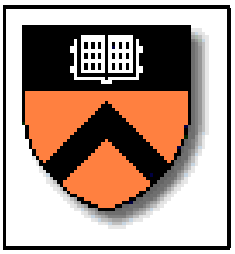
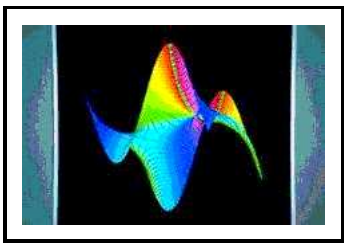




THE AIRY DISTRIBUTION IN ANALYTIC COMBINATORICS

Philippe Flajolet, INRIA Rocquencourt

With P. Poblete (Chili) and A. Viola (Uruguay)

1. Analysis of algorithms and distributions
— pumping method for moments
2. Linear Probing Hashing: average-case
3. Analytic Combinatorics
4. Linear Probing Hashing: generating functions
5. Linear Probing Hashing: limit distribution
6. Combinatorics in the background
— LPH, tree inversions, path length, area; connectivity in RG's
— coalescent saddle points

1 ANALYSIS of ALGORITHMS

goes probabilistic

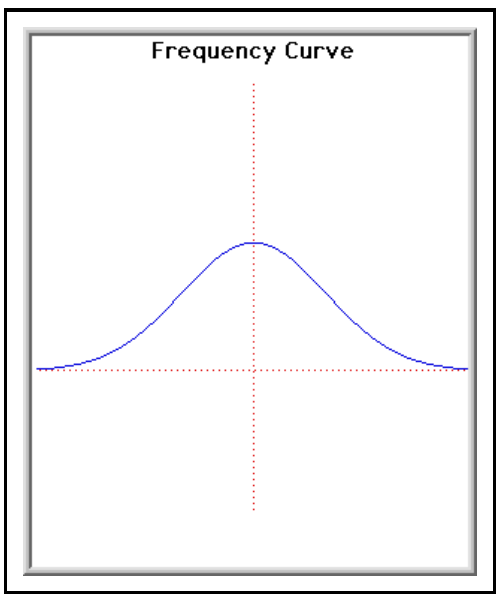
- Probabilistic methods $\mapsto$ probability distributions;
- Average case analysis $\mapsto$ probability distributions.

$X_n :=$ cost on (random) data of size n
with mean μ_n and s-dev σ_n:

$$\Pr\left\{\frac{X_n - \mu_n}{\sigma_n} \leq x\right\} \to F(x).$$

E.g: *bubble sort, insertion sort, 2-1-shellsort, trie sort*
yield a Gaussian law

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



Distributions: Subconfigurations

Number of occurrences of a subconfiguration, esp. for linear time algorithms and parameters

Limit laws and BGF's.

E.g: string matching, runs in permutations, etc.

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

$$[z^n]F(z,u) \approx \left(\frac{1}{\rho(u)}\right)^n.$$

Singularity/Saddlepoint perturbation + Hwang's quasi-powers theory.

Bivariate schemas by: Bender, Richmond, Pittel, Soria, Fl. Drmota, Hwang, . . . , (1974–pres.)

Distributions: Difference equations

For nonlinear parameters, esp. quadratic ones, decomposability implies a *functional equation*.

$$X_n = n + X_{\text{smaller}} + X'_{\text{smaller}}$$

Trie sort (Jacquet, Régnier)

$$F(z,q) = F(q\frac{z}{2},q)^2 + az + b$$

Digital search and compression (J, Szpank, Louchard)

$$\frac{\partial}{\partial z}F(z,q) = F(q\frac{z}{2},q)^2$$

Quicksort (Hennequin + Régnier, Rösler)

$$\frac{\partial}{\partial z}F(z,q) = F(qz,q)^2$$

In situ permutation (Knuth, Prodinger)

$$\frac{\partial}{\partial z}F(z,q) = F(z,q)\cdot F(qz,q)$$

Linear probing hashing (Knuth, Fl-Poblete-Viola)

$$\frac{\partial}{\partial z}F(z,q) = F(z,q)\cdot\frac{F(z,q)-qF(qz,q)}{1-q}$$

Path length in trees (Louchard, Takacs)

$$F(z,q) = \frac{z}{1-F(qz,q)}$$

Distributions: Pumping method

$$F(z,q) = \Phi(F(z,q), F(qz,q), \dots) \qquad \text{nonlinear } \Phi$$

Apply ∂_z repeatedly and set $q = 1$

$$f_k(z) := \left(\frac{\partial}{\partial z}\right)^k F(z,q)$$

the generating function of kth (factorial) moments.

Then, there exists a linear operator $\mathcal{M}$ such that

$$f_k = \mathcal{M}[f_k] + \mathsf{Combi}(f_0, f_1, \dots, f_{k-1})$$

E.g.: Quicksort

$$\mathcal{M}[f](z) = 2\int_0^z f(t)\,\frac{dt}{1-t}$$

This is in particular the average-case iterator.

- Prodinger, Kirschenhofer, Tichy 1985
- Mahmoud and Pittel: m-ary search trees.

Distributions: Pumping method (2)

$$f_k = \mathcal{M}[f_k] + \mathsf{Combi}(f_0, f_1, \dots, f_{k-1})$$

$$f_k = (I - \mathcal{M})^{-1} \mathsf{Combi}(f_0, f_1, \dots, f_{k-1})$$

One can "pump" out moments one after the other; either exactly or asymptotically.

Quicksort normalized costs, as $n \to +\infty$ (Cumulants)

$$\kappa_2 = 7 - 4\zeta(2), \quad \kappa_3 = 16\zeta(3) - 19, \quad \kappa_4 = \frac{937}{9} - 96\zeta(4)$$

Existence of a *limit law*???

Distributions: Moment method

Is a distribution well-characterized by its moments?

The classical moment problem: Knowing

$$\mu_k := E\{X^k\}.$$

Is $w(t)$ determined?

$$\mu_k = \int_I t^k \, dw(t)$$

Moment generating function aka Laplace transform

$$M(s) := \sum_k \mu_k \frac{s^k}{k!} = E\{e^{sX}\} = \int_R e^{st} \, dw(t).$$

Moment theorem. (i) If $M(s)$ exists around 0, then a unique distribution is determined by the moments μ_k.

(ii) If $\mu_{n,k} \equiv E\{{X_n}^k\} \to \mu_k \equiv E\{X^k\}$,
then $X_n \Rightarrow X$ in the sense of convergence in distribution,
$\Pr\{X_n \le x\} \to \Pr\{x \le x\}$.

E.g.: $w(t) = e^{-t}$, $\mu_k = k!$, $M(s) = \frac{1}{1-s}$.

(Moderate growth of moments $\equiv$ sharp tails) $\Longrightarrow$
distribution is determined by its moments.

For a cost parameter:

If

♢ the pumping process can be unrolled, counting on asymptotic simplification;

♢ the limits exist;

♢ these limits determine uniquely a distribution;

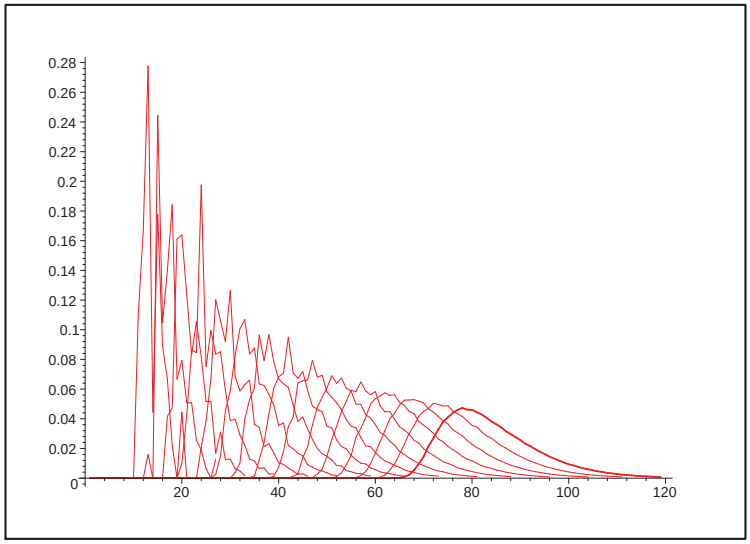
Then,

♡♡ The parameters admit a limit distribition.

Done below for Linear Probing Hashing (L.P.H.)

Quicksort

Theorem. [Hennequin, Régnier, Rösler] *Quicksort has a limit distribution.*



Contrast: "Recursive Method" by Rüschendorf et al.

Path length in Catalan trees = Area below gambler's ruin sequence

Louchard 1984; Takacs ≥ 1990

Theorem. [Louchard, Takacs] *Path length in combinatorial tree models has an Airy limit distribution.*

AIRY distribution defined in terms of Airy functions, i.e., Bessel functions of index $\frac{m}{3}$.

2 Linear Probing: Average-Case

Address calculation methods provide direct access to data.

$$\boxed{\text{SOURCE}} \Longrightarrow \boxed{\text{ADDRESSES}}$$
$$x \mapsto h(x) = \lfloor m \cdot x \rfloor$$

— Place key x at location $h(x)$

— Resolve collisions

- by chaining or other methods.
- by "linear probing"

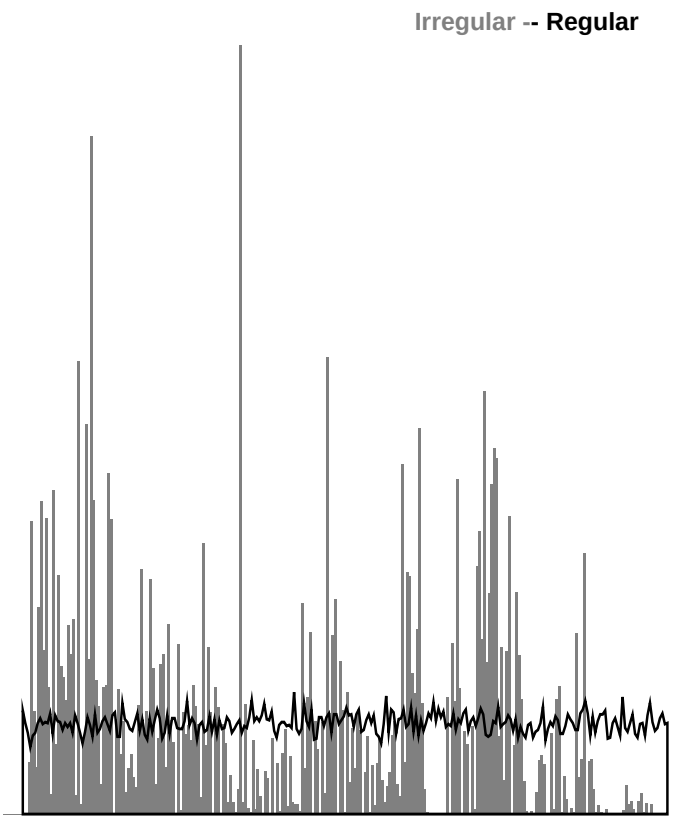
Works well for constrained data or random uniform data.

Natural data are often *not* uniformly random.

Reduction to randomness via

a HASHING FUNCTION $h(x) \ : \ \mathcal{U} \mapsto [0\,..\,m-1]$

(previously $h(x) = \lfloor m \cdot x \rfloor$)

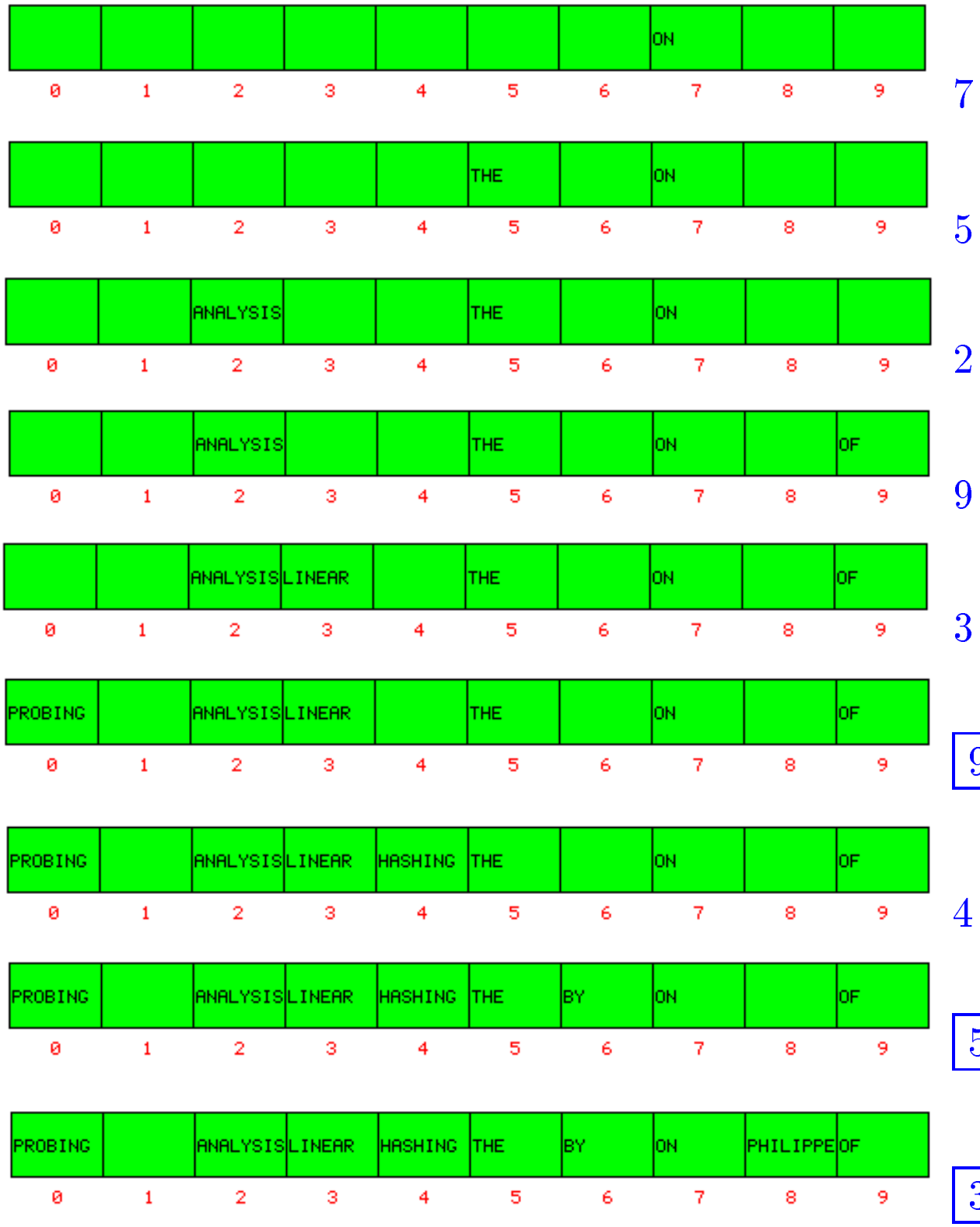


Arithmetic hashing: take $h(x) = \texttt{value}\,(x) \bmod m$

Others: binary, Xor, etc, etc.

Vincent Lum et alii, 1979: works well on real files (DB).

The distribution of natural digrams in a dictionary of the English language with about 25,000 words (in gray), and the smoothing that results from hashing (in black).



http://orange.mcs.dundee.ac.uk:8080/CS202/tuts/mosaic/hashing/HashDemo/index.html

$h(x) =$ ASCII sum mod 10

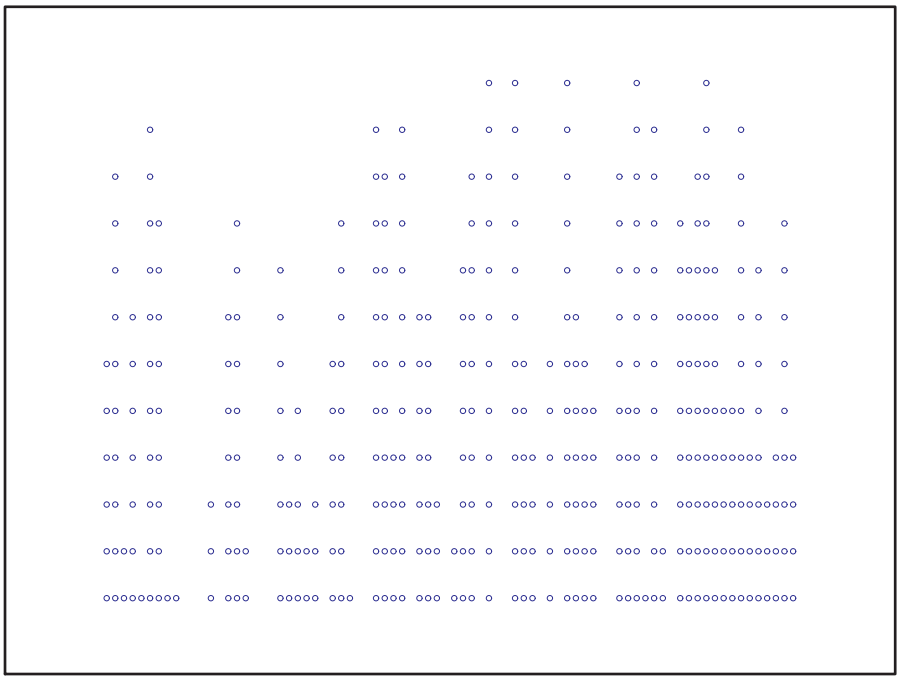
LINEAR PROBING: Physics of "What goes on"?

— For very small α, expect a behaviour like separate chaining (S.C.H.): isolated elements only.

— For moderate α, clusters start to form. These should be a bit larger than S.C.H.

— As $\alpha \to 1$, "clotting" takes place.

— Last element suffers from large displacement $\sim \frac{m}{2}$.



QUANTIFY?

A story that starts with Knuth in 1962

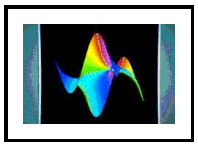
♠ A decomposition based on adding a new element

♠♠♠ Going through Abel identities

$$(x+y)(x+y+n)^{n-1} = xy \sum_{k=0}^{n} \binom{n}{k} (x+k)^{k-1}(y+n-k)^{n-k-1}$$

♠♠♠♠ And concluding with Ramanujan's function

$$Q(n) := 1 + \frac{n-1}{n} + \frac{(n-1)(n-2)}{n^2} + \frac{(n-1)(n-2)(n-3)}{n^3} + \cdots$$



[My first analysis of an algorithm, originally done during summer 1962 in Madison.]

8354

NOTES ON "OPEN" ADDRESSING. D. Knuth 7/22/63

1. Introduction and Definitions. Open addressing is a widely-used technique for keeping "symbol tables." The method was first used in 1954 by Samuel, Amdahl, and Boehme in an assembly program for the IBM 701. An extensive discussion of the method was given by Peterson in 1957 [1], and frequent references have been made to it ever since (e.g. Schay and Spruth [2], Iverson [3]). However, the timing characteristics have apparently never been exactly established, and indeed the author has heard reports of several reputable mathematicians who failed to find the solution after some trial. Therefore it is the purpose of this note to indicate one way by which the solution can be obtained.

We will use the following abstract model to describe the method: N is a positive integer, and we have an array of N variables $x_1, x_2, \ldots, x_N$. At the beginning, $x_i = 0$, for $1 \leq i \leq N$.

To "enter the k-th item in the table," we mean that an integer a_k is calculated, $1 \leq a_k \leq N$, depending only on the item, and the following process is carried out:

1. Set $j = a_k$.
2. "The comparison step." If $x_j = 0$, set $x_j = 1$ and stop; we say "the k-th item has fallen into position x_j."
3. If $j = N$, go to step 5.
4. Increase j by 1 and return to step 2.
5. "The overflow step." If this step is entered twice, the table is full, i.e. $x_i = 1$ for $1 \leq i \leq N$. Otherwise set j to 1 and return to step 2.

Observe the cyclic character of this algorithm.

We are concerned with the statistics of this method, with respect to the number of times the comparison step must be executed. More precisely, we consider all of the N^k possible sequences $a_1, a_2, \ldots a_k$ to be equally probable, and we ask, "What is the probability that the comparison step is used precisely m times when the k-th item is placed?"

2. Non-overflow (self-contained) sequences.

Let $\left[\begin{smallmatrix} n \\ k \end{smallmatrix}\right]$ denote the number of sequences $a_1, a_2, \ldots a_k$ $(1 \leq a_i \leq n)$ in which no overflow step occurs during the entire process of placing k items, if the algorithm is used for $N = n$. (By convention, we set $\left[\begin{smallmatrix} n \\ 0 \end{smallmatrix}\right] = 1$.)

Lemma 1: If $0 \leq k \leq n+1$, then $\left[\begin{smallmatrix} n \\ k \end{smallmatrix}\right] = (n+1)^k - k(n+1)^{k-1}$.

Proof: This proof is based on the fact that $\left[\begin{smallmatrix} n \\ k \end{smallmatrix}\right]$ is precisely the number of sequences $b_1, b_2, \ldots b_k$ $(1 \leq b_i \leq n+1)$ in which, if the algorithm is carried out for $N = n+1$, then $x_{n+1} = 0$ at the end of the operation. This follows because every sequence of the former type is one of the latter, and conversely the condition implies in particular that $1 \leq b_i \leq n$, and that no overflow step occurs.

But sequences of the latter type are easily enumerated, because the algorithm has circular symmetry; of the $(n+1)^k$ possible sequences $b_1, b_2, \ldots b_k$, exactly $k/(n+1)$ of these leave $x_{n+1} \neq 0$. This shows that

$$\left[\begin{smallmatrix} n \\ k \end{smallmatrix}\right] = (n+1)^k \left(1 - \frac{k}{n+1}\right).$$

Theorem K. • *α-sparse tables*, $\alpha = n/m$

L.P.H. $C_n^+ \sim \frac{1}{2}\left(1 + \frac{1}{(1-\alpha)}\right)$ $C_n^- \sim \frac{1}{2}\left(1 + \frac{1}{(1-\alpha)^2}\right)$

Compare

S.C. $C^+ = 1 + \frac{\alpha}{2}$ $C^- = 1 + \alpha$

• *Full tables*, $n = m - 1$

L.P.H. $C_n^+ \sim \sqrt{\frac{\pi n}{2}}$ $C_n^- \sim \frac{n}{2}$.

♡ Fine for α small

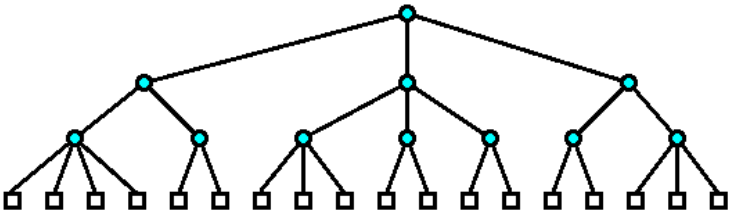
♠ In a full table mean displacement is $O(\sqrt{n})$, distance to nearest gap is $O(n)$.

3 Analytic Combinatorics

Two basic principles $\mapsto$ "dictionaries"

SYMBOLIC METHODS

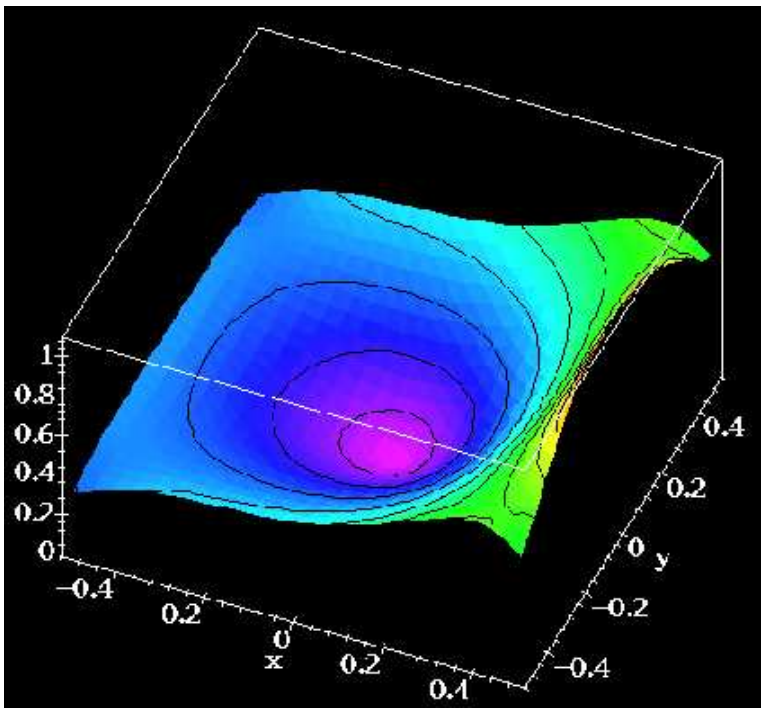
Generating functions



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

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

Analytic functions and singularities



CONSTRUCTIONS

Dictionary (I)

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

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

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

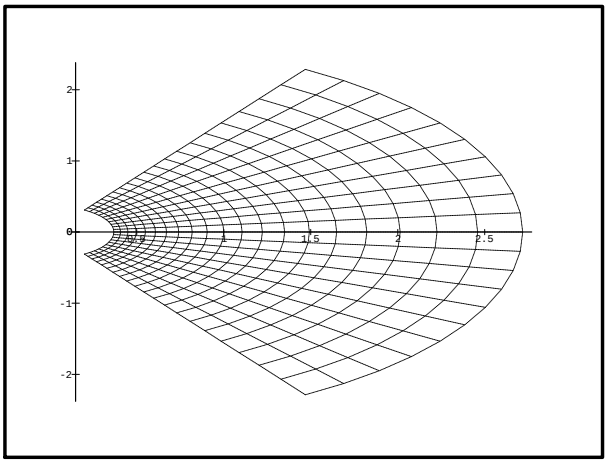
$$\begin{aligned}
\mathbf{A} \cup \mathbf{B} &\mapsto A(z) + B(z) \\
\mathbf{A} \times \mathbf{B} &\mapsto A(z) \times B(z) \\
\text{Seq } \mathbf{A} &\mapsto \frac{1}{1 - A(z)} \\
\text{Set } \mathbf{A} &\mapsto \exp(A(z)) \\
\text{Cycle } \mathbf{A} &\mapsto \log \frac{1}{1 - A(z)}
\end{aligned}$$

COMPLEX ASYMPTOTICS

Dictionary (II)

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

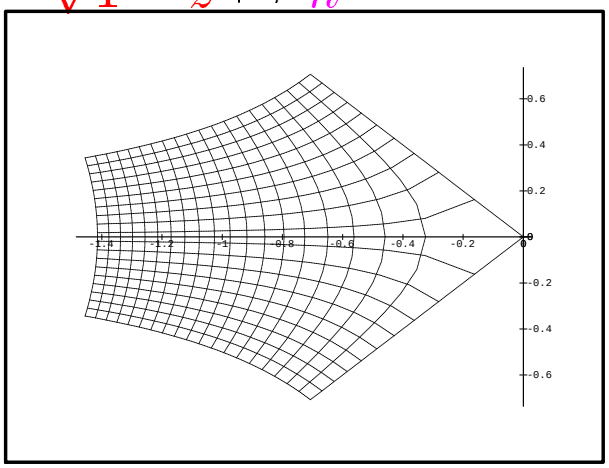
$\exp(z)$



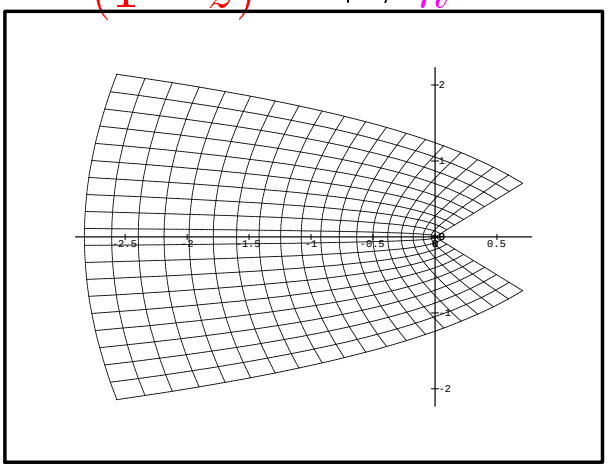
Point of singularity. $\neg\exists \; \frac{d}{dz} f(z)\big|_{z_0}$

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

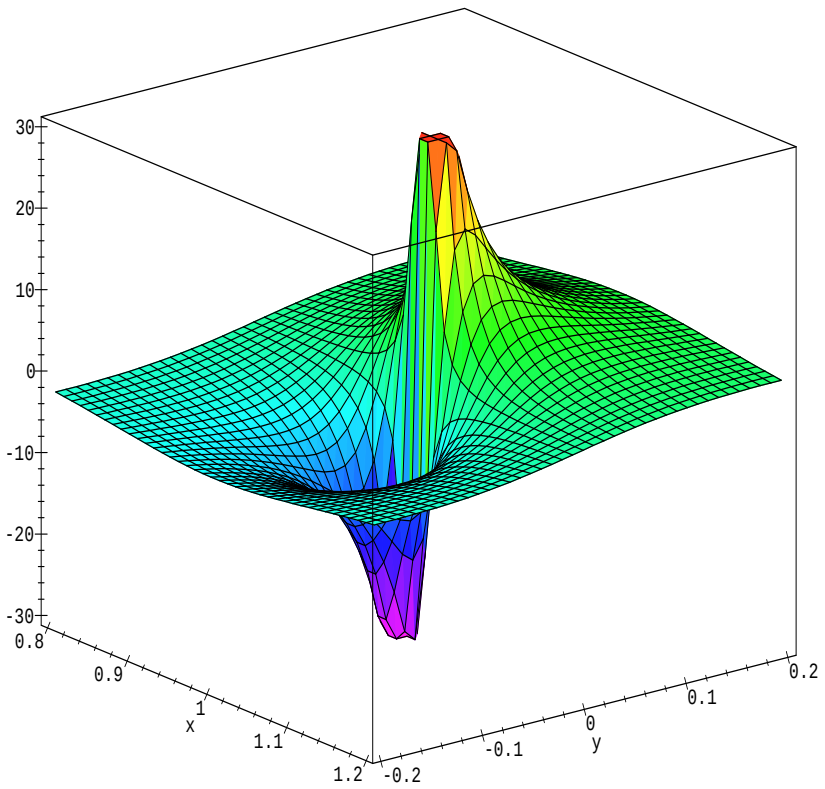
$-\sqrt{1-z} \mapsto n^{-3/2}$



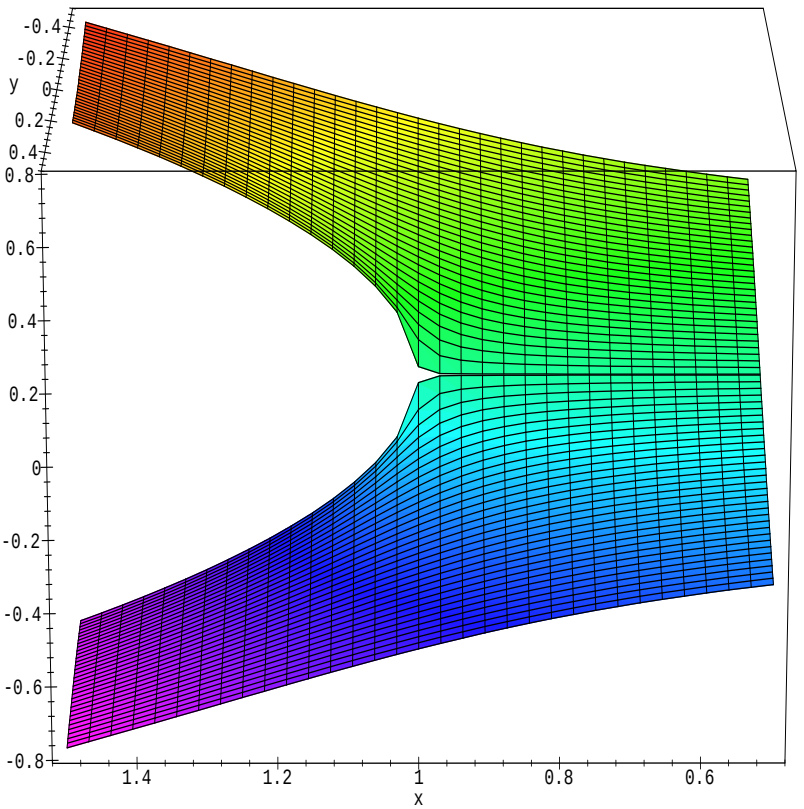
$-(1-z)^{3/2} \mapsto n^{-5/2}$



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



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



4 L.P.H.: Generating functions

Almost-full tables $n = m - 1$ have **tree decomposition**.

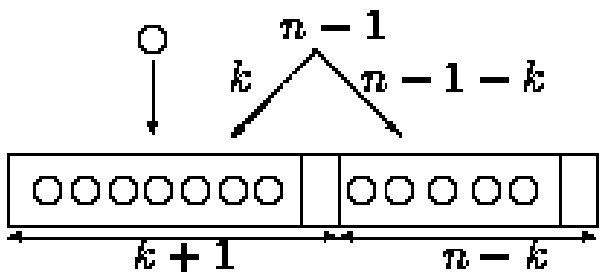


FIGURE 1. The binary tree decomposition of almost full tables.

```
-------------------------------------
|  <Full> := <Full> * <Last> * <Full> |
|             ^^^^^^  with Position   |
-------------------------------------
```

Dictionary: Products $\mapsto$ Products, $C_n = \sum \binom{n}{k} A_k B_{n-k}$

$$\mathcal{C} = \mathcal{A} \star \mathcal{B} \qquad \mapsto \qquad C(z) = A(z) \cdot B(z)$$

Dictionary: Adding an element $\mapsto \int$, $C_n = A_{n-1}$

$$\mathcal{C} = \mathsf{Add}(\mathcal{A}) \qquad \mapsto \qquad C(z) = \int_0^z A(w)\, dw.$$

Dictionary: Choosing a position $\mapsto \partial$, $C_n = (n+1)A_n$

$$\mathcal{C} = \mathsf{Pos}(\mathcal{A}) \qquad \mapsto \qquad \frac{\partial}{\partial z}\left(z\, A(z)\right).$$

A nonlinear ODE translates Linear Probing

$$\begin{aligned} Table &= \underline{Table} \star Table \\ F &= \int (zF)' F \\ T'(1-\tfrac{1}{T}) &= \tfrac{1}{z} \qquad \text{with } T = zF \end{aligned}$$

Lemma 1.

$$F(z) = \frac{1}{z} T(z) \quad \text{where} \quad T = ze^T$$

Lemma 2. [Lagrange + Eisenstein + Cayley]

$$T(z) = z + 2\frac{z^2}{2!} + 9\frac{z^3}{3!} + \cdots = \sum_n n^{n-1} \frac{z^n}{n!}$$

Defines the Tree function $T(z)$

```
Tree = Root * Set(Tree)
```

[Knuth63] The number of almost full tables (n keys)

$$F_n = (n+1)^{(n-1)}$$

Asymptotics

• Find the singularities of an Implicit Function

$$z - Te^{-T} = 0$$

— Implicit function theorem: OK if partial derivative is nonzero

— Singularity when partial derivative equals 0

$$(1 - T)e^{-T} = 0$$

The system gives $\boxed{T = 1,\ z = e^{-1}}$.

Singular dependence between z and T is locally quadratic

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

$$\frac{T_n}{n!} \sim \sqrt{\frac{n}{2\pi}} e^n.$$

Gives back Stirling's formula since $T_n = n^{n-1}$.

$\Longrightarrow$ Analyse any expression involving $T(z)$

Distributional equations (almost full tables)

- *construction cost* = *total displacement*

```
-------------------------------------
|  <Full> := <Full> * <Last> * <Full> |
|             ^^^^^^  with Position   |
-------------------------------------
```

$$F_n(q) = \sum_{k=0}^{n-1} \binom{n-1}{k} F_k(q)(1+q+\cdots+q^k)F_{n-1-k}(q).$$

- q–Calculus: The construction cost is worst-case quadratic

$$\sum_n q^{n^2} z^2$$

Q-calculus

$$
\begin{array}{lcl}
n & \mapsto & [n] = 1 + q + q^2 + \cdots q^{n-1} = \dfrac{1-q^n}{1-q} \\
n! & \mapsto & [n]! = [1] \cdot [2] \cdots [n] \\
\sum (n+1) f_n z^n & \mapsto & \sum [n+1] f_n z^n \\
\dfrac{\partial}{\partial z}(z f(z)) & \mapsto & \mathrm{H}[f(z)] = \dfrac{F(z) - qF(qz)}{1-q}
\end{array}
$$

Many "combinatorial identities" survive in the q-world, with $q^\star \mapsto 1$

EXAMPLE. [Euler] $(\exp(z))^{-1} = (\exp(-z))$

$$\left(\sum \frac{z^n}{n!} \right)^{-1} = \left(\sum \frac{(-z)^n}{n!} \right)$$

$$\left(\sum \frac{z^n}{[n]!} \right)^{-1} = \left(\sum q^{n(n-1)/2} \frac{(-z)^n}{[n]!} \right)$$

$$\partial_z F(z) = \partial_z (z F(z)) \cdot F(z)$$

$$\partial_z F(z,q) = \left(\frac{F(z,q) - qF(qz,q)}{1-q} \right) \cdot F(z,q)$$

$$\boxed{\partial_z F = \mathrm{H}\, F \cdot F}$$

Moments result from

— ∂_q, differentiation w.r.t. q

— U, setting $q = 1$

For the r-th moment, apply $U\, \partial_q^r$ to difference-differential equation

Lemma. Commutation rule with $\mathrm{Z}[f] = z \cdot f$

$$\begin{aligned}
\mathrm{U}\,\mathrm{H} &= \partial_z \mathrm{Z} \\
\mathrm{U}\,\partial_q\,\mathrm{H} &= \mathrm{U}\,\partial_z \mathrm{Z} \partial_q + \frac{1}{2} \mathrm{Z} \partial_z^2 Z, \\
\mathrm{U}\,\partial_q^2\,\mathrm{H} &= \partial_z \mathrm{Z}\,\mathrm{U}\,\partial_q^2 + \mathrm{Z} \partial_z^2 \mathrm{Z}\,\mathrm{U}\,\partial_q + \tfrac{1}{3} \mathrm{Z}^2 \partial_z^3 \mathrm{Z}\,\mathrm{U}
\end{aligned}$$

Proof. $\partial_q [n+1]$, Leibniz, &c.

Lemma. *The generating functions associated with moments* $f_r(z) = \mathrm{U}\,\partial_q^r F(z,q)$ *satisfy a linear ODE that involves* $T(z)$.

$$\begin{aligned} zf_1 &= \frac{1}{2}\frac{T^3}{(1-T)^2}, \\ zf_2 &= \frac{1}{12}\frac{T^4(24-11T+2T^2)}{(1-T)^5} \end{aligned}$$

Theorem. [Full tables, exact form of moments]

$$\begin{aligned} \mathbf{E}[d_{n,n}] &= \frac{n}{2}(Q(n)-1) \\ \mathbf{E}[d_{n,n}^2] &= \frac{n}{12}(5n^2+4n-1-8n\ Q(n)). \end{aligned}$$

$$Q(n) := 1+\frac{n-1}{n}+\frac{(n-1)(n-2)}{n^2}+\frac{(n-1)(n-2)(n-3)}{n^3}+\cdots$$

Theorem. [Full tables, asymptotic form of moments]

$$\begin{aligned} \mathbf{E}[d_{n,n}] &= \frac{\sqrt{2\pi}}{4}n^{3/2}-\frac{2}{3}n+\frac{\sqrt{2\pi}}{48}n^{1/2}-\frac{2}{135}+O(n^{-1}), \\ \mathbf{Var}[d_{n,n}] &= \frac{10-3\pi}{24}n^3+\frac{16-3\pi}{144}n^2+\frac{\sqrt{2\pi}}{135}n^{3/2}-\cdots \end{aligned}$$

Knuth (1962-3); Flajolet-Poblete-Viola (1997); Knuth (1997)

5 L.P.H.: Limit distribution

The method of "pumping" moments

— Start from **nonlinear** combinatorial decomposition (BGF)

$$\Phi[F(z,q)] = 0$$

— Apply derivatives $U\partial_q^r$ to get rth moment.

— Expect **linear** operator $\mathcal{L}$ with

$$\mathcal{L} f_r = \Phi_r[f_0, f_1, \ldots, f_{r-1}]$$

— Solve exactly or asymptotically (singularities)

Path length in Cayley trees: $F(z,q) - ze^{F(qz,q)} = 0$

Pumping moments *ad libidinem*

= for L.P.H..

$$\partial_z F = F \cdot H$$

$$\mathcal{L}\, Y(z) = R_r(z)$$

$$\mathcal{L}\, Y := Y' \cdot (1-T) - Y \cdot \frac{T(2-T)}{z(1-T)}$$

- Homogenous equation

$$\mathcal{L}Y = 0 \quad \Longrightarrow \quad Y = \frac{e^T}{1-T}$$

- Variation-of-constant method

$$Y(z) = \frac{e^{T(z)}}{1-T(z)} \int_0^z R_r(u) e^{-T(u)}\, du$$

Singularity analysis is enough: restrict attention to dominant terms (3 out of $\approx k^2$).

++ [Wright, Janson et al.] Rational function in T.

Moments and reduced singular structure

$$zf_1 \sim \frac{1}{2}\,\frac{1}{(1-T)^2},\quad zf_2 \sim \frac{5}{4}\,\frac{1}{(1-T)^5},\quad zf_3 \sim \frac{45}{4}\,\frac{1}{(1-T)^8}$$

Lemma.

$$\begin{cases} zf_r(z) \sim \dfrac{C_r}{(1-T(z))^{3r-1}} \\ 2C_r = (3r-4)rC_{r-1} + \displaystyle\sum_{j=1}^{r-1} \binom{r}{j} C_j C_{r-j} \quad r \geq 1. \end{cases}$$

Tree decomp. $\mapsto$ Functional Eq. $\mapsto$ Quadratic recurrence

By singularity analysis implies

$$\boxed{E\{(d_{n,n-1})^r\} \sim \frac{-\Gamma(\frac{-1}{2})}{\Gamma(\frac{3r-1}{2})} C_r \left(\frac{n}{2}\right)^{3r/2}}$$

$$\Gamma(s) := \int_0^\infty e^{-t} t^{s-1}\, dt$$

Airy functions

Quadratic recurrence $\mapsto$ Riccati ODE $\mapsto$ Linear ODE

$$y' = y^2 + by + c \quad \mapsto \quad y = -\frac{Y'}{Y}$$

The GF of moment coefficients diverges (taken in asymptotic sense) but it is expressed in terms of Airy function.

$$\sum_r C_r \frac{z^r}{r!} \sim -\frac{I_{2/3}(\frac{1}{3z})}{I_{-1/3}(\frac{1}{3z})}$$

$$I_\nu(z) = (\frac{z}{2})^\nu \sum_{k=0}^{\infty} \frac{(\frac{z^2}{4})^k}{k!\,\Gamma(\nu + k + 1)}$$

Airy solution to $Y'' - zY = 0$

$$\mathrm{Ai}(z) = \frac{1}{\pi} \int_0^\infty \cos\left(\frac{1}{3}t^3 + zt\right)\, dt$$

Use Airy–Bessel asymptotics to get all moments.

Moment problem

Moments grow factorially only, so that the moment problem is determinate. Moment convergence then implies convergence in distribution.

Theorem. For almost full tables, convergence to the Airy distribution,

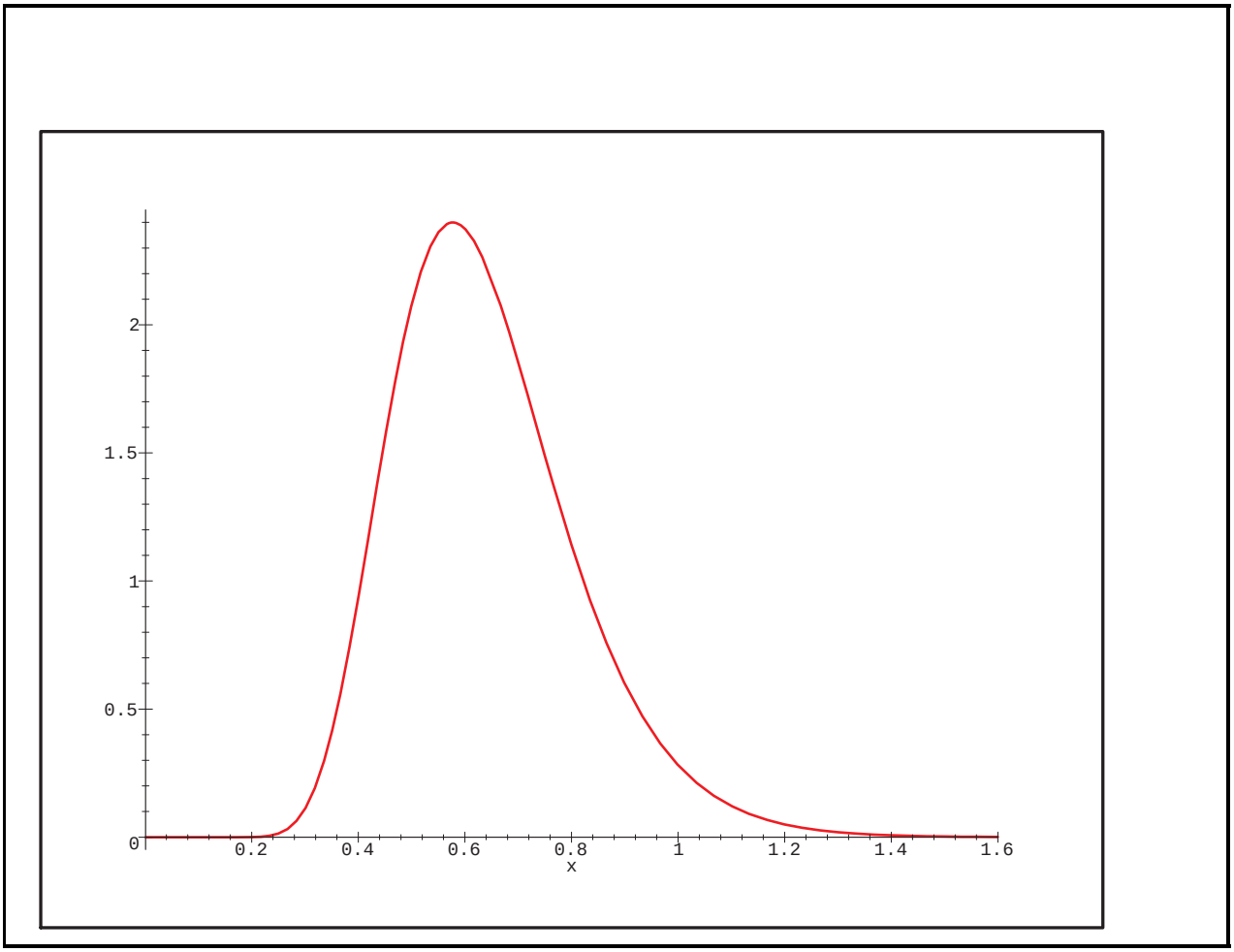
$$\Pr\{\frac{d_{n,n-1}}{(n/2)^{3/2}} \le x\} \to \Pr\{X \le x\} \qquad (n \to \infty),$$

where X is Airy distributed. $\mathbf{E}[X^r] = -\dfrac{\Gamma(-\frac{1}{2})}{\Gamma(\frac{3r-1}{2})}\,\Omega_r$.

$$\sum_{r\ge 0} \Omega_r \frac{w^r}{r!} = -\frac{\Phi_{2/3}(w)}{\Phi_{-1/3}(w)}$$

$$\begin{aligned}\Phi_\nu(w) \quad = \quad & 1 - (4\nu^2-1)\left(\frac{w}{24}\right) + \frac{(4\nu^2-1)(4\nu^2-9)}{2!}\left(\frac{w}{24}\right)^2 \\ & - \frac{(4\nu^2-1)(4\nu^2-9)(4\nu^2-25)}{3!}\left(\frac{w}{24}\right)^3 + \cdots .\end{aligned}$$

The Airy density



Complements

Sparse tables, $\alpha := \frac{n}{m} < 1$

= Gaussian Law (Fl-Poblete-Viola);

= a general schema on "large assemblies" (cf, Mahmoud et al.)

More on analysis of the Airy distribution

= Expressions with Airy zeros (Louchard, Takacs);

= Tail expansions (Louchard, Fl)

6 Combinatorics

The Airy distribution occurs in

— (1) Full tables for L.P.H.
— (2) Inversions in Trees
— (4) Area below random walks
— (5) Path length in random trees

The Airy coefficients C_r, Ω_r occur in

— **(3)** Enumeration of connected graphs

Why?? Foata, Kreweras, Gessel, Knuth, Spencer, Louchard, Takacs, Wright, etc.

Knuth 1997; Mallows–Riordan 1968: $\partial_z F = F \cdot \mathrm{H}\, F$

$$F(z,q) \quad = \quad \frac{\displaystyle\sum_{n=0}^{\infty} q^{n(n+1)/2} \frac{z^n (q-1)^{-n}}{n!}}{\displaystyle\sum_{n=0}^{\infty} q^{n(n-1)/2} \frac{z^n (q-1)^{-n}}{n!}}$$

$$F(z, q+1) = \sum_{n,t} \gamma(n, n+t-1) q^t \frac{z^{n-1}}{(n-1)!},$$

$\Longrightarrow$ "Closed form" but moments are still "hard" to find!

Hashing $\mapsto$ Inversions $\mapsto$ Graphs $\mapsto$ Paths $\mapsto$ Tree P.L.

LPH has Airy distrib.

Hashing $\mapsto$ Inversions $\mapsto$ Graphs $\mapsto$ Paths $\mapsto$ Tree P.L.

Foata ca 1971; Knuth 1997, Kreweras 1980. Almost full table are combinatorially equivalent to trees.

Correspondence: Connect a key/car to (immediately before) where it wanted to land.

Inversions in trees have Airy distrib.

Hashing $\mapsto$ Inversions $\mapsto$ Graphs $\mapsto$ Paths $\mapsto$ Tree P.L.

Correspondence: Depth-first search as a combinatorial correspondence.
Gessel–Wang 1979. Wright 1977 Count connected graphs by excess. Janson et al. 1993 The Giant paper

Graphs by excess counted by Airy coeff.

Hashing $\mapsto$ Inversions $\mapsto$ Graphs $\mapsto$ Paths $\mapsto$ Tree P.L.

Correspondence: Breadth-first search as a combinatorial correspondence.

Spencer 1997. Leads to Poisson walks. Louchard 1984. Analyse area by moments and/or Brownian motion Connects to area of Dyck/Catalan paths by universality of Brownian motion

$$\cfrac{1}{1-\cfrac{qz}{1-\cfrac{q^2z}{1-\cfrac{q^3z}{\ddots}}}} = q\text{-exponentials}$$

Excursion area has an Airy distribution

Hashing $\mapsto$ Inversions $\mapsto$ Graphs $\mapsto$ Paths $\mapsto$ Tree P.L.

Correspondence: Catalan (Dyck) walks as traversal sequences of Catalan trees.

Takacs 1990-1994. Moment methods apply to simple families.

Path length in simple trees has Airy distrib.

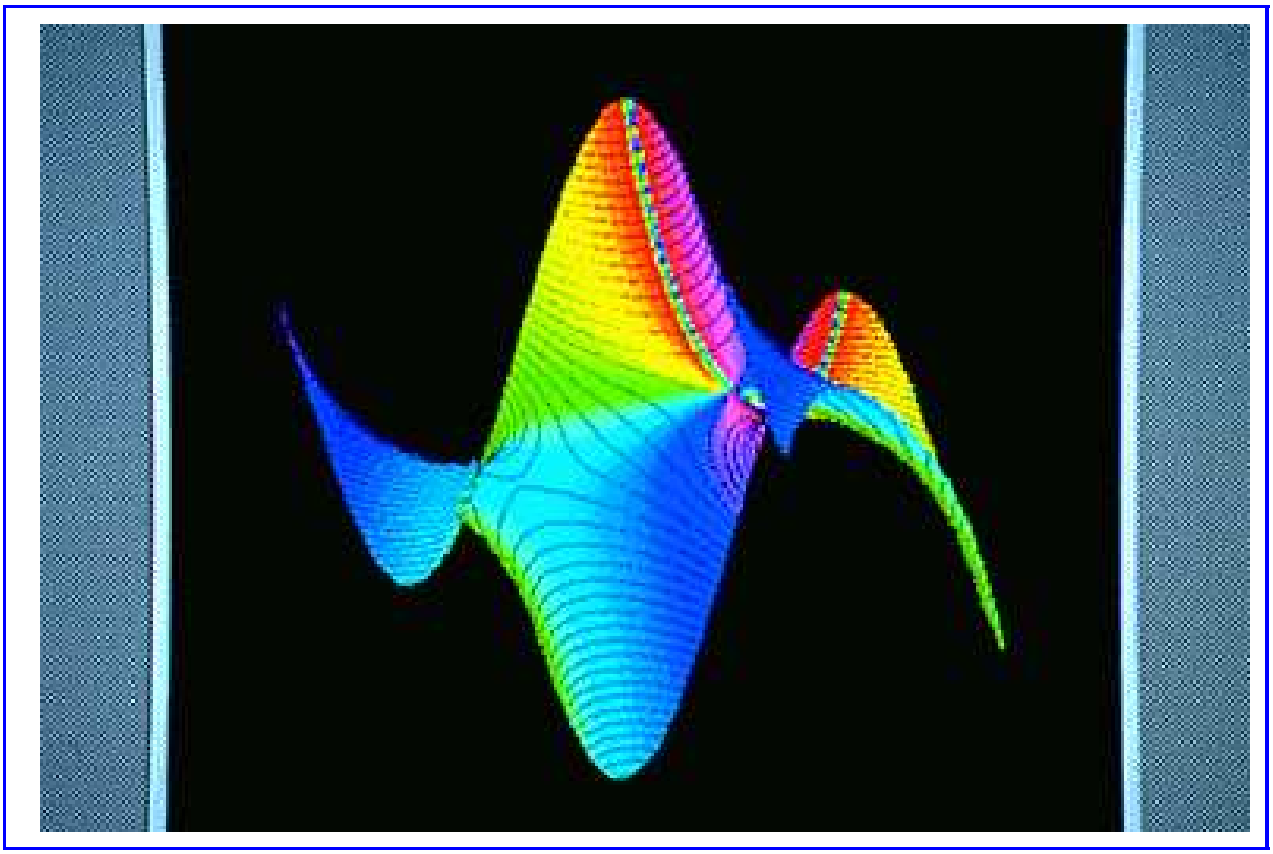
Airy phenomena have a large degree of **universality**

From Flajolet-Salvy (1995): analytic explanation

Lemma. Let $U(z) := \sum u_k z^k$. Then, $q = e^{-t}$,

$$\sum_k q^{k^2/2} u_k = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} U(e^{x\sqrt{t}})\, dx$$

Combinatorics $\mapsto$ Coalescent saddles $\mapsto$ **Airy**



Cf. "Giant paper" of JKLP'93, Prellberg 1993.
Flajolet-Noy 1997 Crossing in chord systems.