CONCENTRATION PROPERTIES OF EXTREMAL PARAMETERS IN RANDOM TREES*

Michael Drmota

Inst. of Discrete Mathematics and Geometry

Vienna University of Technology, A 1040 Wien, Austria

michael.drmota@tuwien.ac.at

www.dmg.tuwien.ac.at/drmota/

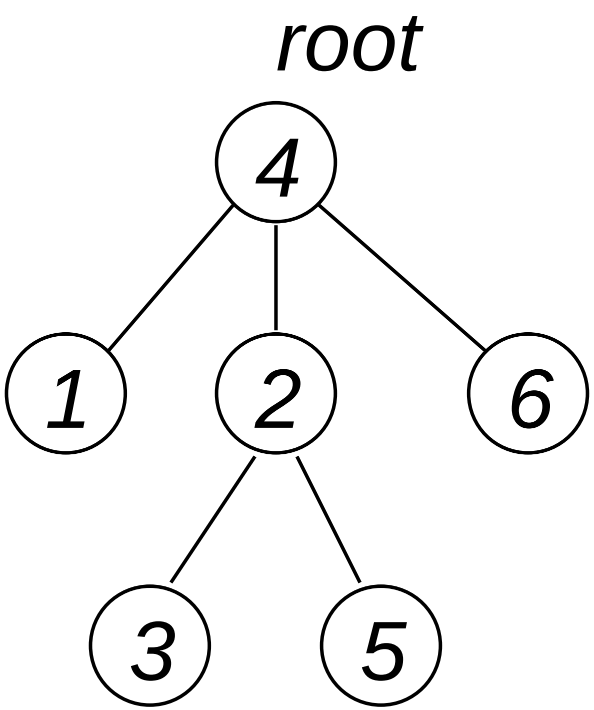
* supported by the Austrian Science Foundation FWF, grant S9600.

Outline of the Talk

- $\sqrt{n}$-Trees
- Recursive Trees
- Plane Oriented Trees
- Extremal Parameters
- Types of Concentration
- Results
- Proof Methods

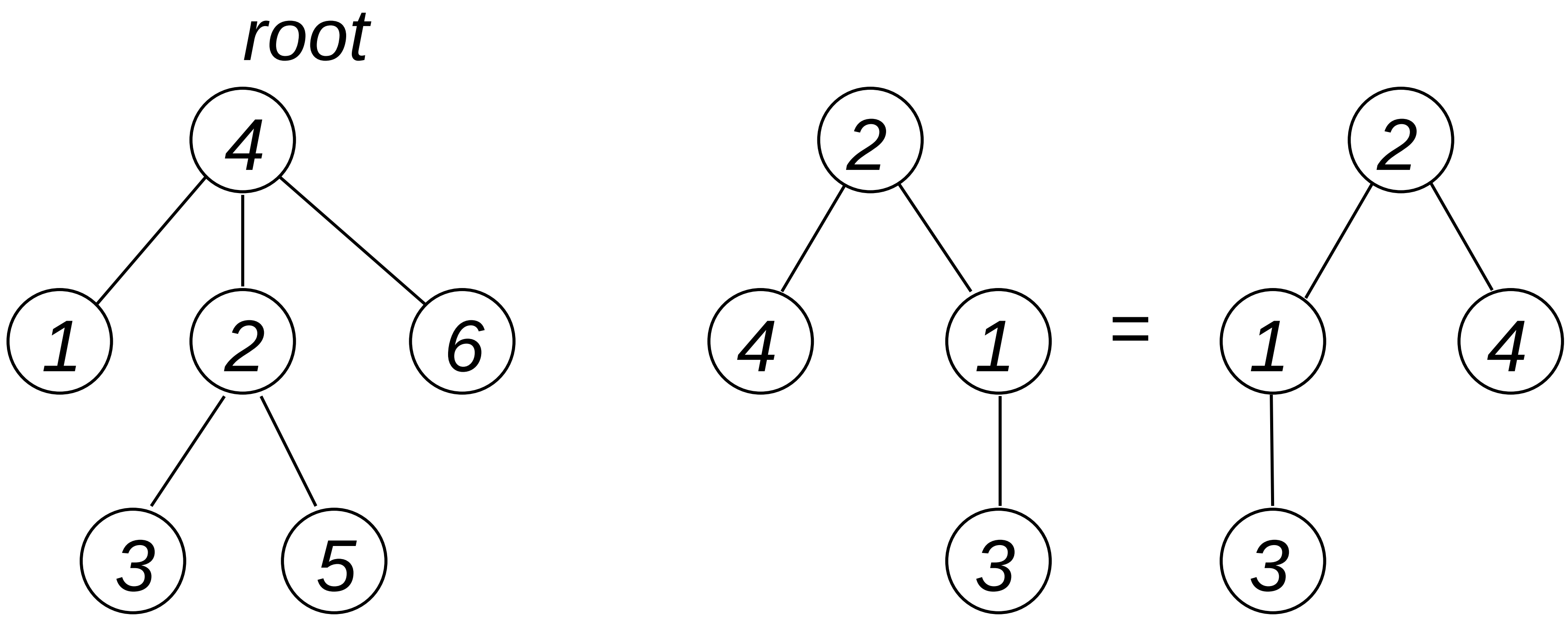
$\sqrt{n}$- Trees

Cayley Trees: labeled, rooted, non-planar



$\sqrt{n}$- Trees

Cayley Trees: labeled, rooted, non-planar



$\sqrt{n}$- Trees

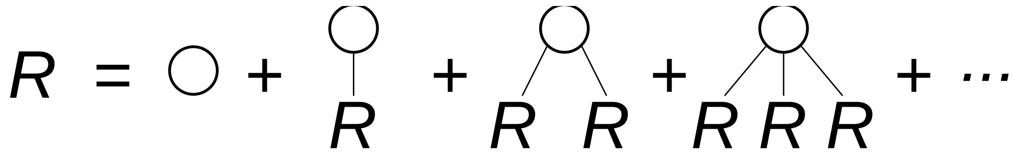
Cayley Trees:

- t_n ... number of Cayley trees of size n
- $\hat{t}(x) = \sum_{n\ge 1} t_n \frac{x^n}{n!}$ ... generating function

$\sqrt{n}$- Trees

Cayley Trees:

- Recursive description: A Cayley tree can be interpreted as a root followed by an unordered sequence of Cayley trees.

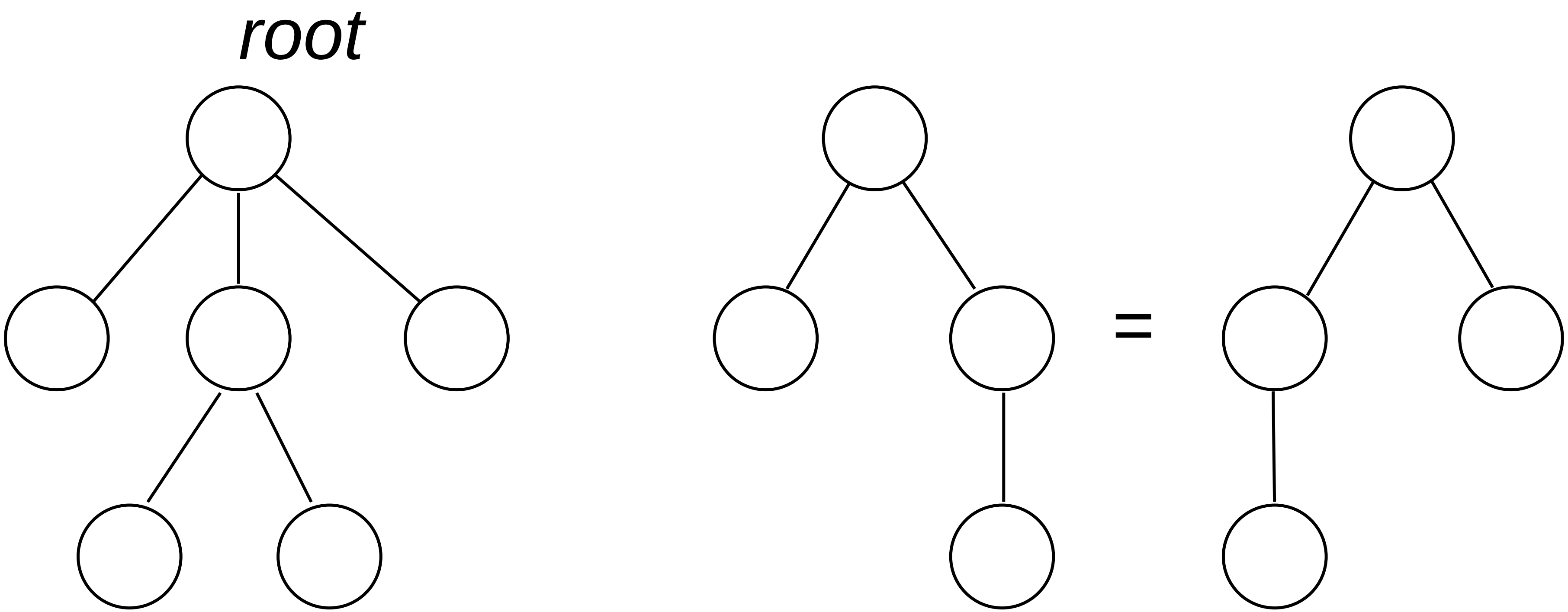


$$\hat{t}(x) = x + x\hat{t}(x) + x\frac{\hat{t}(x)^2}{2!} + x\frac{\hat{t}(x)^3}{3!} + \cdots = x\, e^{\hat{t}(x)}$$

- $t_n = n^{n-1}$ ... by Lagrange inversion

$\sqrt{n}$- Trees

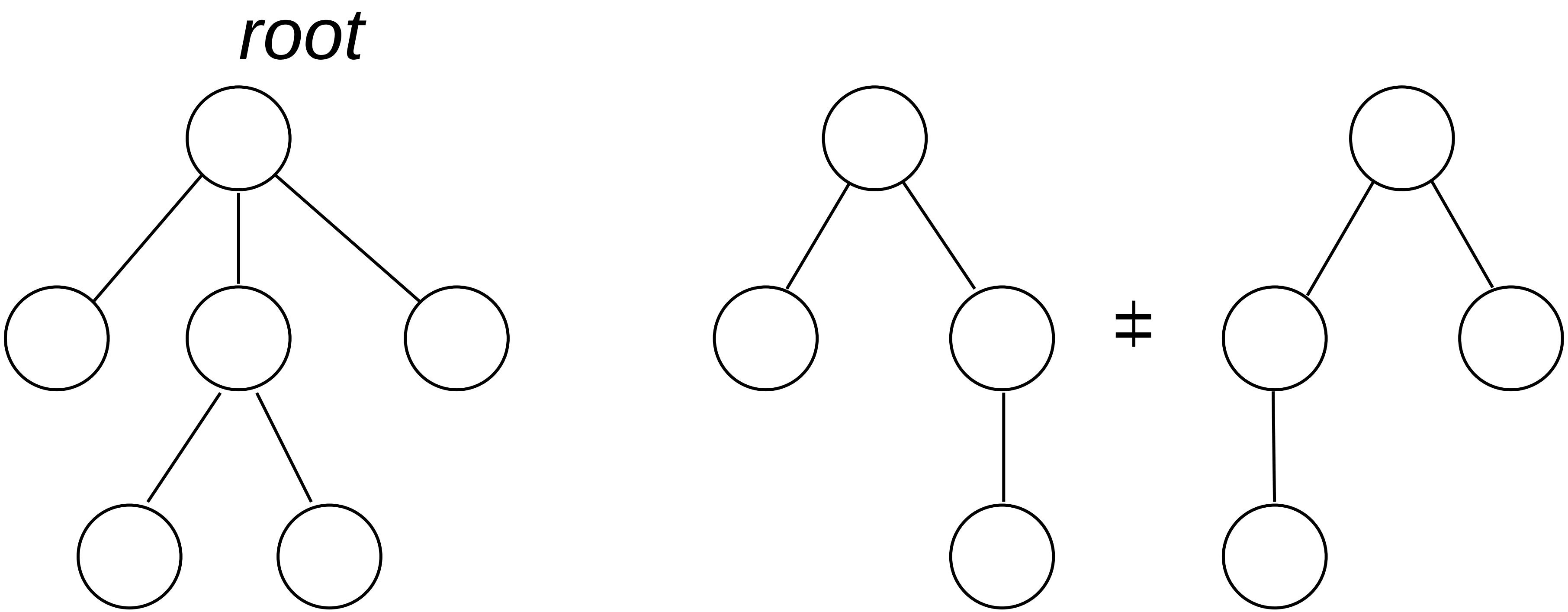
Polya Trees: unlabeled, rooted, non-planar



$$t(x) = \sum_{n \geq 1} t_n \, x^n \qquad t(x) = x \, e^{t(x) + \frac{1}{2} t(x^2) + \frac{1}{3} t(x^3) + \cdots}$$

$\sqrt{n}$- Trees

Planted Plane Trees: unlabeled, rooted, planar



$$t(x) = \sum_{n \geq 1} t_n \, x^n \qquad t(x) = \frac{x}{1 - t(x)} \qquad t_n = \frac{1}{n+1}\binom{2n}{n}$$

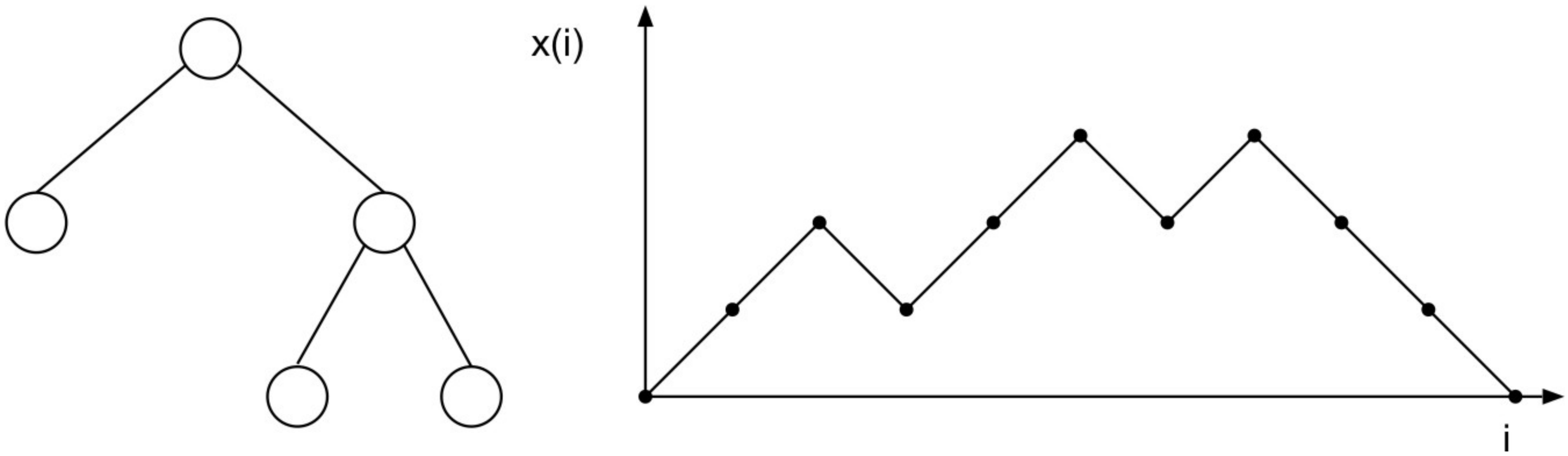
$\sqrt{n}$- Trees

Common Properties

- recursive description
- functional equation for generating function
- height and width are of order $\sqrt{n}$
- stochastic approximation by Brownian excursion

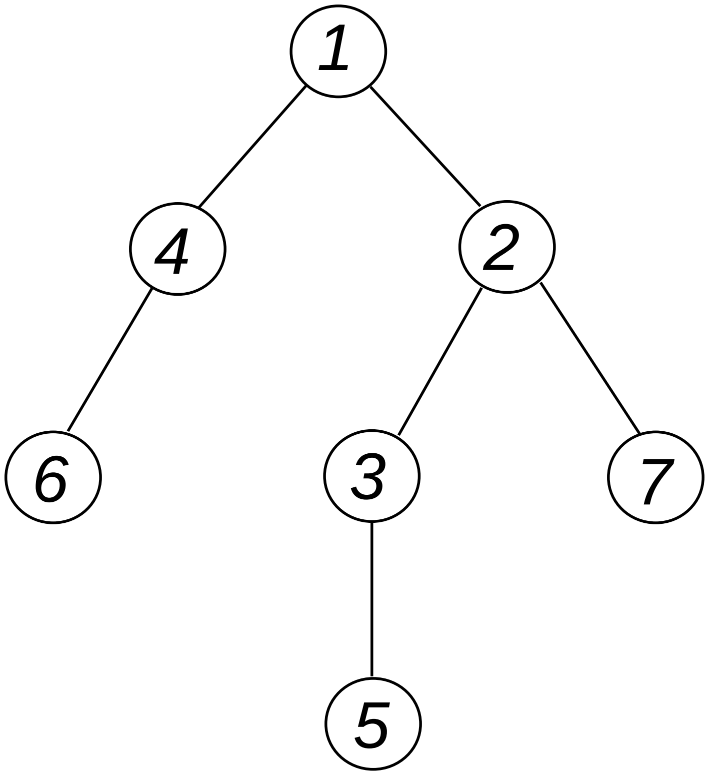
$\sqrt{n}$- Trees

Depth-first search.



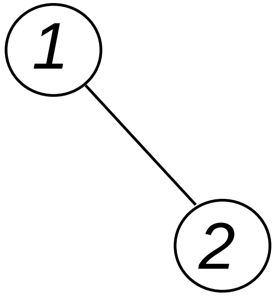
$$\frac{T_n(\lfloor 2nt \rfloor)}{c_2\sqrt{n}} \to e(t) \quad \ldots \text{ Brownian excursion}$$

Recursive Trees

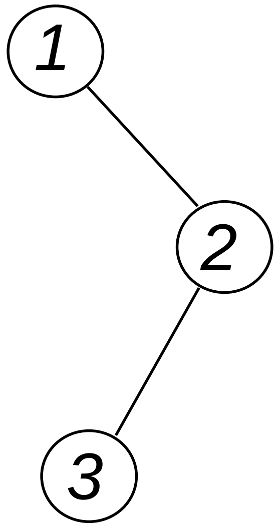


Recursive Trees

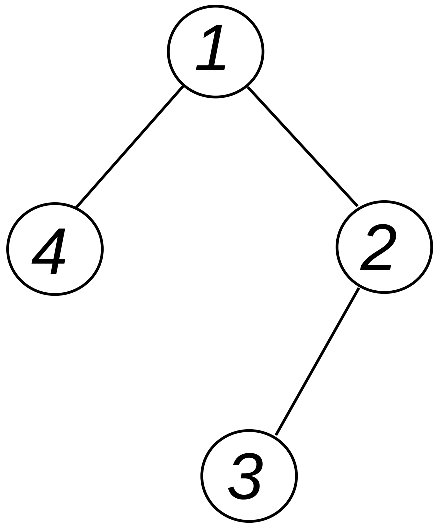
Recursive Trees



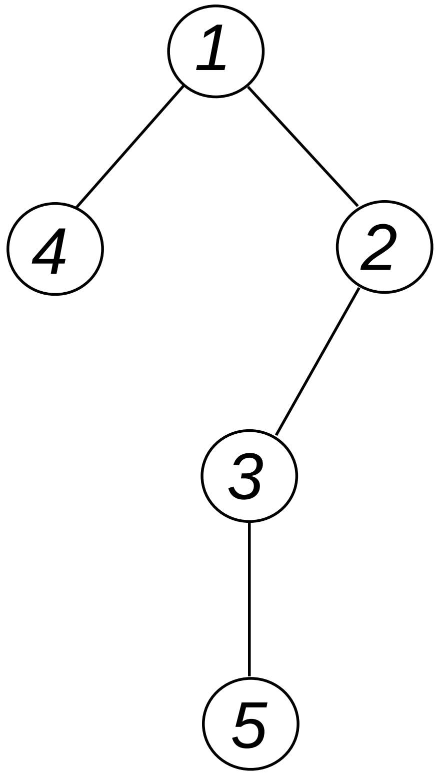
Recursive Trees



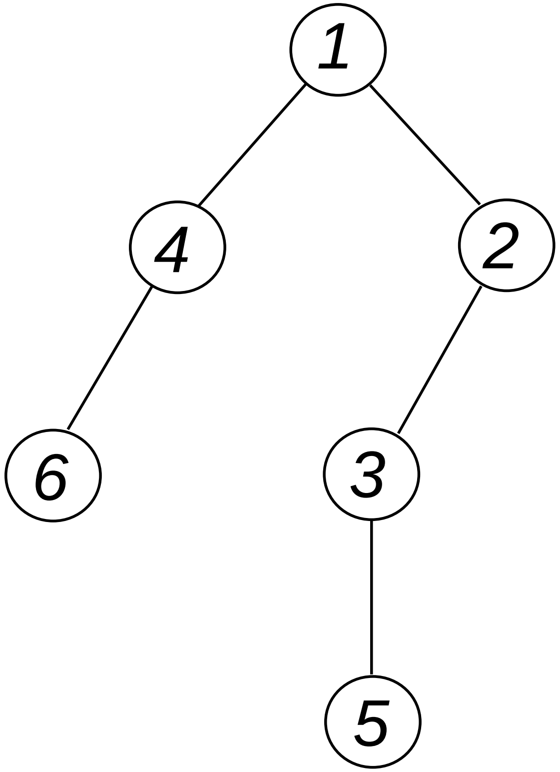
Recursive Trees



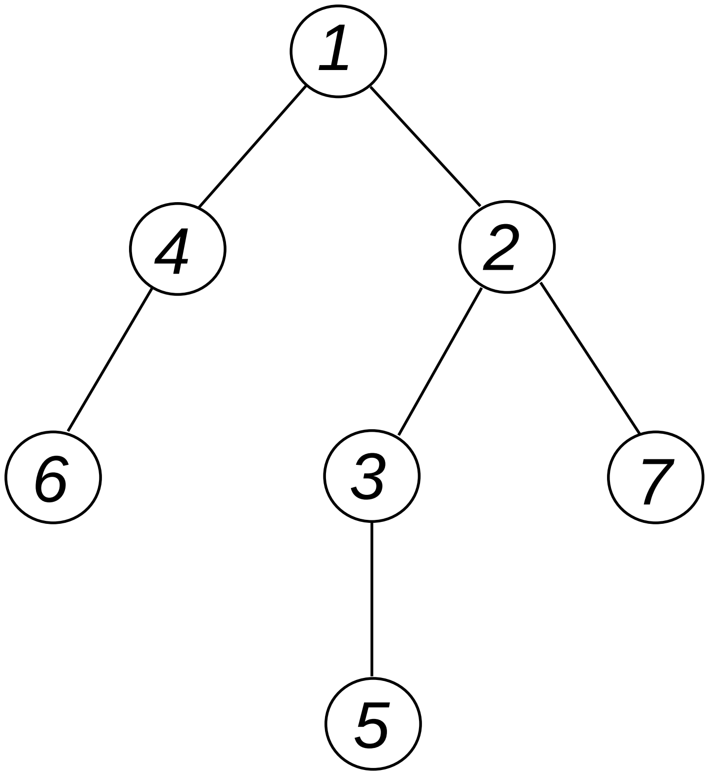
Recursive Trees



Recursive Trees



Recursive Trees



Recursive Trees

Combinatorial Description:

- labeled rooted tree
- labels are strictly increasing (starting at the root)
- no left-to-right order (non-planar)

Recursive Trees

Number of Recursive Trees:

$$\begin{aligned} y_n &= \text{number of recusive trees of size } n \\ &= (n-1)! \end{aligned}$$

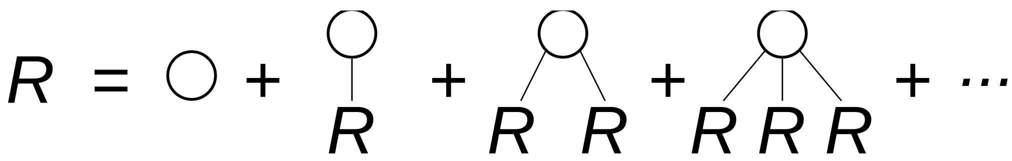
The node with label j has exactly $j-1$ possibilities to be inserted $\Longrightarrow y_n = 1 \cdot 2 \cdots (n-1)$.

Recursive Trees

Generating Functions:

$$y(x) = \sum_{n \geq 1} y_n \frac{x^n}{n!} = \sum_{n \geq 1} \frac{x^n}{n} = \log \frac{1}{1-x}$$

$$\boxed{y'(x) = 1 + y(x) + \frac{y(x)^2}{2!} + \frac{y(x)^3}{3!} + \cdots = e^{y(x)}}$$



A recursive tree can be interpreted as a root followed by an unordered sequence of recursive trees. ($y'(x) = \sum_{n \geq 0} y_{n+1} x^n / n!$)

Recursive Trees

Probability Model:

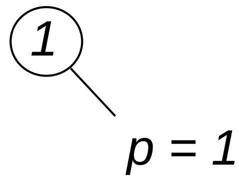
Process of growing trees

- The process starts with the root that is labeled with 1.
- At step j a new node (with label j) is attached to any previous node with probability $1/(j-1)$.

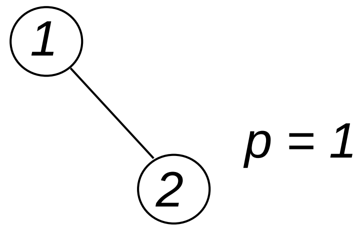
After n steps every tree (of size n) has equal probability $1/(n-1)!$.

Recursive Trees

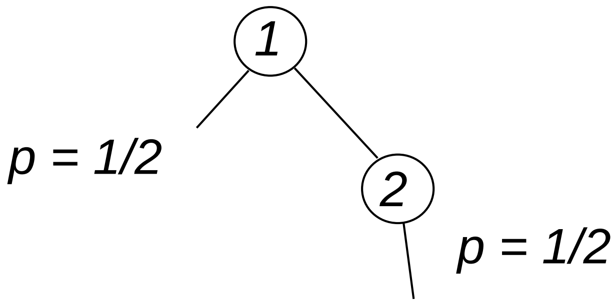
Recursive Trees



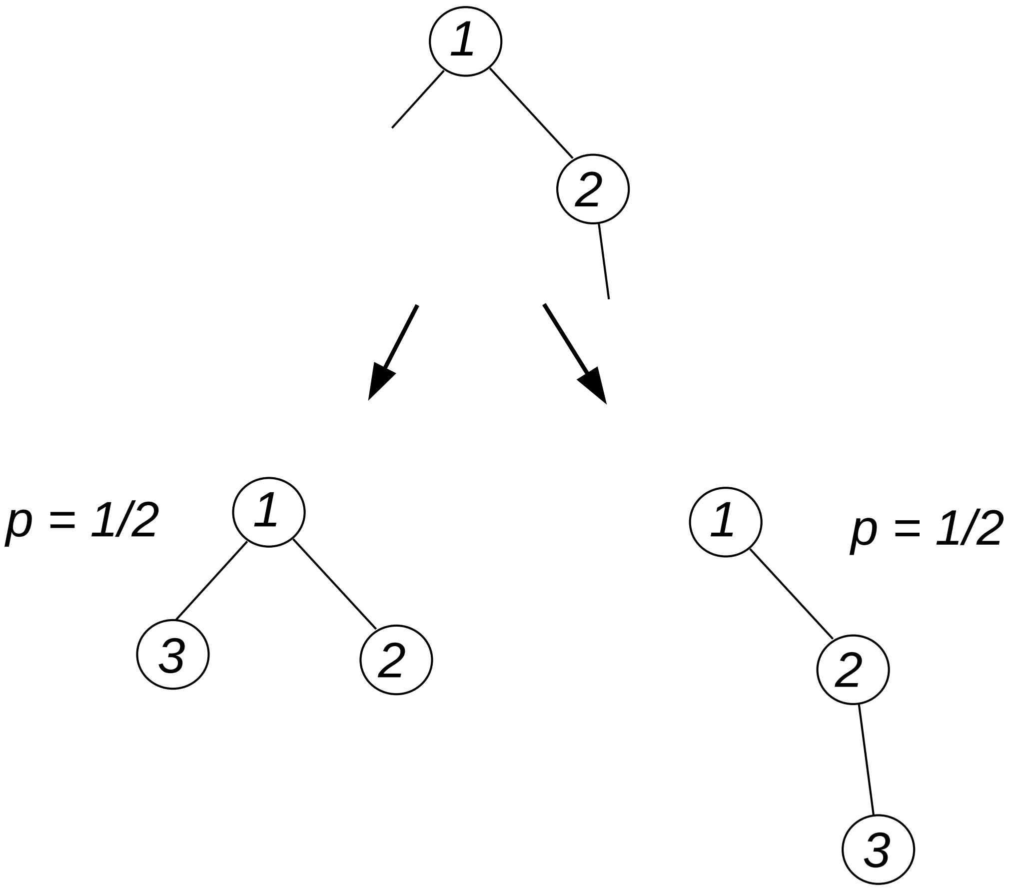
Recursive Trees



Recursive Trees

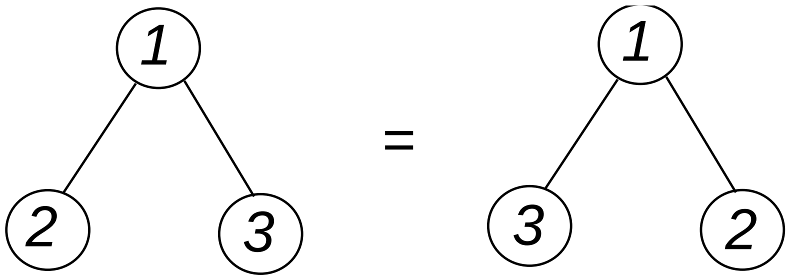


Recursive Trees



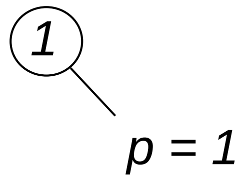
Recursive Trees

Remark

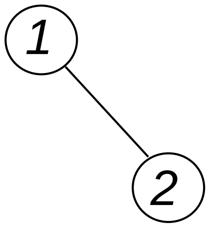


Plane Oriented Trees

Plane Oriented Trees

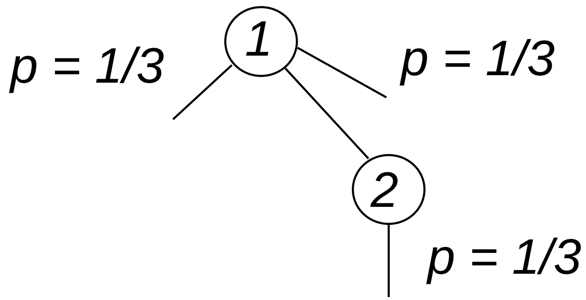


Plane Oriented Trees

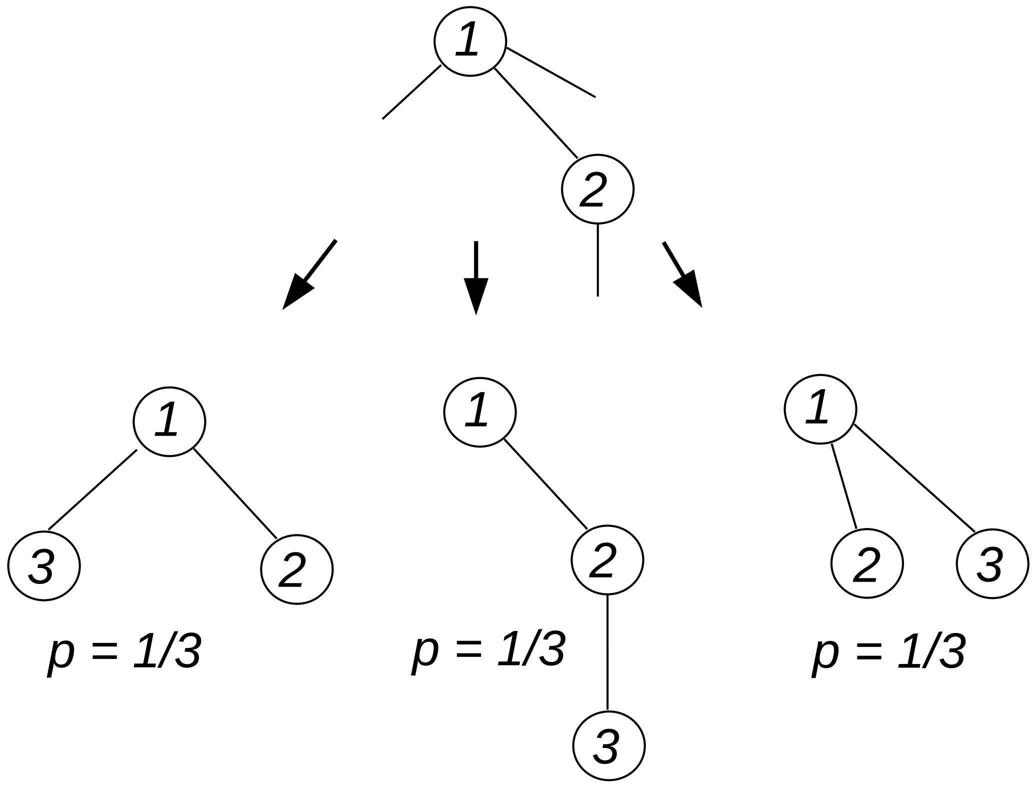


$p = 1$

Plane Oriented Trees

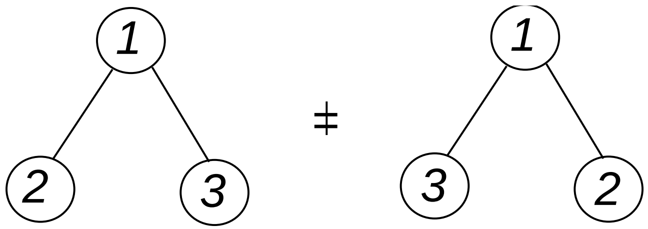


Plane Oriented Trees



Plane Oriented Trees

Remark



Plane Oriented Trees

Number of Plane Oriented Trees:

$$\begin{aligned} y_n &= \text{number of plane oriented trees of size } n \\ &= 1 \cdot 3 \cdot 5 \cdots (2n-3) = (2n-3)!! \\ &= \frac{(2n-2)!}{2^{n-1}(n-1)!} \end{aligned}$$

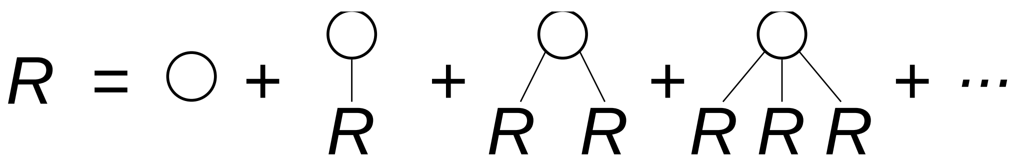
The node with label j has exactly $2j-3$ possibilities to be inserted $\Longrightarrow y_n = 1 \cdot 3 \cdots (2n-3)$.

Plane Oriented Trees

Generating Functions:

$$y(x) = \sum_{n\geq 1} y_n \frac{x^n}{n!} = \sum_{n\geq 1} \frac{1}{2^{n-1}}\binom{2(n-1)}{n-1}\frac{x^n}{n} = 1 - \sqrt{1-2x}$$

$$\boxed{y'(x) = 1 + y(x) + y(x)^2 + y(x)^3 + \cdots = \frac{1}{1-y(x)}}$$



A plane oriented tree can be interpreted as a root followed by an **ordered** sequence of plane oriented trees. $(y'(x) = \sum_{n\geq 0} y_{n+1} x^n/n!)$

Plane Oriented Trees

Probability Model:

Process of growing trees

- The process starts with the root that is labeled with 1.

- At step j a new node (with label j) is attached to any previous node of outdegree d with probability $(d+1)/(2j-3)$.

After n steps every tree (of size n) has equal probability $1/(2n-3)!!$.

Scale Free Trees

Probability Model:

Process of growing trees

- The process starts with the root that is labeled with 1.
- At step j a new node (with label j) is attached to any previous node of outdegree d with probability proportial to $\boxed{d + r}$ (for some $r > 0$).

For $d = 1$ we get plane oriented trees.

Scale Free Trees

Generating Functions

y_n ... weighted sum of plane oriented trees (according to probability distribution)

$y(x) = \sum_{n \ge 1} y_n \frac{x^n}{n!}$ ... generating function

$$\boxed{y'(x) = \frac{1}{(1 - y(x))^r}}$$

$$\Longrightarrow \quad \boxed{y(x) = 1 - (1 - (r+1)x)^{\frac{1}{r+1}}}$$

Scale Free Trees

Degree distribution

Set

$$\begin{aligned}\lambda_d &= \lim_{n\to\infty} \mathbf{P}\,(\text{a random node in a tree of size } n \text{ has out-degree } d) \\ &= \lim_{n\to\infty} \frac{\text{expected number of nodes with out-degree } d}{n}\end{aligned}$$

Then

$$\boxed{\lambda_d = \frac{(r+1)\Gamma(2r+1)\Gamma(r+d)}{\Gamma(r)\Gamma(2r+d+2)}}$$

We have a **scalefree distribution**

$$\lambda_d \sim \frac{(r+1)\Gamma(2r+1)}{\Gamma(r)} \cdot d^{-2-r}.$$

Extremal Parameters

We focus on the following two extremal tree parameters:

- $D_n =$ maximum degree in a tree of size n
- $H_n =$ height of a tree of size n

Further interesting (extremal) parameters: diameter, width, ...

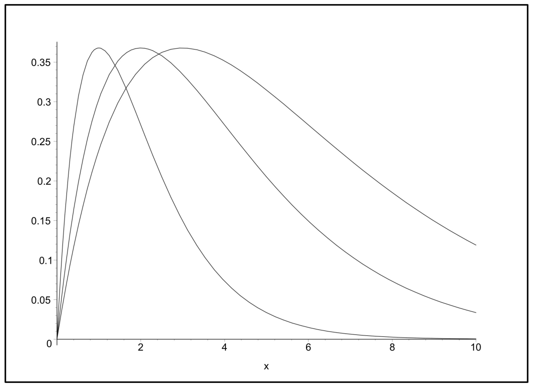
Concentration Properties

X_n ... non-negative, integer valued random variable with $\boxed{\mathbf{E}\, X_n \to \infty}$

Type 1: **No Concentration:**

$$\boxed{\frac{X_n}{\mathbf{E}\, X_n} \to Y \quad \ldots \text{ not concentrated at } 1}$$

Typically: $\mathbf{E}\, X_n^2 \sim c \cdot (\mathbf{E}\, X_n)^2$ for some $c > 1$.



Concentration Properties

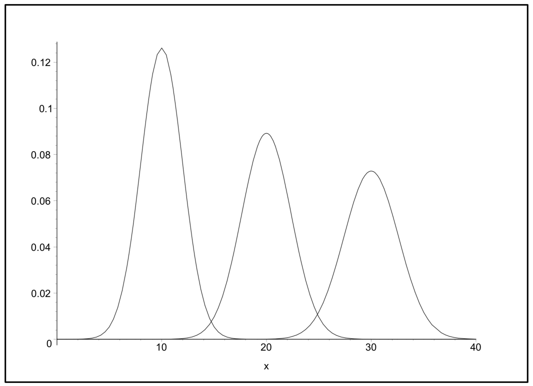
Type 2: Weak Concentration:

$$\boxed{\frac{X_n}{\mathbf{E}\, X_n} \to \delta_1 \quad \ldots \text{ concentrated at } 1}$$

Typically: $\mathbf{E}\, X_n^2 \sim (\mathbf{E}\, X_n)^2$.
(This condition implies weak concentration via Chebyshev's inequality.)

E.g. **Central Limit Theorem**

$$\frac{X_n - \mathbf{E}\, X_n}{\sqrt{\mathbf{Var}\, X_n}} \to N(0,1).$$



Concentration Properties

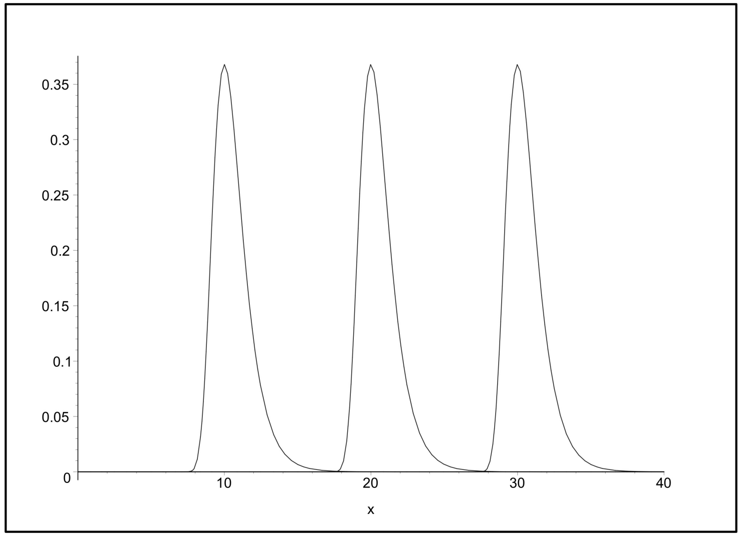
Type 3: Strong Concentration:

$$\boxed{X_n - \mathbf{E}\,X_n \ \ldots \ \text{bounded moments}}$$

Typically: **travelling wave** $F(x)$

$$\boxed{\mathbf{P}\{X_n \le k\} = F(k - m(n)) + o(1)}$$

($m(n)$ is close to the median of X_n)



Concentration Properties

Type 4: Very Strong Concentration:

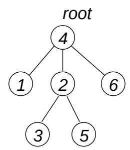
Concentration on two (or finitely many values):

$$\boxed{\mathbf{P}\{X_n = m(n) \text{ or } X_n = m(n) + 1\} = 1 + o(1).}$$

with $m(n) \to \infty$.

Results

Cayley trees



- [Meir & Moon, Carr & Goh & Schmutz]

$$\boxed{\mathbf{P}\{D_n = m(n) \text{ or } X_n = m(n) + 1\} = 1 + o(1).}$$

for some $m(n)$ with $m(n) \sim \dfrac{\log n}{\log \log n}$.

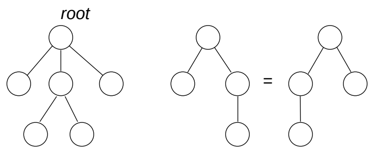
- [Flajolet & Odlyzko]

$$\boxed{\frac{H_n}{\sqrt{n}} \to Y}$$

with $Y = c_2 \cdot \max_{0 \le t \le 1} e(t)$.

Results

Polya trees



- [Goh & Schmutz]

$$\boxed{\mathbf{P}\{D_n \le k\} = \exp\left(-c_0 \eta^{k-\mu_n}\right) + o(1)}$$

with $c_0 = 3.262\ldots$, $\eta = 0.3383\ldots$, and $\mu_n = 0.9227\ldots \cdot \log n$.

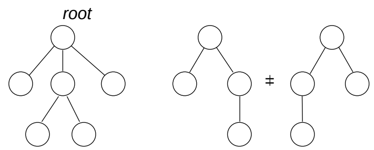
- [D. & Gittenberger]

$$\boxed{\frac{H_n}{\sqrt{n}} \to Y}$$

with $Y = c_2 \cdot \max_{0 \le t \le 1} e(t)$.

Results

Planted plane trees



- $$\boxed{\mathbf{P}\{D_n \le k\} = \exp\left(-2^{-(k-\log_2 n+1)}\right) + o(1)}$$

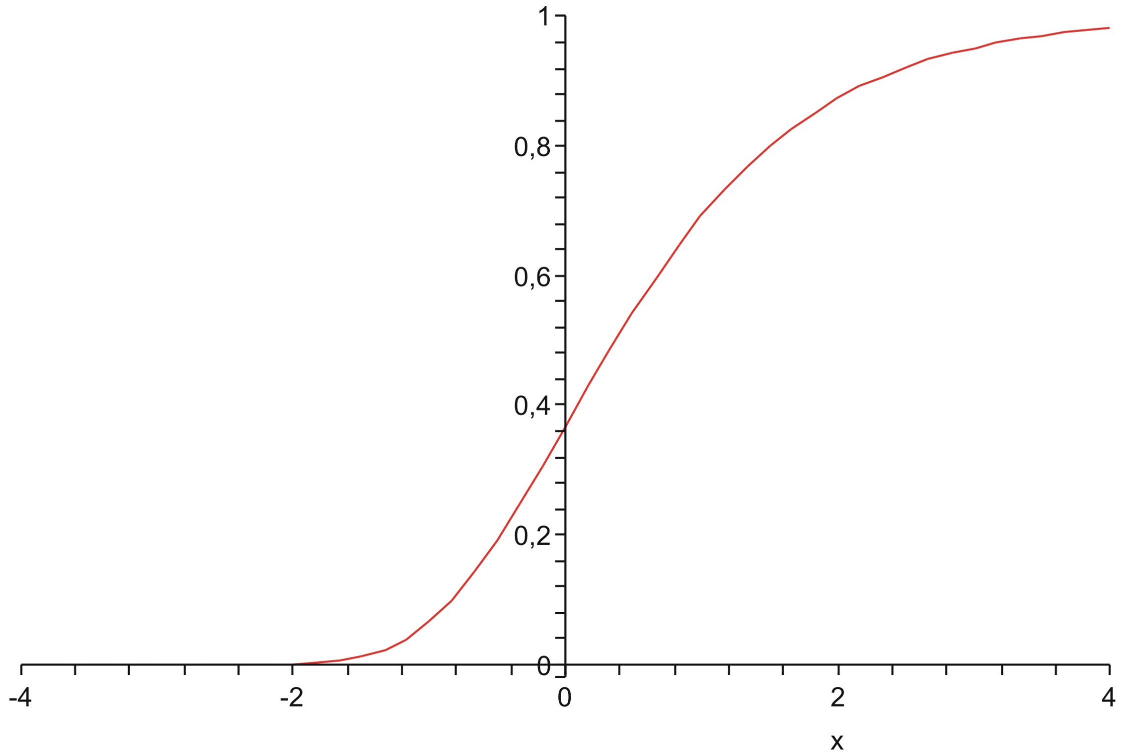
- [De Bruijn & Knuth & Rice]

$$\boxed{\frac{H_n}{\sqrt{n}} \to Y}$$

with $Y = c_2 \cdot \max_{0 \le t \le 1} e(t)$.

Results

Travelling wave for maximum degree: $F(x) = \exp(-e^{-x})$
(Extreme Value Distribution)



Results

Non-concentration of height

$$H_n = \max_{j \geq 0} T_n(j), \qquad \frac{T_n(\lfloor c_1 n t \rfloor)}{c_2 \sqrt{n}} \to e(t)$$

$$\Longrightarrow \qquad \frac{H_n}{\sqrt{n}} \to c_2 \cdot \max_{0 \leq t \leq 1} e(t)$$

$$\Longrightarrow \qquad \boxed{\text{no concentration}}$$

Results

Recursive trees

- [Goh & Schmutz]

$$\boxed{\mathbf{P}\{D_n \le k\} = \exp\left(-2^{-(k-\log_2 n+1)}\right) + o(1)}$$

- [D.]

$$\boxed{\mathbf{P}\{H_n \le k\} = F(c_k - e\log n) + o(1),}$$

where $c_k = k + O(\log k)$, $F(x) = \Psi(e^{-x})$, and $\Psi(y)$ satisfies the integral equation

$$\Psi(y/e^{1/e}) = \frac{1}{y}\int_0^y \Psi(z/e^{1/e})\Psi(y-z)\,dz.$$

Results

Scale free trees (with parameter $r > 0$)

- [Mori]

$$\boxed{\frac{D_n}{n^{\frac{1}{1+r}}} \longrightarrow \mu \quad (a.s.)}$$

and

$$\boxed{\frac{D_n - \mu n^{\frac{1}{1+r}}}{\sqrt{\mu n^{\frac{1}{1+r}}}} \longrightarrow N(0,1)}$$

for some random variable μ (related to degree distribtion).

Results

Scale free trees (with parameter $r > 0$)

- [D.]
  Suppose that $r = \frac{A}{B} > 0$ is rational. Then

$$\boxed{\mathbf{P}\{H_n \le k\} = F(c_k - d_r \log n) + o(1),}$$

where $c_k = k + O(\log k)$ and Set $d_r = 1/((r+1)s)$ with $r\,s\,e^{s+1} = 1$.

Further, $F(x) = \Psi(e^{-x})$, where $\Psi(y)$ is calculated by the following procedure.

Results

Let $\Phi(y)$ be the solution of

$$y^{\frac{1}{A+B}}\Phi(ye^{-1/d_r}) = \frac{\Gamma\left(1+\frac{1}{A+B}\right)}{\Gamma\left(\frac{1}{A+B}\right)^{A+B+1}} \times$$
$$\times \int_{y_1+\cdots+y_{A+B+1}=y,y_j\geq 0} \prod_{j=1}^{B+1}\left(\Phi(y_je^{-1/d_r})y_j^{\frac{1}{A+B}-1}\right)$$
$$\times \prod_{\ell=B+2}^{A+B+1}\left(\Phi(y_\ell)y_\ell^{\frac{1}{A+B}-1}\right)\, d\mathbf{y}$$

Then

$$\Psi(y) = \frac{\Gamma\left(\frac{A}{A+B}\right)}{\Gamma\left(\frac{1}{A+B}\right)^{A}} \int_{z_1+\cdots+z_A=1,z_j\geq 0} \prod_{j=1}^{A}\left(\Phi(yz_j)z_j^{\frac{1}{A+B}-1}\right)\, d\mathbf{z}$$

Proof Methods

Maximum degree in planted plane trees:

$t_{n,k}$ ... number of planted plane trees of size n with degrees $\le k$

$t_k(x) = \sum_{n\ge 1} t_{n,k}x^n$ ... generating function

$$\boxed{t_k(x) = x + xt_k(x) + xt_k(x^2) + \cdots + xt_k(x)^k = x\frac{1-t_k(x)^{k+1}}{1-t_k(x)}}$$

Proof Methods

Theorem

$y(x) = \sum_{n\ge 1} y_n x^n$ satisfies functional equation of the form

$$y(x) = x \cdot \varphi(y(x))$$

(+ some technical conditions)

$$\Longrightarrow \quad \boxed{y_n = \sqrt{\frac{\varphi(\tau)}{2\pi\varphi''(\tau)}}\frac{\rho^{-n}}{n^{3/2}}\left(1 + O(n^{-1})\right),}$$

where $\tau > 0$ is given by $\tau\varphi'(\tau) = \varphi(\tau)$ and $\rho = 1/\varphi'(\tau)$.

Proof Methods

Maximum degree in planted plane trees:

$$\varphi_k(u) = \frac{1-u^{k+1}}{1-u}, \qquad \varphi(u) = \frac{1}{1-u}$$

$$\begin{aligned}
\Longrightarrow \qquad & \tau_k = \frac{1}{2} - \frac{k}{2^{k+1}} + O(2^{-k}), \quad \tau = \frac{1}{2} \\
\Longrightarrow \qquad & \rho_k = \frac{1}{4}\left(1 + \frac{1}{2^{k+1}} + O(k^2 4^{-k})\right), \quad \rho = \frac{1}{4} \\
\Longrightarrow \qquad & t_{n,k} = \frac{1}{\sqrt{\pi}}\rho_k^{-n} n^{-3/2}\left(1 + O(k2^{-k}) + O(n^{-1})\right), \\
& t_n = \frac{1}{\sqrt{\pi}}\rho^{-n} n^{-3/2}\left(1 + O(n^{-1})\right) \\
\Longrightarrow \qquad & \mathbf{P}\{D_n \le k\} = \frac{t_{n,k}}{t_n} \sim \frac{\rho^n}{\rho_k^n} \sim e^{-n/2^{k+1}} = \exp\left(-2^{-(k-\log_2 n+1)}\right).
\end{aligned}$$

Proof Methods

Height of recursive trees:

$$y_n = (n-1)!, \qquad y(x) = \sum_{n\ge 1} y_n \frac{x^n}{n!} = \sum_{n\ge 1} \frac{x^n}{n} = \log \frac{1}{1-x}.$$

$$y'(x) = e^{y(x)}$$

$y_{n,k}$ ... number of recursive trees of size n with degrees $\le k$

$$= \mathbf{P}\{H_n \le k\} \cdot (n-1)!$$

$$y_k(x) = \sum_{n\ge 1} y_{n,k} \frac{x^n}{n!} = \sum_{n\ge 0} \mathbf{P}\{H_n \le k\} \frac{x^n}{n}$$

$$\boxed{y'_{k+1}(z) = e^{y_k(z)}}$$

Proof Methods

Alternate recurrence:

$$Y_k(x) = y_k'(x) = \sum_{n\geq 0} \mathbf{P}\{H_{n+1} \leq k\}\, x^n$$

$$\boxed{Y_{k+1}'(z) = Y_{k+1}(z)Y_k(z)}$$

$(Y_{k+1}(0) = 1)$

Proof Methods

Integral equation:

$$y\,\Psi(y/e^{1/e}) = \int_0^y \Psi(z/e^{1/e})\Psi(y-z)\,dz$$

$$L(u) = \int_0^\infty \Psi(y)e^{-yu}\,dy$$

$$\begin{aligned} \overline{Y}_k(x) &= e^{k/e}\cdot L\left(e^{k/e}(1-x)\right) \\ &= \int_0^\infty \Psi(ve^{-k/e})e^{-v}e^{xv}\,dv \end{aligned}$$

Proof Methods

Auxiliary functions:

$$\overline{Y}_k(x) = e^{k/e} \cdot L\left(e^{k/e}(1-x)\right)$$

- $1 - \overline{Y}_k(0) \sim Ck\left(\frac{2}{e}\right)^k, \qquad \overline{Y}_k(1) = e^{k/e}.$

- $$\boxed{\overline{Y}'_{k+1}(x) = \overline{Y}_{k+1}(x)\overline{Y}_k(x)}$$

- For all integers $\ell \geq 0$ and for all reals $k > 0$ the difference

  $$Y_\ell(x) - \overline{Y}_k(x)$$

  has exaclty one zero (**“Intersection Property”**)

Proof Methods

Auxiliary functions:

- $\overline{Y}_k(x) = \sum_{n\ge 0} \overline{Y}_{k,n} x^n$ is an entire function with coefficients

$$\overline{y}_{k,n} = \int_0^\infty \Psi\left(ve^{-k/e}\right) v^n e^{-v}\, dv$$

that are asymptotically given by

$$\boxed{\overline{Y}_{k,n} = \Psi\left(ne^{-k/e}\right) = F(k - e\log n) + o(1)}$$

$(\Psi(x) = F(e^{-x}))$

Proof Methods

Comparsion between $Y_k(x)$ and $\overline{Y}_k(x)$:

- $Y_k(x)$ is approximated by $\overline{Y}_{c_k}(x)$ by choosing c_k in a way that

$$Y_k(1) = \overline{Y}_{e_k}(1) \quad \Longleftrightarrow \quad c_k = e \cdot \log \tilde{Y}_k(1) \sim k.$$

- $Y_k(x) \approx \overline{Y}_{e_k}(x)$ in a neighbourhood of $x = 1$

$$\Longrightarrow \quad \boxed{\begin{aligned} \mathbf{P}\{H_n \le k\} \approx \overline{y}_{n,c_k} &= \Psi\Big(n/Y_k(1)\Big) + o(1) \\ &= F(c_k - e\log n) + o(1) \end{aligned}}$$

$(\Psi(x) = F(e^{-x}))$

Thank You!