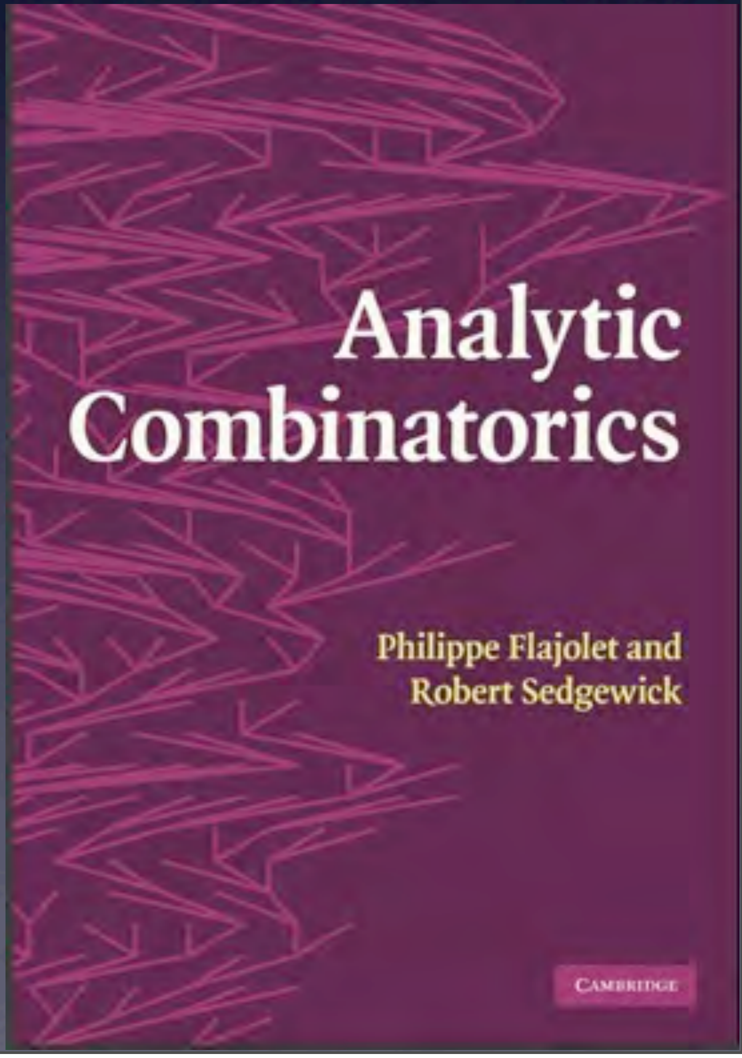
Analyses of Tree Height

Philippe Flajolet, Algorithms, INRIA, France

How tall ("high") are random trees?

- **Combinatorial Tree Models:**
  - General Catalan Trees
  - Binary Trees
  - Simple Varieties & nonplane trees
  - diameter &c



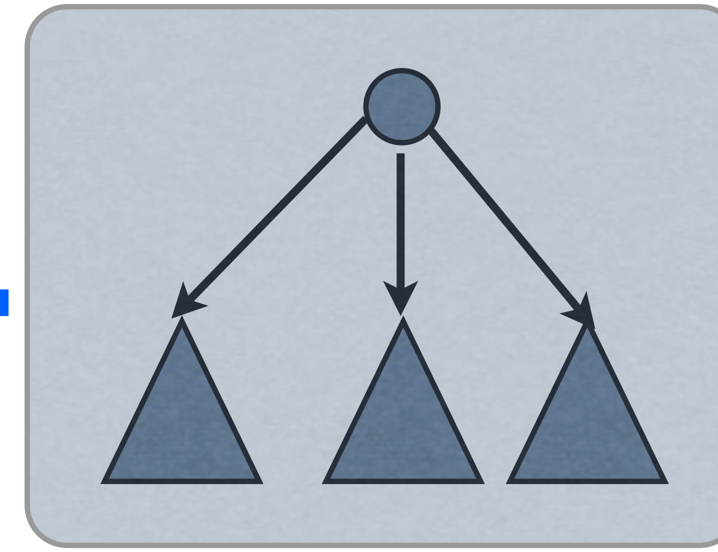
1. General Trees

- "General" trees & Catalan numbers
- De Bruijn, Knuth, & Rice (1972)
- Explicit and limit laws --central/local
- Theta transformations; Continued fractions

Basics

General Catalan trees = plane + all degrees allowed

$$\mathcal{G} = \mathcal{Z} \times \textsc{Seq}(\mathcal{G})$$



- generating function $G(z) := \sum G_n z^n$ is

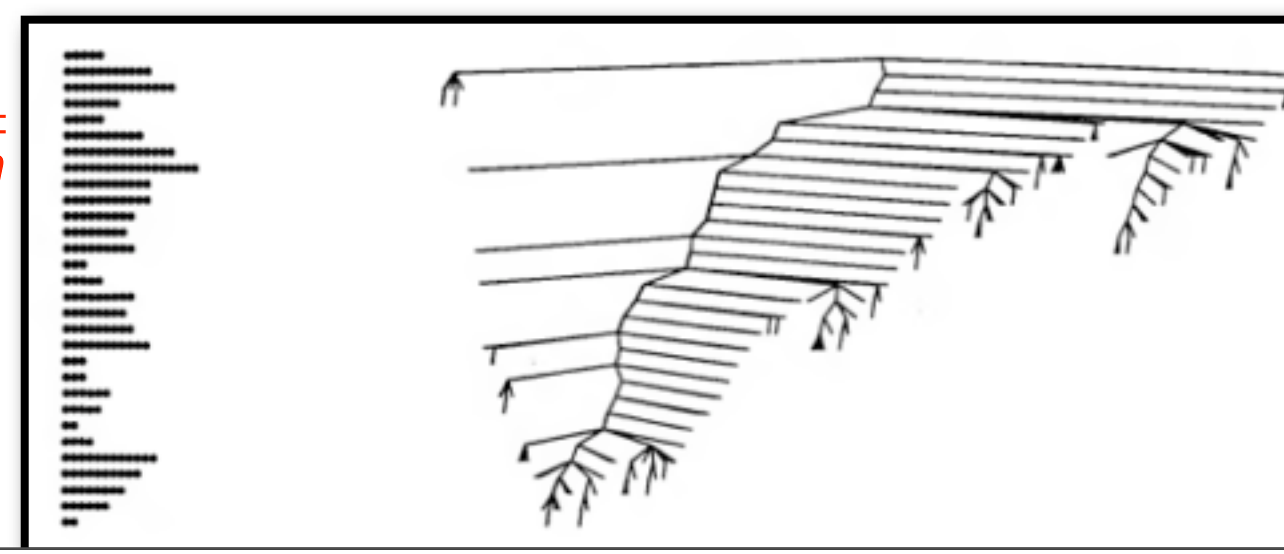
$$G(z) = \frac{z}{1 - G(z)} \quad \Longrightarrow \quad G(z) = \frac{1}{2}\left(1 - \sqrt{1-4z}\right)$$

- coefficients are *Catalan numbers*

$$G_{n+1} = \frac{1}{n+1}\binom{2n}{n}$$

- asymptotically

$$G_{n+1} \sim \frac{4^n}{\sqrt{\pi n}}$$



Tree height: exact forms

De Bruin, Knuth, and Rice 1972 : $\mathcal{G}^{[h]}$:= trees of height $\leq h$

$$\mathcal{G}^{[h+1]} = \mathcal{Z} \times \textsc{Seq}(\mathcal{G}^{[h]}); \qquad \mathcal{G}^{[0]} = \mathcal{Z}$$

$$\begin{cases} G^{[0]}(z) = z \\ G^{[h+1]}(z) = \dfrac{z}{1 - G^{[h]}(z)} \end{cases} \Longrightarrow \text{rational f.} \Longrightarrow G^{[h]} = \cfrac{z}{1 - \cfrac{z}{1 - \cfrac{z}{\ddots \atop 1 - z}}} \ (h \text{ stages})$$

$\Longrightarrow$ **Fibonacci polynomials**: $G^{[h]}(z) = z\dfrac{F_{h+1}}{F_{h+2}}$, $\qquad F_{h+1} = F_h - zF_{h-1}$

$$G^{[h]}(z) = z\frac{F_{h+1}}{F_{h+2}}, \qquad F_{h+1} = F_h - zF_{h-1}$$

— Fibonacci polynomials satisfy a *linear recurrence*;

— characteristic equation is $\rho^2 = \rho - z \implies \rho, \bar{\rho} = \frac{1}{2}\left(1 \mp \sqrt{1-4z}\right)$

— thus $F_h = \dfrac{\rho^h - \bar{\rho}^h}{\rho - \bar{\rho}}$.

Everything is expressible as function of $\rho \equiv G(z)$ alone and **Lagrange Inversion** [ballot numbers] applies:

Theorem (Trees of bounded height, GF and coeff.)

$$G^{[h-1]} = z\frac{\rho^h - \bar{\rho}^h}{\rho^{h+1} - \bar{\rho}^{h+1}}; \qquad G_{n+1} - G_{n+1}^{[h-1]} = \sum_{j\geq 1} \Delta^2 \binom{2n}{n-jh}.$$

= *a sampled sum of line $\underline{2n}$ of Pascal's triangle* *(via Δ^2)*

Pascal

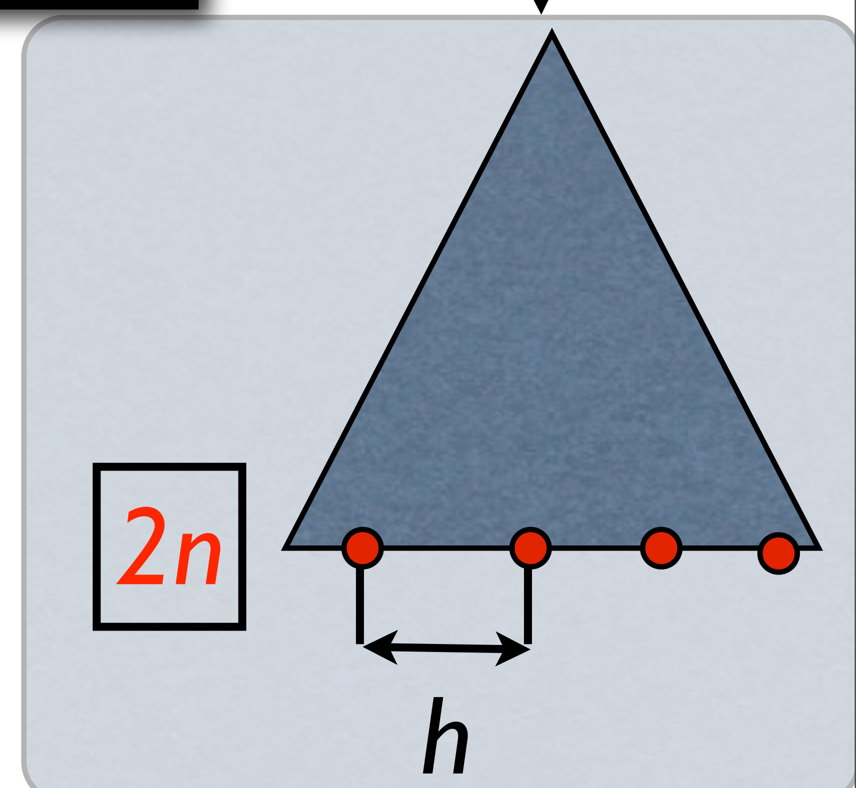
1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1

Theorem (Trees of bounded height, GF and coeff.)

$$G^{[h-1]} = z\frac{\rho^h - \bar{\rho}^h}{\rho^{h+1} - \bar{\rho}^{h+1}}; \quad G_{n+1} - G_{n+1}^{[h-1]} = \sum_{j\geq 1} \Delta^2 \binom{2n}{n-jh}.$$

= a sampled sum of line $\underline{2n}$ of Pascal's triangle (via Δ^2)

$$\Delta^2 f(x) = f(x+1) - 2f(x) + f(x-1).$$



Tree height and trigonometry

"All second-order linear recurrences are the same and are equivalent to multiplication formulae for sin, cos*."*

$$F_h\left(\frac{1}{4\cos^2\theta}\right) = \frac{1}{(2\cos\theta)^{h-1}}\frac{\sin h\theta}{\sin\theta}, \qquad z := \frac{1}{4\cos^2\theta}.$$

Thus Fibonacci $\simeq$ Chebyshev
Thus the roots of $F_h(z) = 0$ are $z = \frac{1}{4\cos^2\theta}$, where $\sin h\theta = 0$.
Thus we know the partial fraction expansion of $G^{[h]}(z)$!

Theorem (Trees of bounded height, trig forms)

$$G^{[h-2]}_{n+1} = \frac{4^n}{h}\sum_{1\leq j<h/2} \sin^2\frac{j\pi}{h}\cos^{2n}\frac{j\pi}{h}.$$

- Lagrange (1775; cf DBKR) had the trig forms (!!!)
- Lord Kelvin (1824–1907; cf Feller) had the sampled binomial sums
- Delannoy (1833–1915; cf Lucas) had the sampled binomial sums*

---

* Henri-Auguste Delannoy et la publication des oeuvres posthumes d'Edouard Lucas. By Autebert, Décaillot, Schwer. In *Gaz. SMF* 1995. Cf Cyril Banderier.

RECHERCHES

SUR

LES SUITES RÉCURRENTES

$$y_{x,t} = 1 - (2\sqrt{pq})^x \left(\sqrt{\frac{q}{p}}\right)^t$$
$$\times \left[(1)\left(\cos\frac{\pi}{n}\right)^x \sin\frac{t\pi}{n} + (2)\left(\cos\frac{2\pi}{n}\right)^x \sin\frac{2t\pi}{n} + (3)\left(\cos\frac{3\pi}{n}\right)^x \sin\frac{3t\pi}{n} + \ldots + (n-1)\left(\cos\frac{(n-1)\pi}{n}\right)^x \sin\frac{(n-1)t\pi}{n}\right];$$

ARTICLE V. — *Application des méthodes précédentes à la solution de différents Problèmes de l'Analyse des hasards.*

PROBLÈME I.

49. *Un joueur parie d'amener un événement donné, b fois au moins, en un nombre a de coups, la probabilité de l'amener à chaque coup étant p; on demande le sort de ce joueur.*

[Thanks: NUMDAM/Gallica]

Limit distributions

Easy by either **binomial forms** or **trig forms**; not in [DBKR].

If $h = x\sqrt{n}$: $\dfrac{\binom{2n}{n-kh}}{\binom{2n}{n}} \sim e^{-k^2x^2}$; $\cos^{2n}\dfrac{j\pi}{h} \sim e^{-j^2\pi^2/x^2}$.

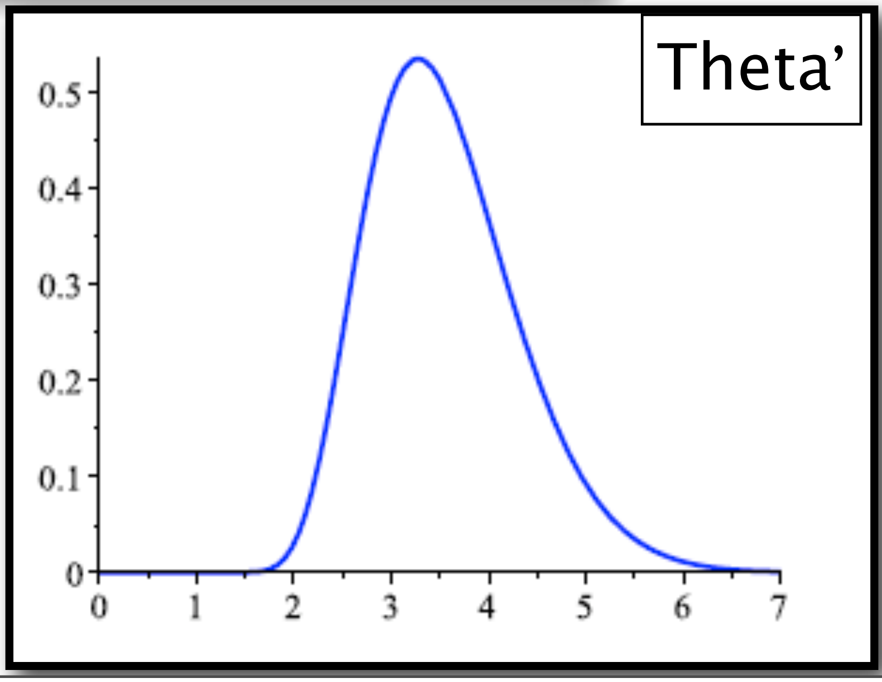
Theorem (Local limit law)

$$\mathbb{P}_{\mathcal{G}_n}\left(H = \lfloor x\sqrt{n}\rfloor\right) \sim \frac{1}{\sqrt{n}}\Theta'(x); \quad \Theta(x) \simeq \begin{cases} \sum e^{-k^2x^2} \dots \\ \sum e^{-k^2\pi^2/x^2} \dots \end{cases}$$

Theorem (Central limit law)

$$\mathbb{P}_{\mathcal{G}_n}\left(H \leq \lfloor x\sqrt{n}\rfloor\right) \to \Theta(x)$$

$$\Theta(x) := \sum_{j\geq 1} e^{-j^2x^2}(4j^2x^2 - 2).$$



Moments

[DBKR] have them:

Theorem (Moments of height)

$$\begin{cases} \mathbb{E}_{\mathcal{G}_n}(H) = \sqrt{\pi n} - \frac{3}{2} + O\left(\frac{1}{\sqrt{n}}\right) \\ \mathbb{E}_{\mathcal{G}_n}(H^r) = r(r-1)\Gamma(r/2)\zeta(r)n^{r/2}. \end{cases}$$

- Need $\sum_h h^r \Theta'(ht)$, with $t = \frac{1}{\sqrt{n}} \to 0$.
- [DBKR] $\simeq$ with **Mellin transforms♡** (can be done with $\sum \mapsto \int$).

$$f(t) \quad \rightsquigarrow \quad f^\star(s) = \int_0^\infty f(t)t^{\,s-1}\,dt.$$

$$\boxed{\mathbb{E}_n(H) \rightsquigarrow \sum d(m)e^{-m^2x^2}}$$

Corollary 1: the theta transformation

By comparing binomial and trig forms of height, get:

$$\frac{1}{\sqrt{\pi x}} \sum_{k=-\infty}^{+\infty} e^{-k^2 x^2} = \sum_{k=-\infty}^{+\infty} e^{-k^2 \pi^2 / x^2}$$

$=$ a well-known elliptic **theta function identity**.

Reverse-engineering from the height of Catalan trees:

- *Forget height, Fibonacci, &c*. Start from $f(z) = (1+z)^{2n}$.
- *Multisection* of series $f(z)$: $\sum_h f_{nh} = \frac{1}{h} \sum_{\omega^h=1} f(\omega)$.
- Analyse asymptotically when $h = x\sqrt{n}$ the two equivalent forms.

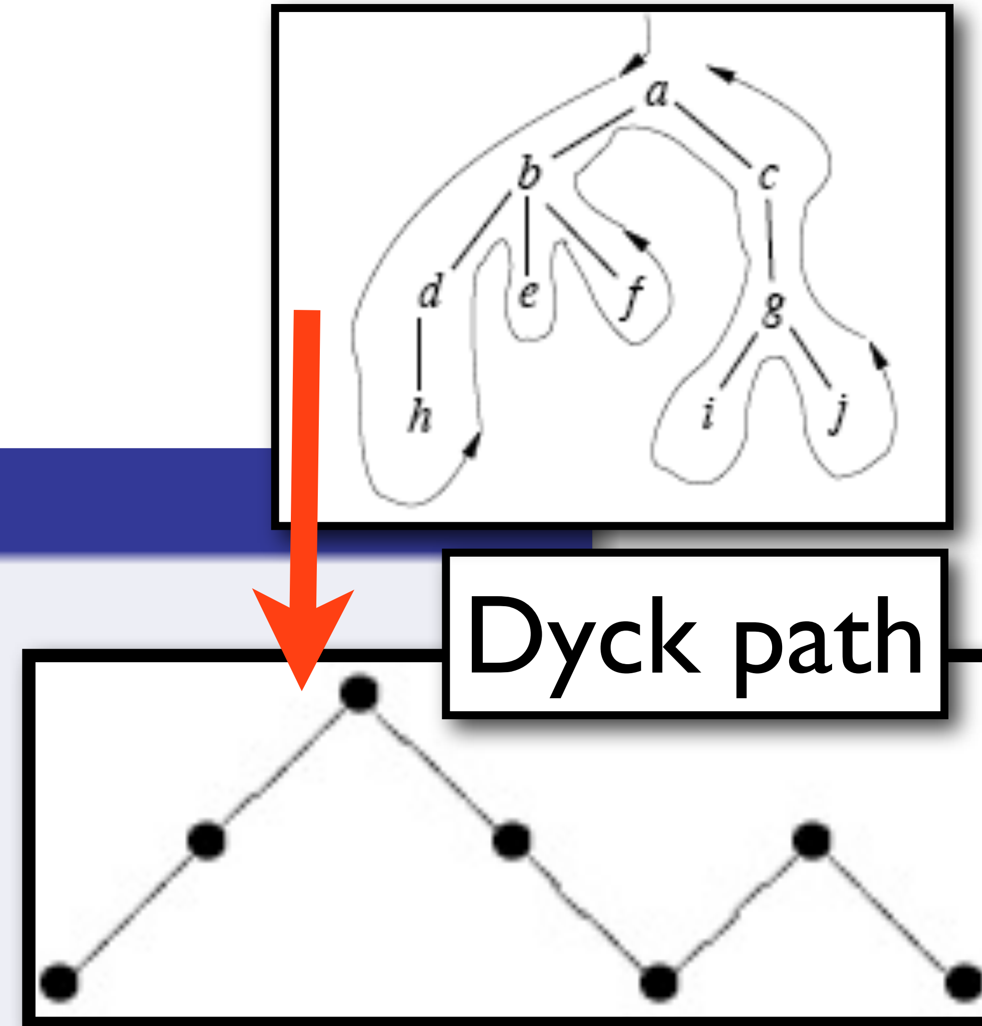
Pólya (1927) "Elementarer Beweis einer Thetaformel". Sitzungsberichten der Preuß. Akad. des Wissenschaften, pp. 157–161.

cf also [Biane-Pitman-Yor, 2001]

Corollary 2: the continued fraction theorem

By inspection of the GFs of height, get the GF of trees, with u_j marking nodes at level j as

$$\cfrac{zu_0}{1-\cfrac{zu_1}{1-\cfrac{zu_2}{\ddots}}}$$



Theorem (Dyck paths and levels of steps)

With u_j marking descents from level j, the GF is

$$\cfrac{1}{1-\cfrac{zu_1}{1-\cfrac{zu_2}{1-\cfrac{zu_3}{\ddots}}}}$$

F. "Combinatorial Aspects of Continued fractions", *Discr. Math.*, 1980 & 2006.

[Good-Touchard-Lenard-Jackson-Flajolet-Read]

2. Binary trees

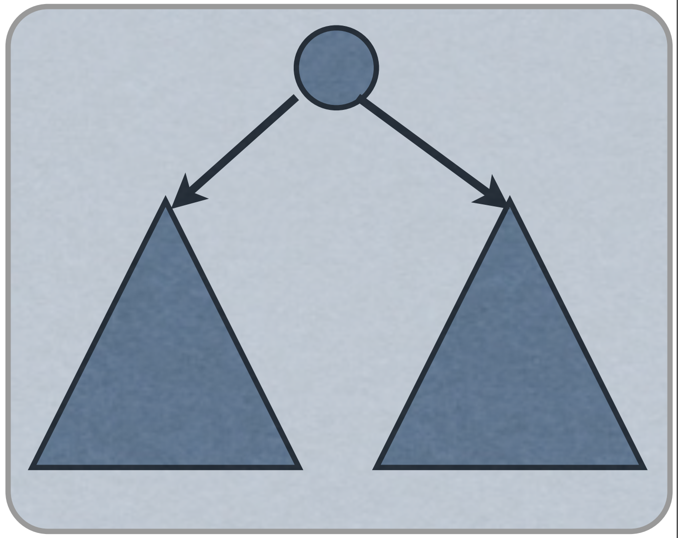
- Iteration of GFs at a fixed point
- Singularity analysis
- Local and central limits

Basics

Binary Catalan trees = plane + degrees $\{0,2\}$ allowed

Size = # leaves

$$\mathcal{B} = \mathcal{Z} + \mathcal{B} \times \mathcal{B}$$



- generating function $B(z) := \sum B_n z^n$ is

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

- $\mathcal{B}^{[h]}$ trees of height $\leq h$ with GF $B^{[h]}(z)$:

$$B^{[0]} = z; \qquad B^{[h+1]} = z + B^{[h]}(z)^2.$$

We have **polynomials determined by a quadratic recurrence.**

Degree **double** at each iteration: $\deg(B^{[h]}) = 2^h$.

The two Catalan's

	General	**Binary**
GF	algebraic	algebraic
bounded height	rational	polynomial
	(explicit, lin. degree)	"implicit" (exponential degree)
coeff.	binomial & trigs	??
asymptotics	direct	via **singularities**

On the real line

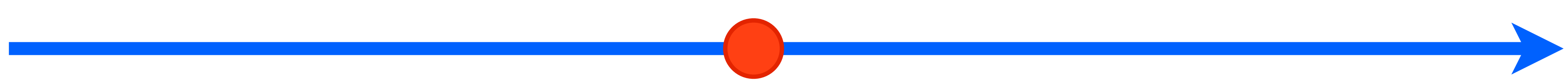
$$B^{[0]} = z; \qquad B^{[h+1]} = z + B^{[h]}(z)^2.$$

- For $z \in (0, \frac{1}{4})$, we have $B(z) - B^{[h]}(z)$ dominated by $\sum_{n>h+1} B_n z^n$. Implies **geometric convergence**.

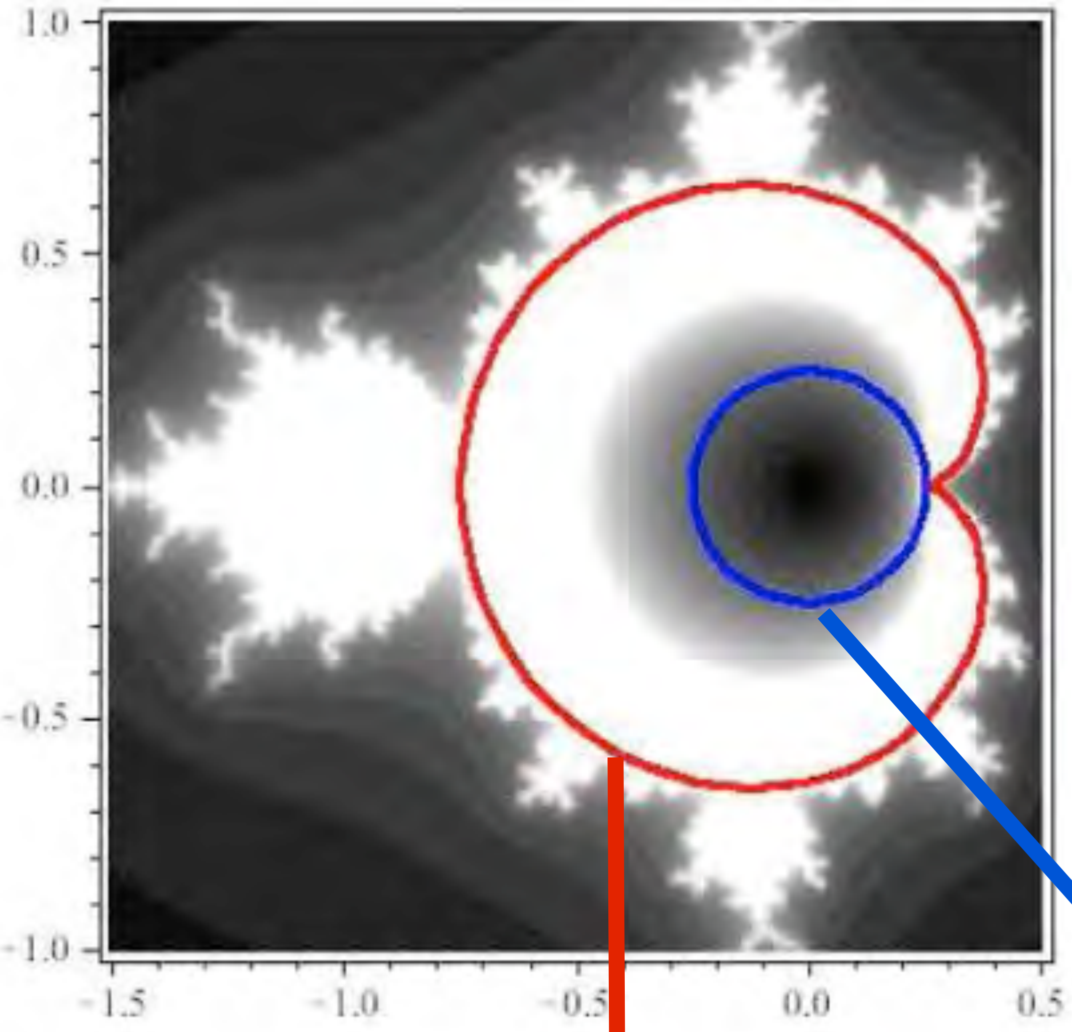
- For $z > \frac{1}{4}$, the $B^{[h]}(z) \nearrow$ and cannot have limit. Thus, unbounded. Thus **blow up doubly exponentially**.

- For $z = \frac{1}{4}$, what goes on??? This is the information at the **singularity** of $B(z)$, hence needed!

$0 \leq z < \frac{1}{4}$	$x = \frac{1}{4}$	$x > \frac{1}{4}$
geometric convergence	??	double exp. divergence



In complex plane



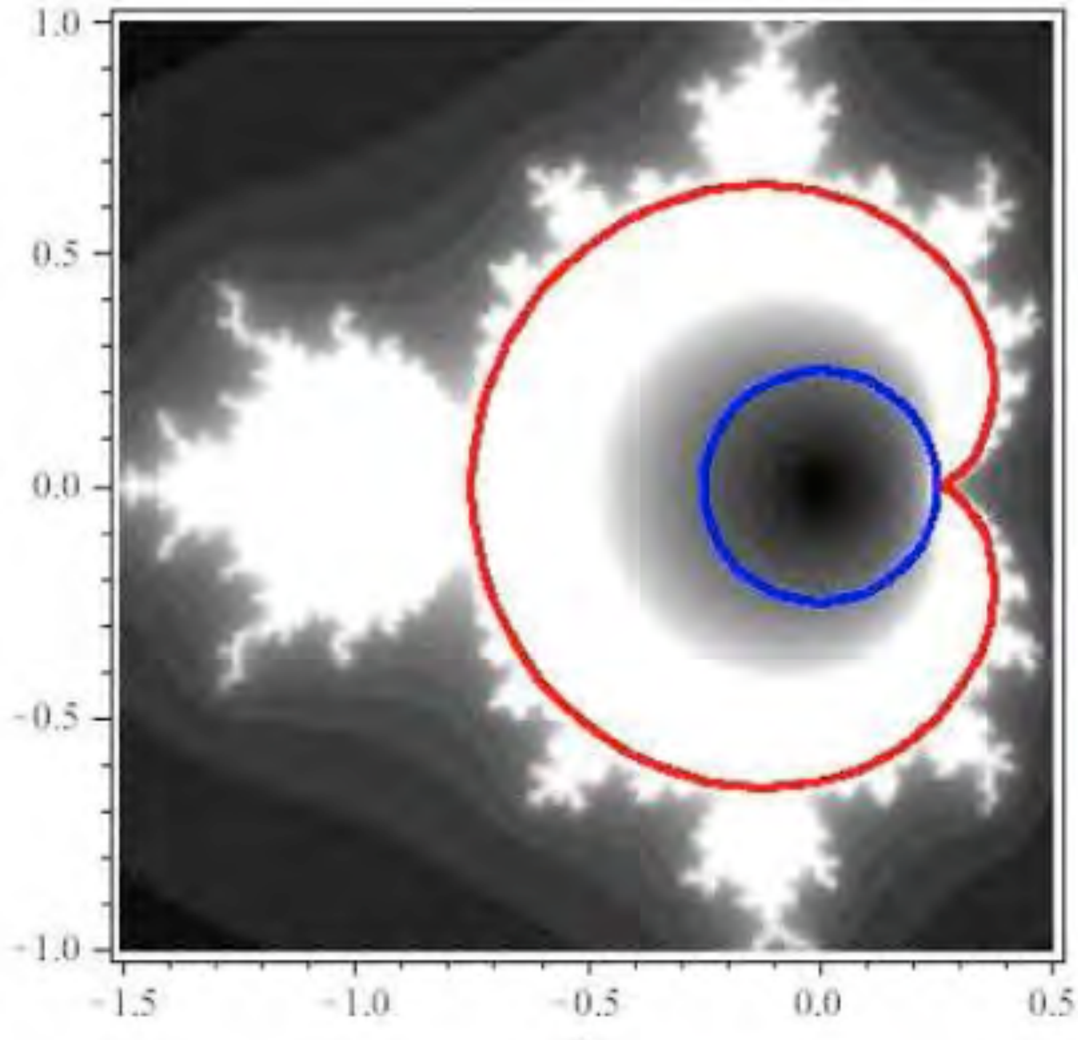
- gray level indicates speed of convergence (to fixed point or to infinity)

$$B^{[h+1]} = z + B^{[h]}(z)^2$$

Disc |z|=1/4

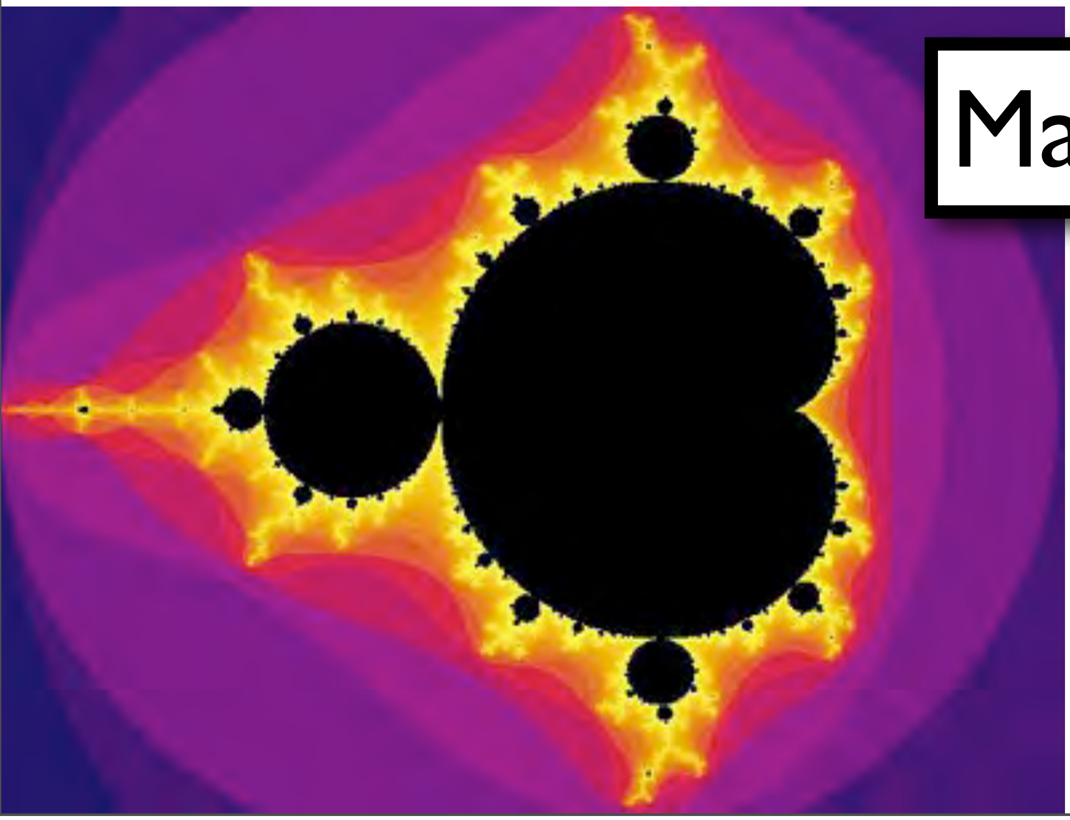
Cardioid

In complex plane

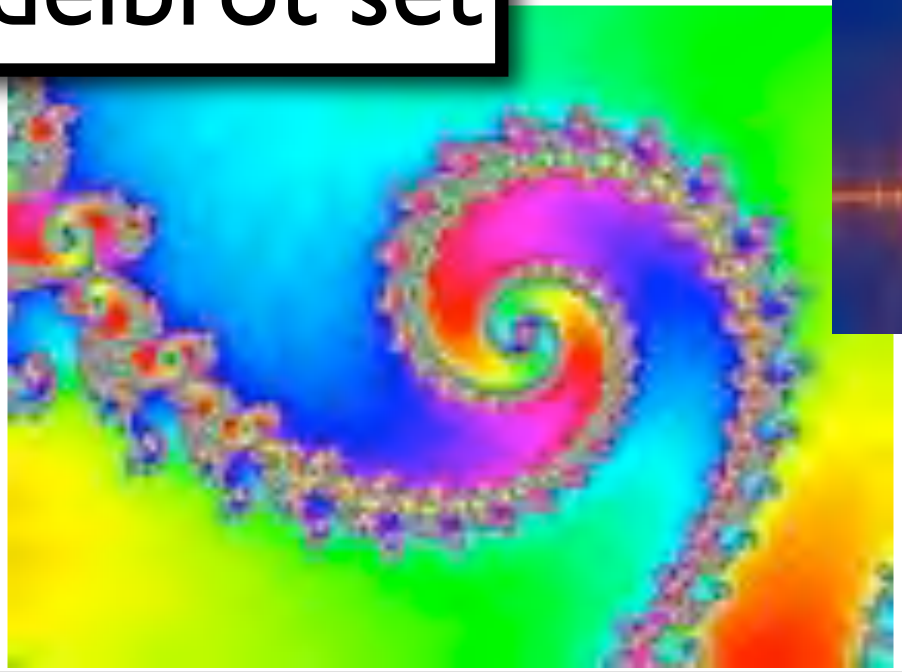


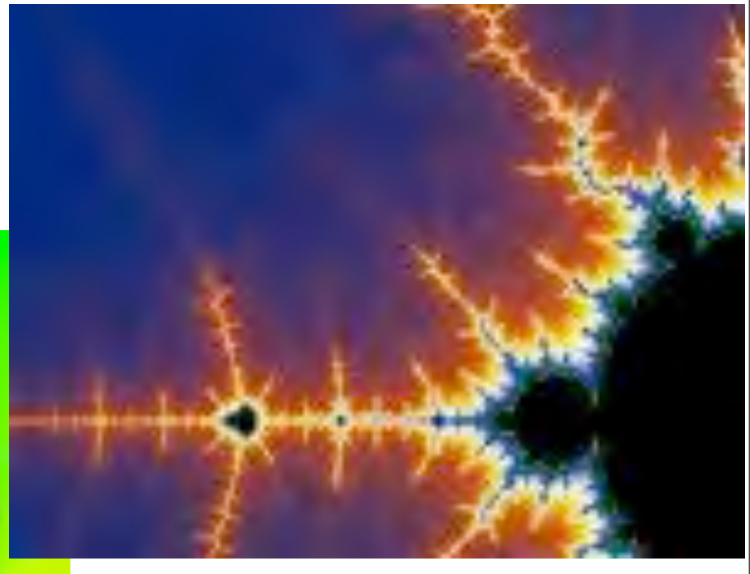
- gray level indicates speed of convergence (to fixed point or to infinity)

$$B^{[h+1]} = z + B^{[h]}(z)^2$$



Mandelbrot set



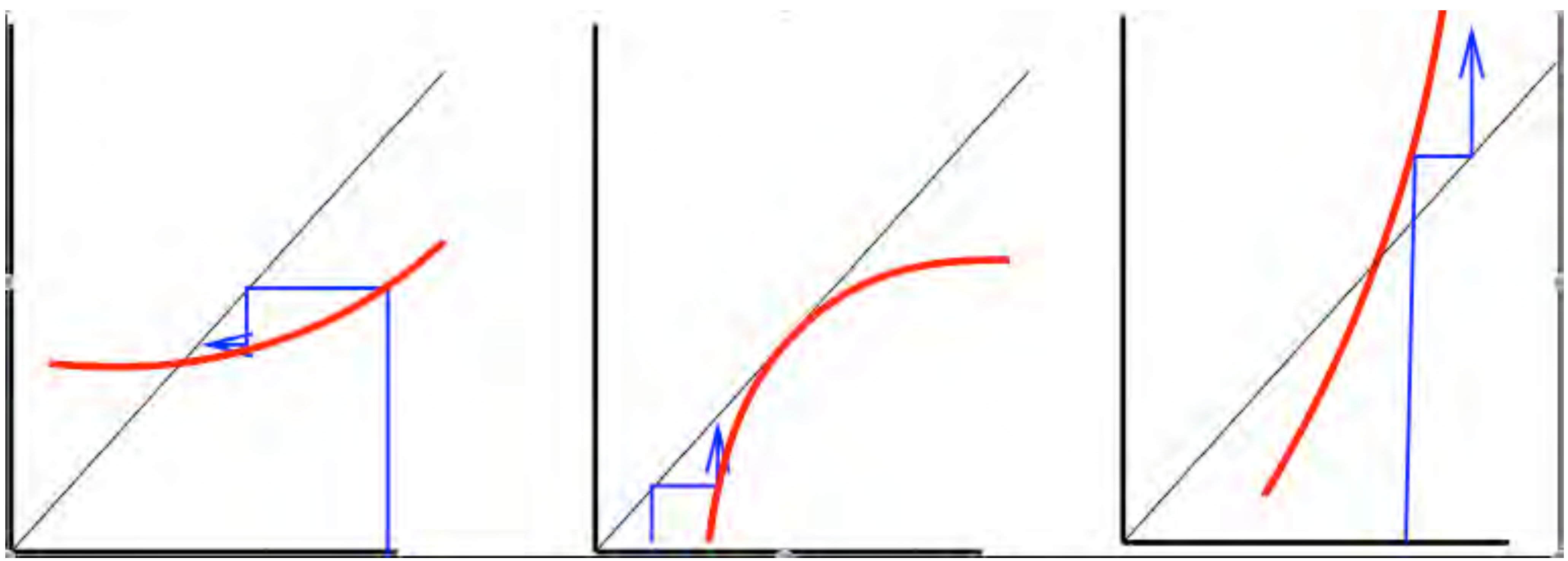


Elementary fixed-point theory

- A function $y \mapsto f(y)$.
- A **fixed point** $\xi = f(\xi)$.
- The **multiplier** $\kappa := f'(\xi)$.

locally!

attractive fixed point	**indifferent** fixed point	**repulsive** fixed point
$\lvert\kappa\rvert < 1$	$\kappa = 1$	$\lvert\kappa\rvert > 1$
$(u_{n+1} - \xi) \sim \kappa(u_n - \xi)$	$(u_{n+1} - \xi) \sim (u_n - \xi)$	—
geometric conv.	near-stationarity	divergence



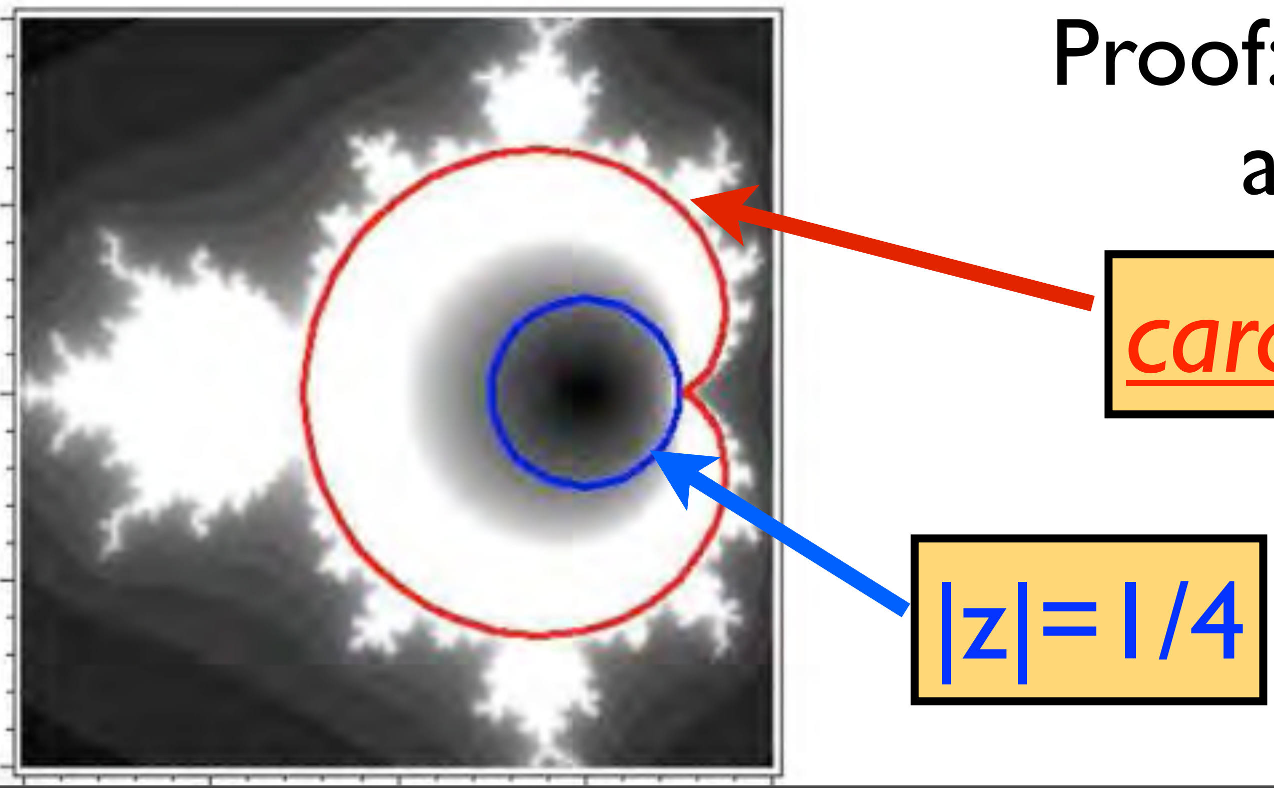
Attractive fixed point & geometric convergence

$$u_0 = z; \qquad u_{h+1} = z + u_h^2.$$

Function is $f(y) = z + y^2$; **fixed point** is $\frac{1}{2}(1 - \sqrt{1-4z})$;
multiplier is $f'(\xi) = 2\xi = 1 - \sqrt{1-4z}$.

Lemma

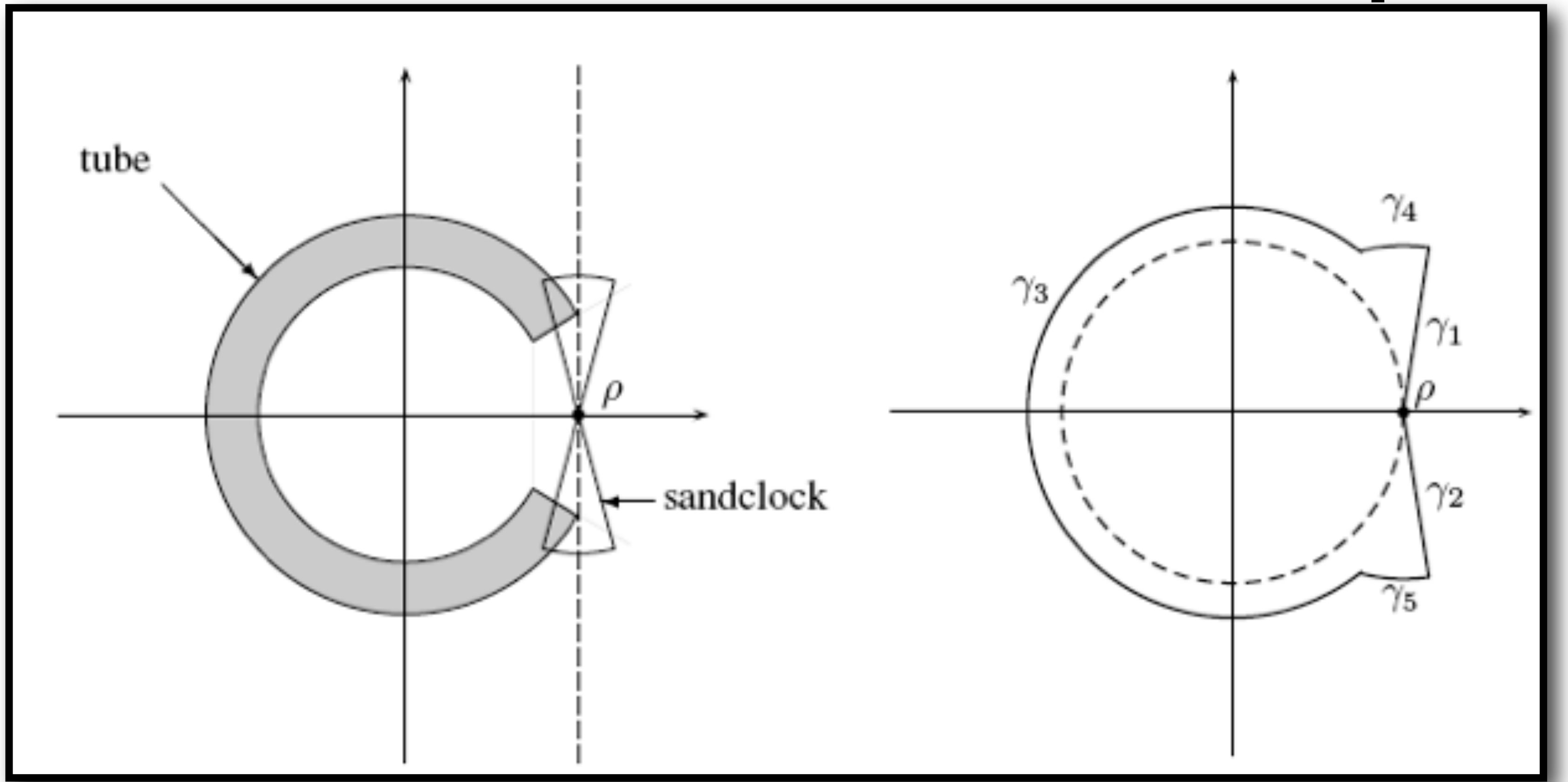
Local convergence is granted inside cardiod $|1 - \sqrt{1-4z}| < 1$.
Convergence starting from $u_0 = z$ is granted around all points of $|z| = \frac{1}{4}$, $z \neq \frac{1}{4}$ and is geometric.



Proof: convergence of B(z) on circle and continuity argument...

The tube & sandclock paradigm

From [Broutin-F., 2008-2010; height of Otter trees]



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

At the singularity 1/4

Set $e_h := B(z) - B^{[h]}(z)$ = the GF of trees with **height** $> h$**.**

At the **singularity** $z = 1/4$**:** $\boxed{e_{h+1} = e_h(1 - e_h)}$.

Convexity implies **convergence to 0**, but how fast???

$$e_{h+1} \sim e_h \quad (!!)$$

At the singularity 1/4

At the **singularity** $z = 1/4$: $\boxed{e_{h+1} = e_h(1 - e_h)}$.

♡ The trick is to **take inverses:** [De Bruijn helps]

$$\begin{cases} \dfrac{1}{e_{h+1}} & = \dfrac{1}{e_h} \cdot \dfrac{1}{1 - e_h} \\ & = \dfrac{1}{e_h} \cdot \left(1 + e_h + e_h^2 + \cdots\right) \\ & = \dfrac{1}{e_h} + \mathbf{1} + e_h + e_h^2 + \cdots . \end{cases}$$

Thus we can **bootstrap!!**: **Lower bounds** $\leftrightarrow$ **Upper bounds**.

$$\frac{1}{e_h} \sim h + \log h + \mathbf{C}(e_0) + \cdots .$$

$$\boxed{e_h \sim \frac{1}{h} - \frac{\log h}{h^2} - \frac{\mathbf{C}}{h^2} + \cdots}$$

Relation to branching processes

- An event $\mathcal{E}$ with counting generating function $E(z)$
- *Probability of* E *under critical branching process is* $\boxed{2E(1/4)}$.

(Critical B.P. $\equiv$ critical Boltzmann model.)

$$\mathbb{P}^{\text{B.P.}}(\text{tree } \tau) = \frac{1}{2^{2|\tau|+1}}.$$

Corollary (Critical —binary— branching process)

$$\mathbb{P}(\mathit{Height} \geq h) \sim \frac{2}{h}; \qquad \mathbb{P}(\mathit{Height} = h) \sim \frac{2}{h^2}.$$

Near the singularity 1/4, in **sandclock**

The "écarts" $e_h = y - u_h$ ={trees of height$>h$} satisfy:

$$\left.\begin{array}{lcl} y^2 & = & z + y^2 \\ u_{h+1} & = & z + u_h^2 \end{array}\right\} \implies e_{h+1} = 2y\left(1 - \frac{e_h}{2y}\right)\mathbf{e}_h$$

Their normalized version $e_h = (2y)^h f_h$ satisfies

$$f_{h+1} = f_h(1 - (2y)^{h+1} f_h)$$

Same player plays again: **take inverses**. . .

$$\frac{1}{f_{h+1}} = \frac{1}{f_h} + (2y)^{h+1} + \cdots$$

Lemma (Main approximation lemma: height $> h$)

$$B - B^{[h]} \approx \varepsilon \frac{(1-\varepsilon)^h}{1-(1-\varepsilon)^h}, \qquad \varepsilon := \sqrt{1-4z}.$$

Main approximation

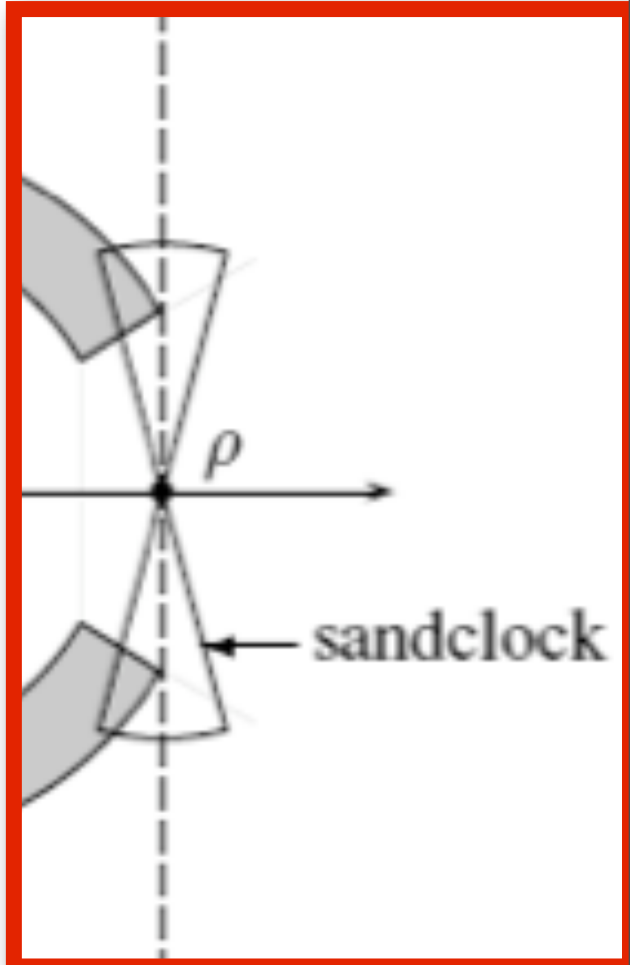
Lemma (Main approximation lemma: height $> h$)

$$B - B^{[h]} \approx \varepsilon \frac{(1-\varepsilon)^h}{1-(1-\varepsilon)^h}, \qquad \varepsilon := \sqrt{1-4z}.$$

Binary tree ~~ general Catalan trees

Perturbation of parameter
near an indifferent fixed-point
"Interpolation formula"

- For fixed $z \neq 1/4$ gives **geometric convergence**.
- For $z = 1/4$ gives **harmonic convergence**.
- With some work ... shown to hold in a sandclock.



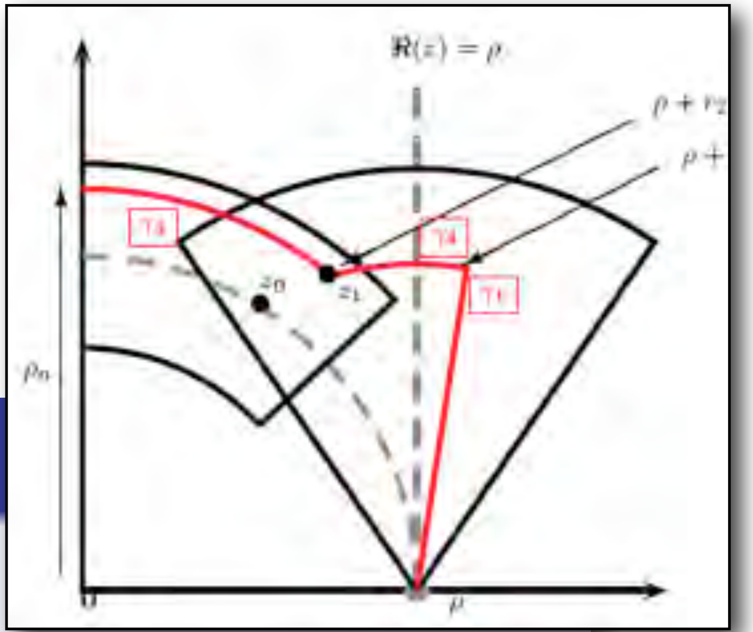
+ uniform error terms

Local & central limit law

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

Cf **SINGULARITY ANALYSIS** = Hankel-like contour near singularity.

$$\Theta(x) := \sum_{j\geq 1} e^{-j^2x^2}(4j^2x^2 - 2).$$



Theorem (Local limit law)

$$\mathbb{P}_{\mathcal{B}_n}\left(H = \lfloor 2x\sqrt{n}\rfloor\right) \sim \frac{1}{\sqrt{n}}\Theta'(x).$$

Theorem (Central limit law:)

$$\mathbb{P}_{\mathcal{B}_n}\left(H \leq \lfloor 2x\sqrt{n}\rfloor\right) \to \Theta(x)$$

$$H[\mathcal{B}_n] \approx 2H[\mathcal{G}_n]$$

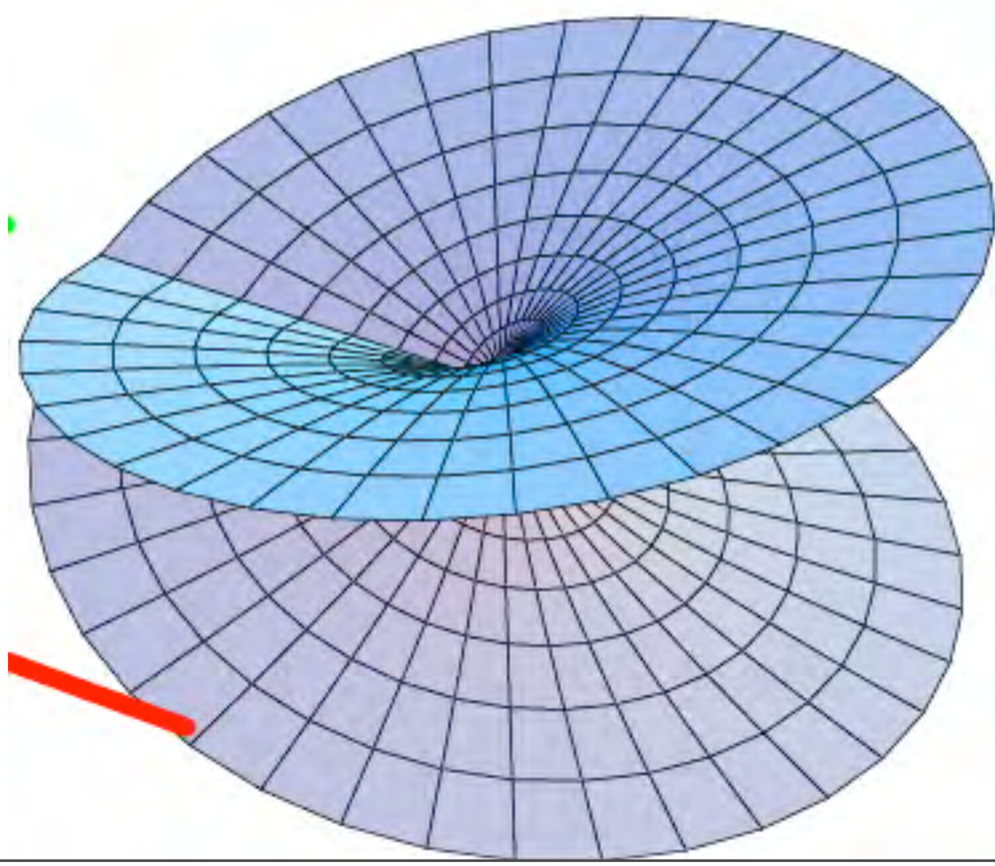
[F, Gao, Odlyzko, Richmond 1993]

3. Other stories

- Simple varieties of trees (like binary!)
- Non-plane binary trees
- Speed of convergence, Large deviations
- Balanced structures

Simple varieties of trees

- Only certain node degrees allowed
- **Universality of SQRT singularity**
- Perturbation of singular iteration succeeds



- Works also for non-plane binary trees [Broutin-F. 2008-2010]

Cf [Renyi-Szekeres 1967], for Cayley

[F, Gao, Odlyzko, Richmond]

Theorem 1.2. *Consider a simple family of trees corresponding to the equation*

$$y = z\phi(y), \;\; \phi(y) = \sum c_r y^r$$

and restrict to

$$n \equiv 1 \pmod{d} \text{ with } d = \gcd\{r : c_r \neq 0\}.$$

Let $y_n = \sum_h (y_n^{[h]} - y_n^{[h-1]})$, τ *be the smallest positive solution of*

$$\phi(\tau) - \tau\phi'(\tau) = 0$$

and set

$$c = (2\phi(\tau)\phi''(\tau))^{1/2}/\phi'(\tau) \text{ and } \beta = 2\sqrt{n}/(ch).$$

Then for any $\delta > 0$, *we have the relation*

$$\frac{y_n^{[h]} - y_n^{[h-1]}}{y_n} \sim \begin{cases} 2c\pi^{1/2}n^{-1/2}\beta^4 \sum_{m\geq 1} (m\pi)^2(2(m\pi\beta)^2 - 3)e^{-(m\pi\beta)^2} \\ 2c/(\beta\sqrt{n}) \sum_{m\geq 1} m^2(2(m/\beta)^2 - 3)e^{-(m/\beta)^2} \end{cases}$$

uniformly as $n \to \infty$, *for* $\delta^{-1}(\log n)^{-1/2} \leq \beta \leq \delta(\log n)^{1/2}$.

Speed of convergence...

- Previous methods give speed ~ $\dfrac{\log n}{\sqrt{n}}$
- Mean height is, e.g., for binary trees

$$\mathbb{E}_{\mathcal{B}_n}[H] \sim 2\sqrt{\pi n} + c \log n + c' + \frac{c'' \log n}{\sqrt{n}} + \cdots$$

[Broutin-F, in prep.]

Large deviations

- *Probability of small or large height is exponentially small:*

Theorem 1.3. *There is a $\delta > 0$ such that the number of binary trees with n internal nodes and height h, for $1 \le h \le n$, satisfies*

$$B_n - B_n^{[h]} = O\left(B_n n^{3/2} e^{-h^2/(4n)}\right),$$

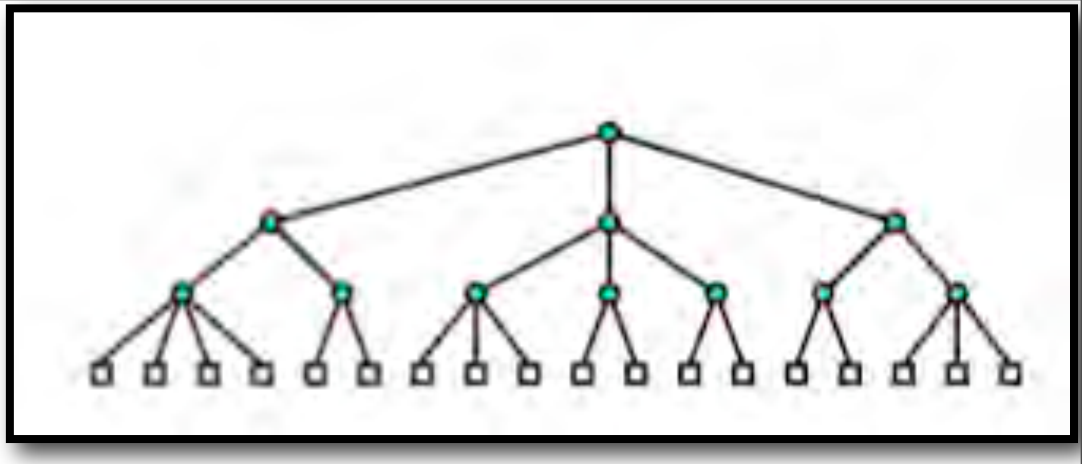
and

$$B_n^{[h]} = O\left(B_n n^{3/2} e^{-\delta n/h^2}\right).$$

Theorem 1.4.

$$B_n^{[h]} - B_n^{[h-1]} \sim \frac{4\epsilon^2 A(\epsilon)}{(1-\epsilon)^2\sqrt{\pi(1+\epsilon)n}} \left((1-\epsilon)^{(1-\epsilon)}(1+\epsilon)^{(1+\epsilon)}\right)^{-h/2\epsilon} 4^n$$

uniformly for all h such that $h/n = 2\epsilon/(1+\epsilon)$ with $\epsilon \in [\delta', 1-\delta']$, where δ' is a positive constant, which can be arbitrarily small, and $A(\epsilon)$ is a positive and continuous function for $\epsilon \in [\delta', 1-\delta']$.

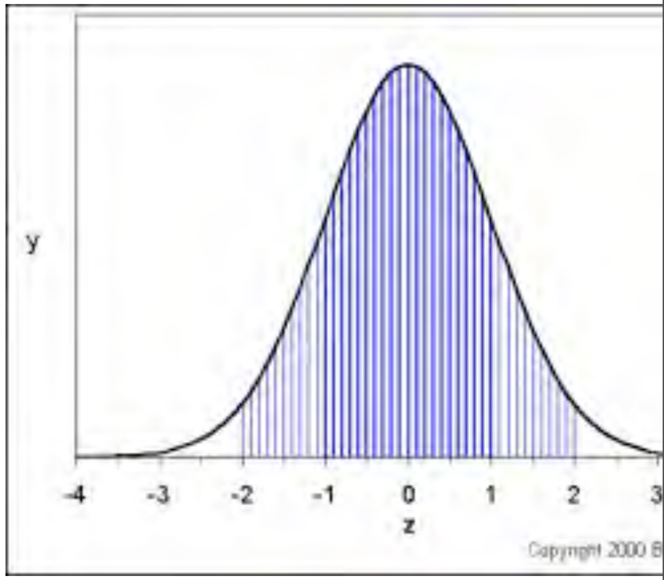


★ $\boxed{p_{h+1} = z + p_h^2; \quad p_0 = z}$

≡ Take a random binary tree of height h. Distribution of size?

★ $\boxed{p_{h+1} = p_h^2 + p_h^3; \quad p_0 = z}$

≡ Take a random balanced 2–3 tree of height h. Size?



How are the coefficients of $p_h(z)$?

Answer: Just like $\boxed{p_{h+1} = p_h^2}$, i.e., $p_h = (1+z)^{2^h}$

Technique: $p_h(1)$ grow doubly exponentially fast and satisfy exact formula $\boxed{p_h(1) = \lfloor \alpha^{2^h} \rfloor}$. Then, **perturbation + saddle point**

Gaussian coeffs

Theorem (F-Odlyzko, ~~2094~~) 1984 (!!)

Coefficients of polynomials that satisfy $p_{h+1} = P(z, p_h)$, for P a nonlinear and positive polynomial obey a local Gaussian law.

Diameter

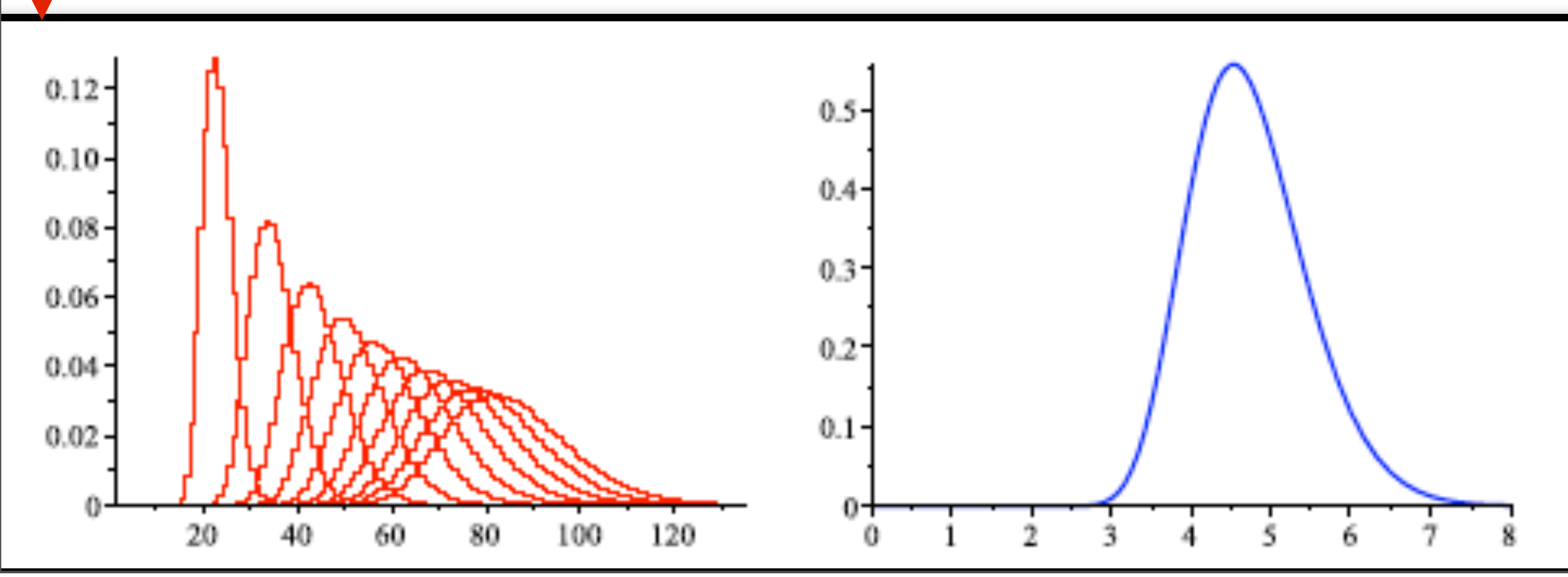
- Done by [Szekeres 1982] for unrooted Cayley trees
- Done by [Broutin-F. 2010] for binary(Otter) trees

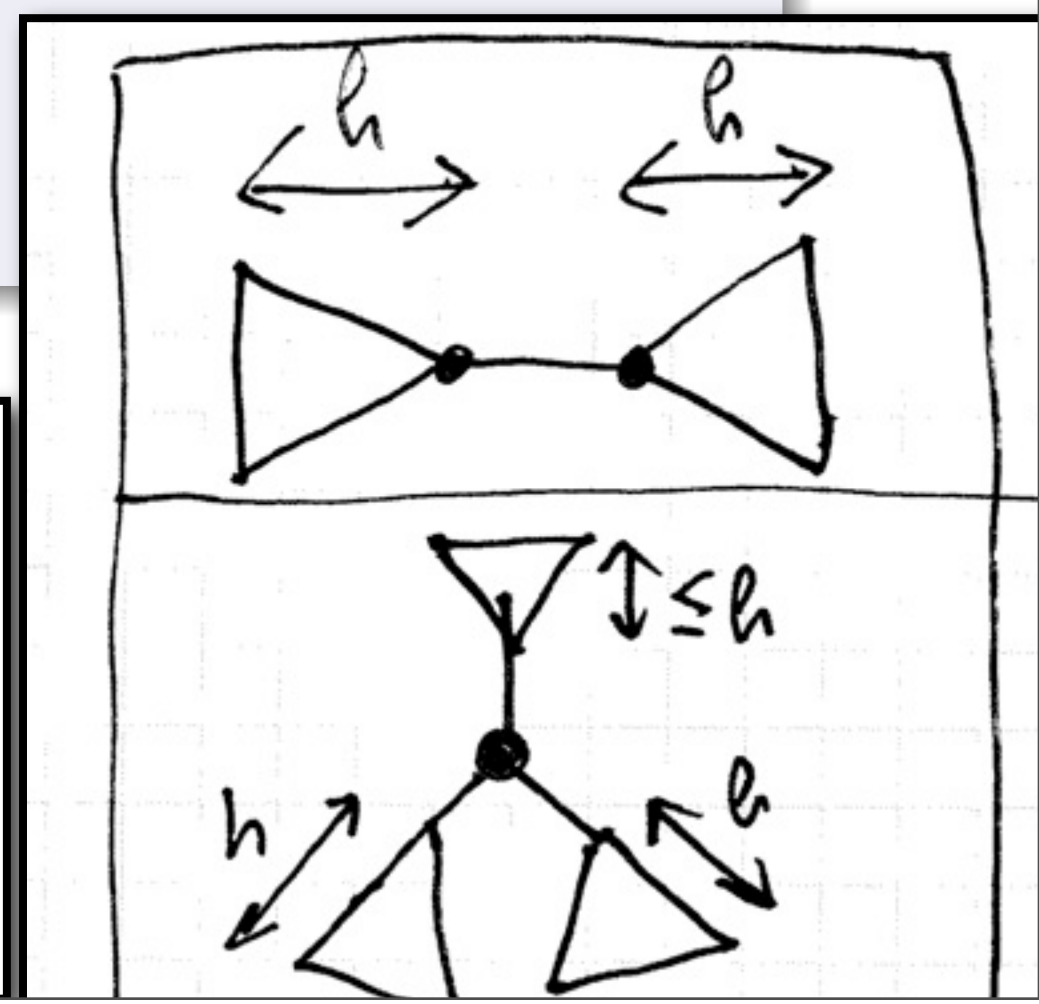
A theta-like distribution. Also quantify proportion of central/bicentral trees = agrees with Aldous' model of CRT. Cf Haas, Miermont, Marckert 2009–2010.

Theorem

The ratio of expected diameter (D) to expected height satisfies

$$\lim_{n\to\infty} \frac{\mathbb{E}_n(D)}{\mathbb{E}_n(H)} = \frac{4}{3}.$$

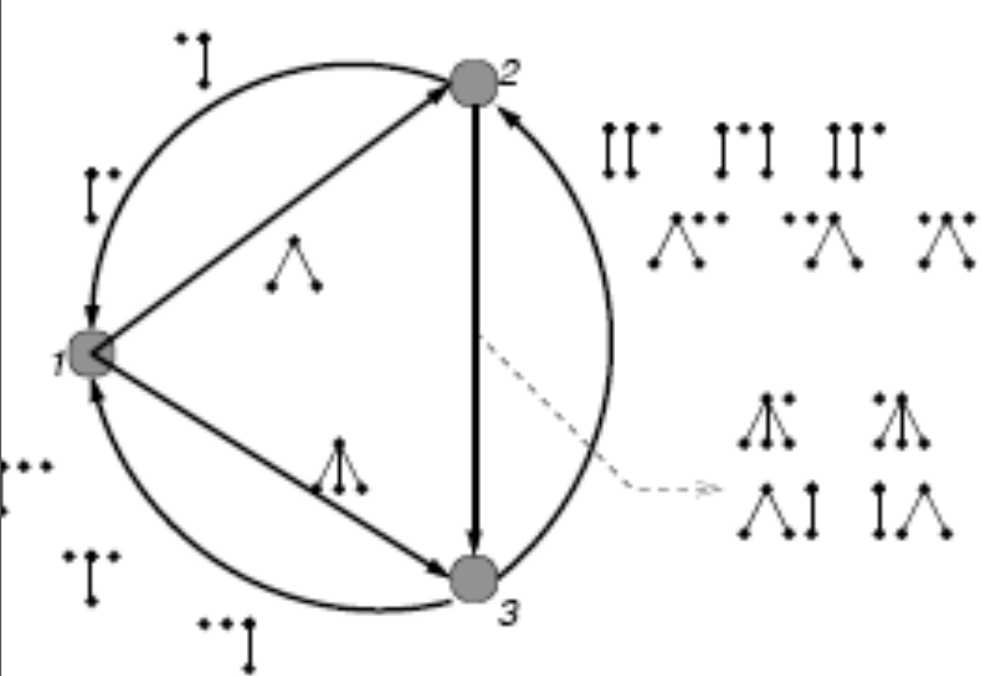




Width?

$$\mathbb{E}_n(W) = \sqrt{\frac{\pi n}{2}} + O\left(n^{1/4}\sqrt{\log n}\right), \qquad \mathbb{P}_n(\sqrt{2}W \le x) \to 1 - \Theta(x),$$

- *Width* is accessible by properties of Brownian motion (as is height) : A definitive treatment is [Chassaing-Marckert-Yor]
- Analysis? Transfer matrix methods:



▷ **V.45.** *A question on width polynomials.* It is unknown w
true. The smallest positive root ρ_k of the denominator of $Y^{[k]}$

$$\rho_k = \rho + \frac{c}{k^2} + o(k^{-2}),$$

Entre deux vérités du domaine réel, le chemin le plus facile et le plus court passe bien souvent par le domaine complexe.

PAUL PAINLEVÉ [467, p. 2]

It has been written that the shortest and best way between two truths of the real domain often passes through the imaginary one[1].

— JACQUES HADAMARD [316, p. 123]

Analytic methods are extremely powerful and when they apply, they often yield estimates of unparalleled precision.

— ANDREW ODLYZKO [461]