathrm{Set}(\mathcal{K}),\quad \mathcal{K}=\mathrm{Cycle}(\mathcal{T}),\quad \mathcal{T}=\mathcal{Z}\star\mathrm{Set}(\mathcal{T}).$$

$$F(z)=e^{K(z)},\quad K(z)=\log\frac{1}{1-T(z)},\quad T(z)=ze^{T(z)}.$$

But the Tree function $T(z)$ has singularity at e^{-1}:

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

Gaussian Law with $\mu_n\sim\frac{1}{2}\log n$, $\sigma_n^2\sim\frac{1}{2}\log n$

4) Functional Graphs (Mappings Patterns)

$$F(z)=\exp\left(K(z)+\frac{1}{2}K(z^2)+\frac{1}{3}K(z^3)+\cdots\right),\quad etc.$$

EXAMPLE 5) Polynomials over finite fields

— are a sequence of coefficients

$\Longrightarrow$ GF has a pole

Coeffs grow like q^n

— are a set of irreducibles (primes)

$\Longrightarrow$ GF has a LOG sing.

Coeffs grow like $\dfrac{q^n}{n}$

A Prime Number Theorem: *The density of irreducibles is* $\dfrac{1}{n}$.

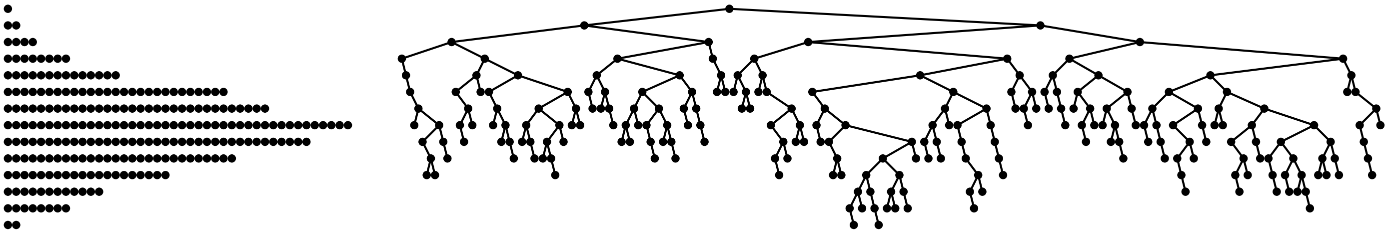
— Bivariate Analytic Schema

$\exp(u \log)$

$\Longrightarrow$ Gaussian law

An Erdős-Kac Theorem: *The number of irreducibles is asymptotically Gaussian.*

Example 6) Internal nodes in BST's [Louchard, etc]



Tree-Node = Root or Left-Node or Right-Node

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

$$F(z,u) = (1-z)^{-2u}$$

♡ More generally order statistics and differential equations (later!)

3 Implicit Schemes & Functional Equations

THREE TYPES OF FUNCTIONAL EQUATIONS
imply Gaussian Laws via singularities

- I. Rational Systems
- II. Algebraic/Implicit Systems
- III. Differential Systems

EXAMPLE. Coin Fountains (Odlyzko-Wilf)

$$F(z,u) = \cfrac{1}{1 - \cfrac{zu}{1 - \cfrac{z^2}{1 - \cfrac{z^3}{1 - \cfrac{z^4}{\ddots}}}}}$$

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

$G(z)$ is analytic in $|z| < \rho$, with $\frac{1}{4} < \rho < 1$ and with $G(\rho) = \infty$

(Also connected to Hardy-Ramanujan identities, easy)

$\Longrightarrow$ Gaussian Law

- $\boxed{\text{I. Positive Rational Systems}}$

$$Y = V + T \cdot Y$$

with $V = V(z,u)$ and $T = T(z,u)$ nonnegative.

Technical conditions: irreducible, aperiodic, unicity of dominant pole of $Y(z,1)$.

$$\mu_n \sim c_1 n, \quad \sigma_n^2 \sim c_2 n, \qquad \text{Gaussian Law}$$

1) Limit theorems for Markov chains

2) Mixture with some renewal theory

3) Patterns in strings

4) Transfer Matrices.

Bender 1973

- $\boxed{\text{II. Positive Polynomial Systems}}$

[F, Soria'93] $\mapsto$ Drmota, RS&A, 1997

1) Simple families of trees (Meir-Moon): node types

   Branching Processes

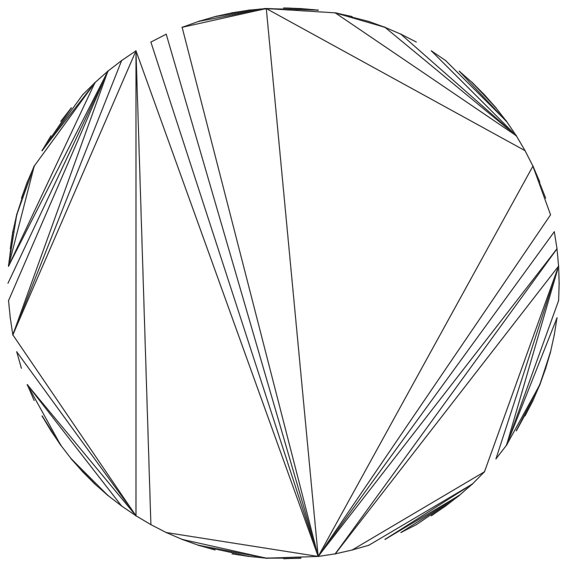
2) Same with Pólya operators (Gittenberger)

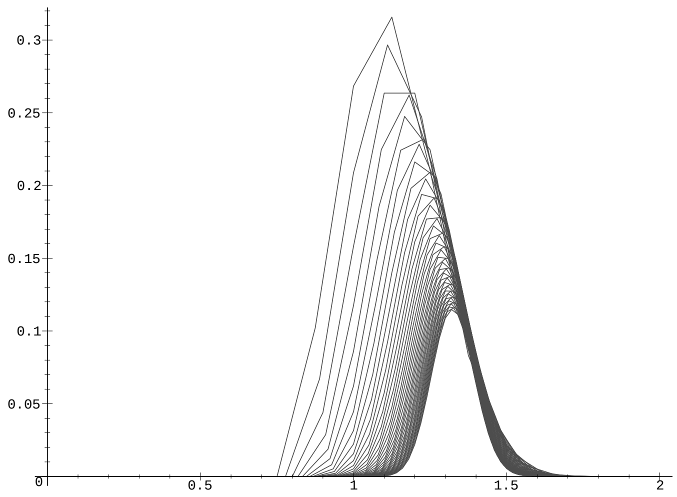
3) Noncrossing graphs [F, Noy]

4) Context-free objects

5) Leaves etc in random mappings

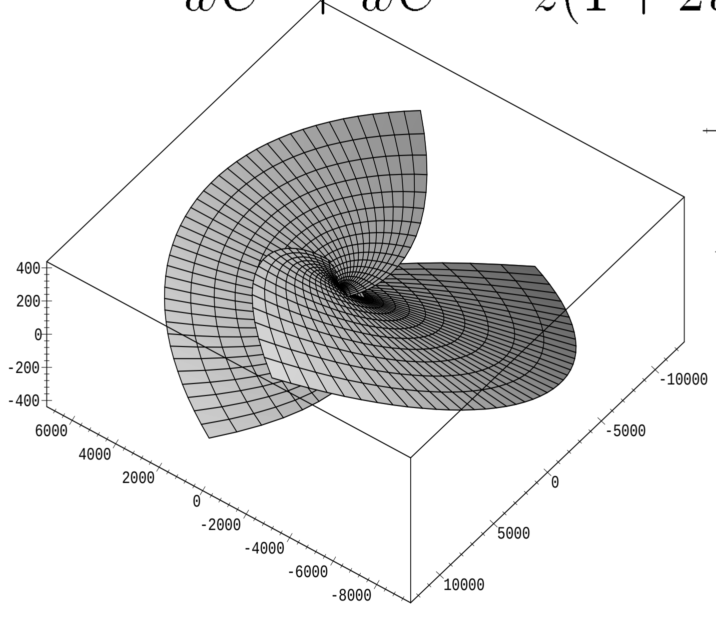
6) Perturbations of the implicit fn thm.

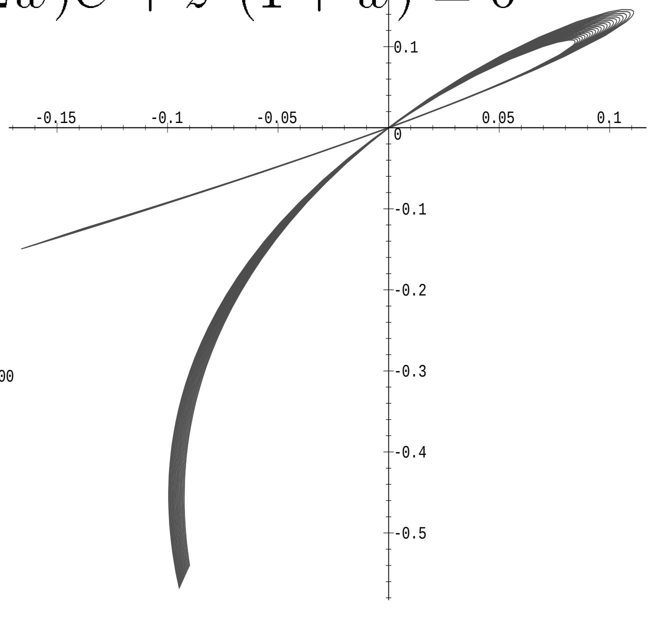




$$C^3 + C^2 - 3zC + 2z^2 = 0$$

$$wC^3 + wC^2 - z(1+2w)C + z^2(1+w) = 0$$





- III. Differential Systems

Singularity perturbation

1) Linear ODE [F, Lafforgue, 1994]

2) Riccati [F, Gourdon, Martinez]

3) Others, some [Bergeron-F-Salvy]

Linear ODE's Level in BST's

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

Level in Quadtrees

$$(d=2) \qquad Y = 1 + 4u\mathbf{I} \cdot \mathbf{J}Y$$

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

PRINCIPLE. ***Univariate.*** Know singularities of linear ODE's

$$(1-z)^2(1-z/2)y''(z) + (1-z)\cosh(z)y'(z) - \frac{1}{\cos z}y(z) = 0$$

$1,\quad 2,\quad \pi/2$

+Substitute $(1-z/\rho)^{-\alpha}$ and get an indicial polynomial that determines possible singular exponents:

$$\boxed{\chi(\alpha) = 0}$$

Singularity Perturbation, e.g,

- ρ fixed, $\alpha(u)$ varies: EXP-LOG scheme
- $\rho(u)$ varies, α fixed: ALG-LOG scheme
- $\rho(u), \alpha(u)$ vary: Gao-Richmond

APPLICATIONS. 1) Quadtrees; 2) Quickselect [Prodinger] = pebbling of BST's

Non-linear ODE's Riccati

$$Y' = aY^2 + bY + c$$

$$Y = \frac{1}{a}\frac{W'}{W} \qquad \Leftrightarrow \qquad W'' = AW' + BW.$$

Use singularity perturbation on linear case: usually poles and a movable singularity.

Patterns in Binary Search Trees [FGM, RSA'97]

$$Y(z) = E(z,u) + \int_0^z Y^2(t)\,dt$$

$\Longrightarrow$ Gaussian Limits for: leaves, paging, patterns and sets of patterns. *Cf* Devroye.

Others. Increasing trees [Bergeron-F-Salvy'92]

Pyramids, Recursive trees, etc...

4 Conclusions

§. Left aside Saddle Point Schemes:

Set partitions, integer partitions, etc, Exp(FAST).

Digital "tries" (Jacquet, Régnier, Szpankowski)

Sachkov: "almost quasipowers". Analytic Depoissonization [J&S]

§ 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}

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

STRONG COMBINATORIAL
DECOMPOSABILITY

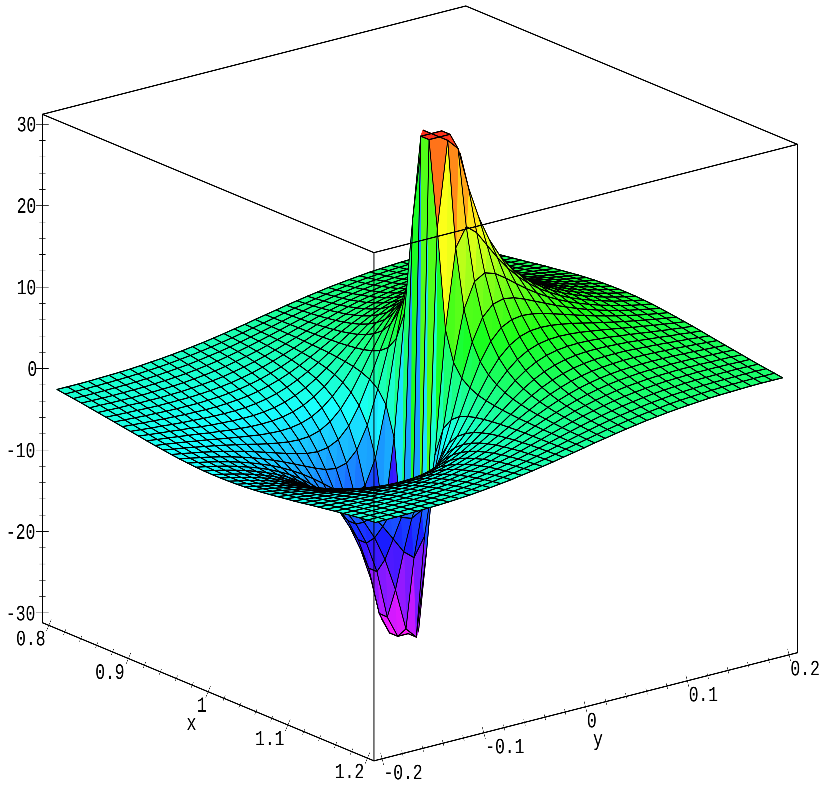
STRONG ANALYTICITY

SMOOTH SINGULARITY
PERTURBATION

GAUSSIAN LAWS
+ Local limits, large deviations, etc

$\boxed{\textbf{\textit{ANALYTIC COMBINATORICS}}}$

Permutations: $(1-z)^{-1}$



Trees: $1-\sqrt{1-z}$