RECURSIVE TREES

Michael Drmota

Institue of Discrete Mathematics and Geometry

Vienna University of Technology, A 1040 Wien, Austria

michael.drmota@tuwien.ac.at

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

2008 Enrage Topical School on GROWTH AND SHAPES,

Paris, IHP, June 2–6, 2008

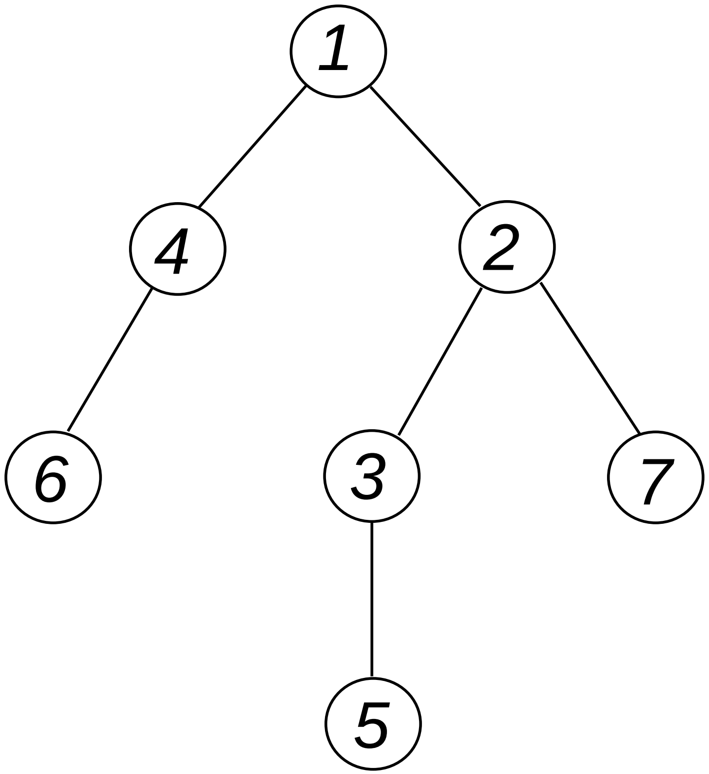
Contents

- Combinatorics on Recursive Trees
- The Shape or Random Recursive Trees
- Cutting down Recursive Trees
- Plane Oriented Recursive Trees

Methodology

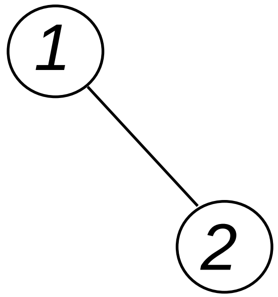
- Mixture of combinatorial, analytic and probabilistic methods
- Heavy use of generating function

Recursive Trees

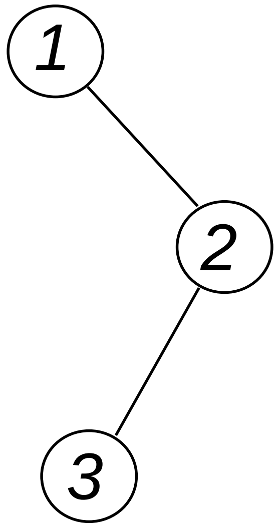


Recursive Trees

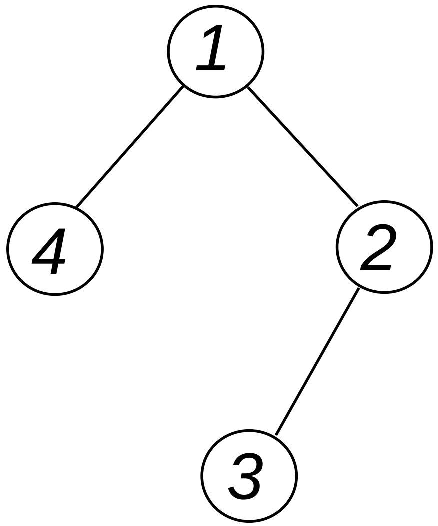
Recursive Trees



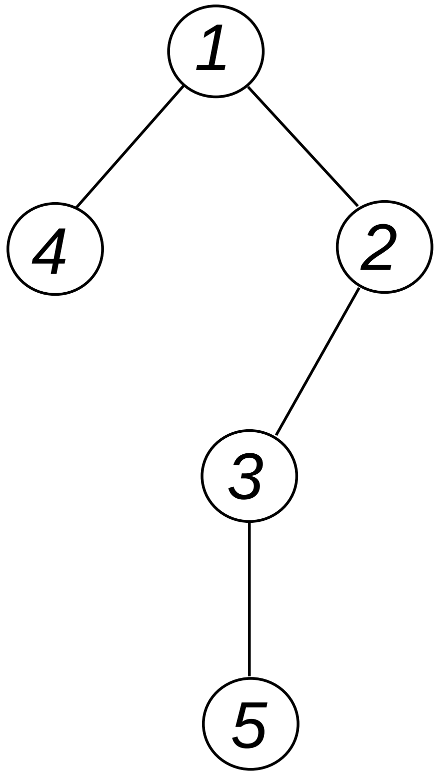
Recursive Trees



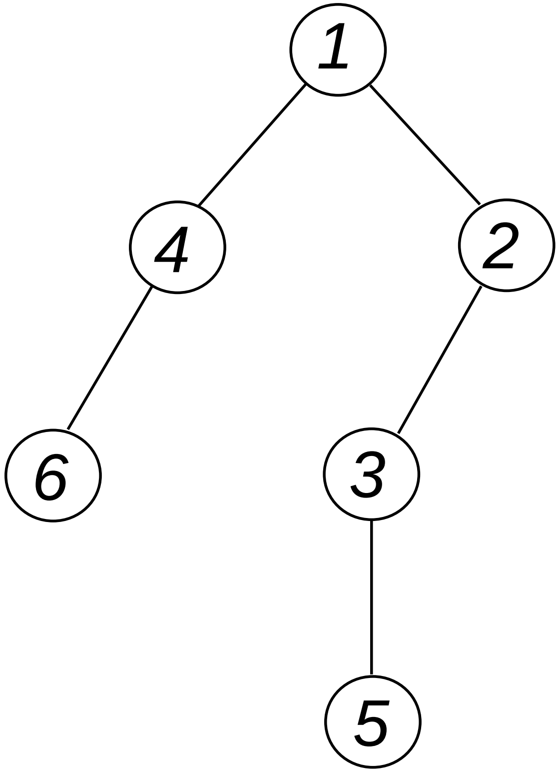
Recursive Trees



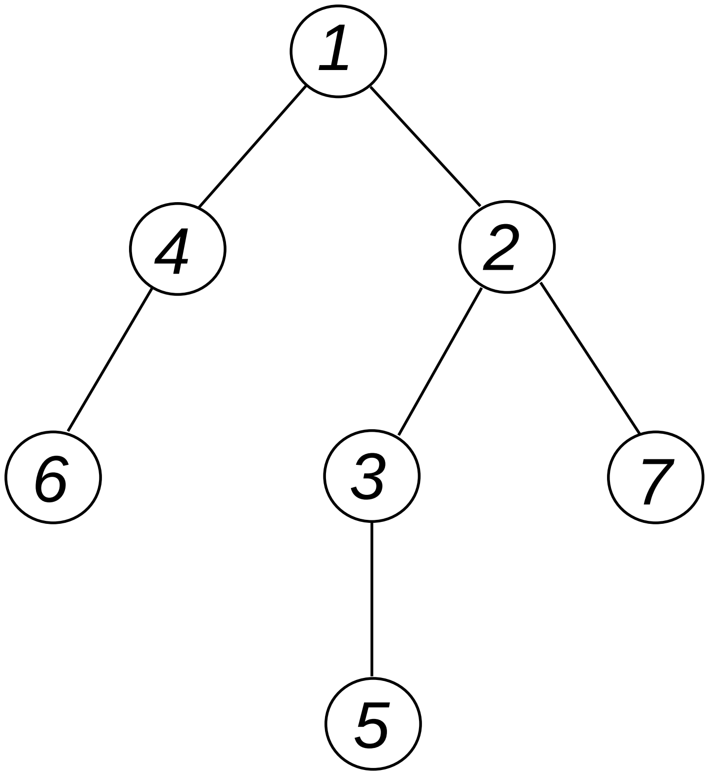
Recursive Trees



Recursive Trees



Recursive Trees



Recursive Trees

Combinatorial Description

- labelled rooted tree
- labels are strictly increasing
- no left-to-right order (non-planar)

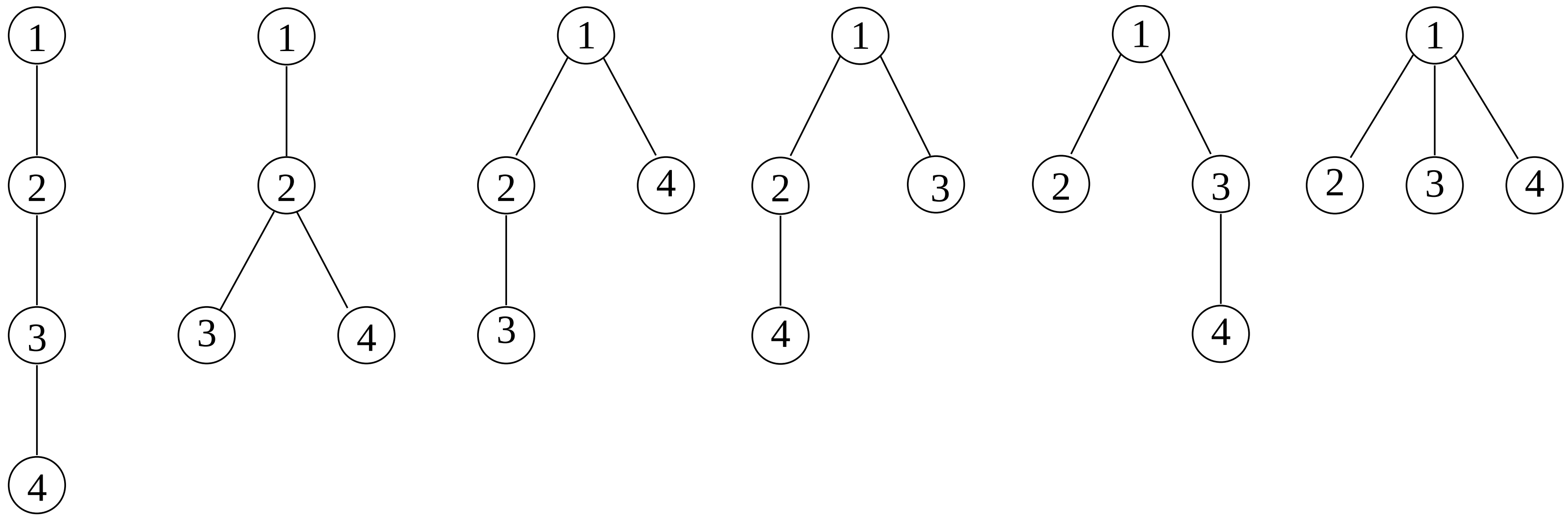
Recursive Trees

Motivations

- spread of epidemics
- pyramid schemes
- familiy trees of preserved copies of ancient texts
- convex hull algorithms
- ...

Enumeration of Recursive Trees

All recursive trees of size 4:



Enumeration of Recursive Trees

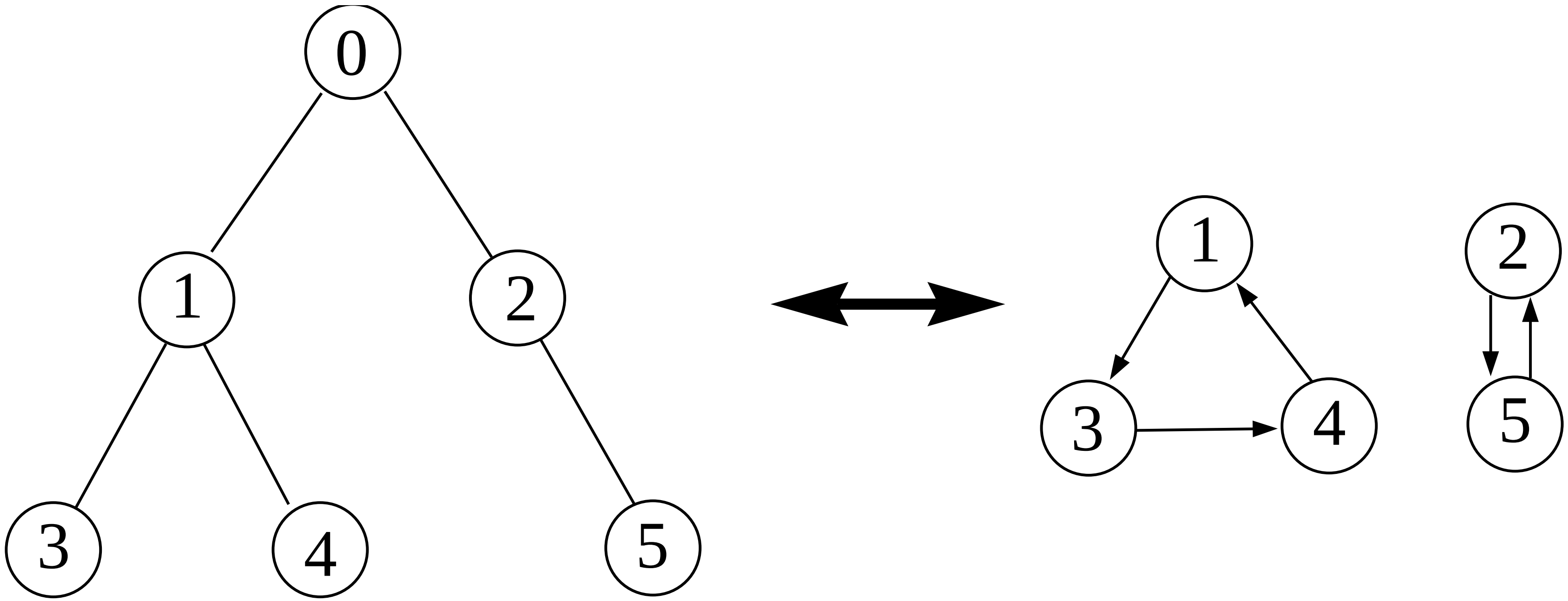
Number of recursive trees

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

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

Enumeration of Recursive Trees

Bijection to permutations



root degree = number of cycles

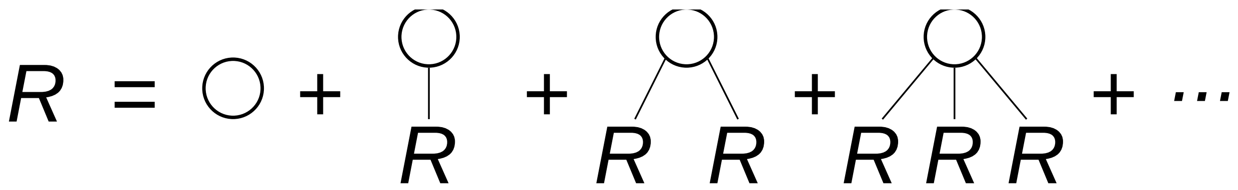
subtree sizes = cycle lengths

Enumeration of Recursive Trees

Generating Functions:

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

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



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

Random Recursive Trees

Probability Model:

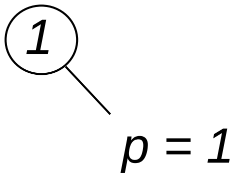
Process of growing trees:

- 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 $1/(j-1)$.

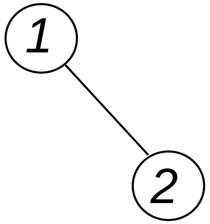
After n steps every tree (of size n) has equal probability $1/(n-1)!$.

Random Recursive Trees

Random Recursive Trees

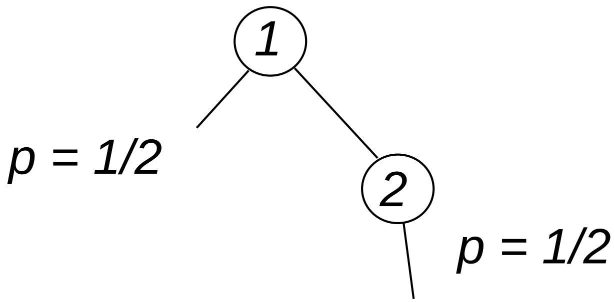


Random Recursive Trees

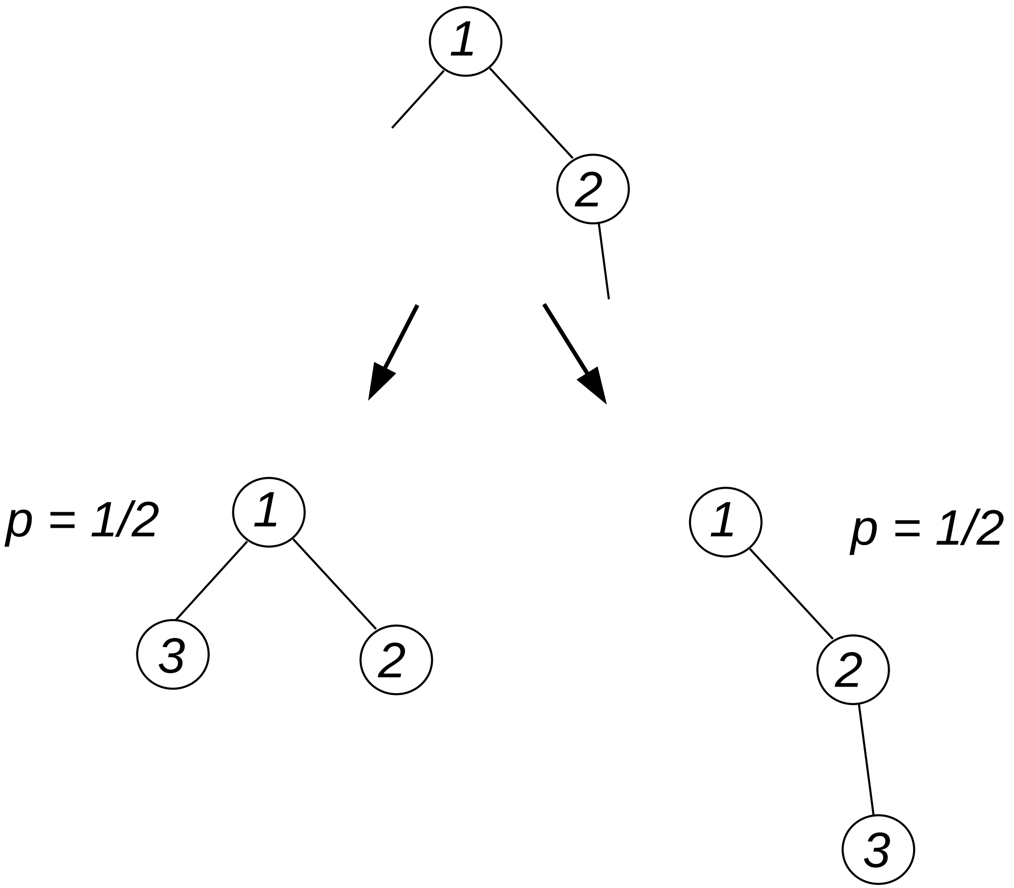


p = 1

Random Recursive Trees

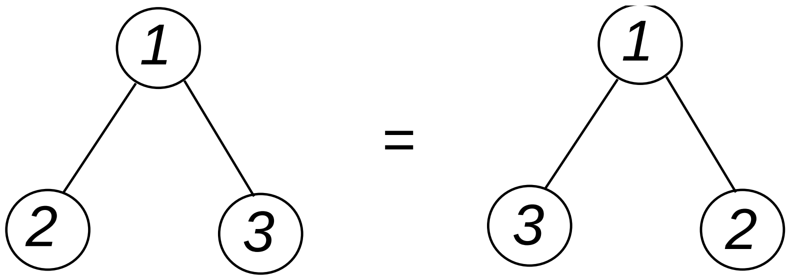


Random Recursive Trees



Random Recursive Trees

Remark: left-to-right order is irrelevant



Profile of Recursive Trees

First sample of shape parameters:

- **insertion depth** of the n-th node: D_n
- **path length**: I_n (sum of alle distances to the root)
- **height** H_n (maximal distance to the root)
- **degree distribution**
- **profile** $X_{n,k}$ (number of nodes at level k)

Profile of Recursive Trees

First sample of shape parameters:

- **insertion depth** of the n-th node: D_n
- **path length**: I_n (sum of alle distances to the root)
- **height** H_n (maximal distance to the root)
- **degree distribution**
- **PROFILE** $X_{n,k}$ (number of nodes at level k)

Profile of Recursive Trees

Relevance of the profile $X_{n,k}$:

- $\mathbb{P}\{D_n = k\} = \dfrac{1}{n-1}\,\mathbb{E}\,X_{n-1,k-1}$

- $I_n = \sum_{k \geq 0} k X_{n,k}$

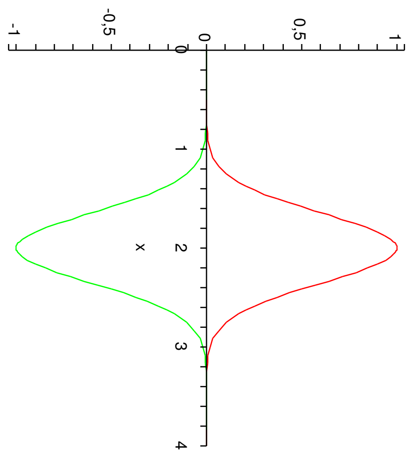
- $H_n = \max\{k \geq 0 : X_{n,k} > 0\}$

- The **profile** descibes the **shape** of the tree.

Profile of Recursive Trees

Average profile:

$$\mathbb{E}\,X_{n,k} = \frac{n}{\sqrt{2\pi \log n}} \left(e^{-\frac{(k-\log n)^2}{2\log n}} + \mathcal{O}\left(\frac{1}{\sqrt{\log n}} \right) \right).$$



Profile of Recursive Trees

Central limit theorem for the insertion depth [Devroye, Mahmoud]

$$\boxed{\frac{D_n - \log n}{\sqrt{\log n}} \to N(0,1)}$$

$$\mathbb{E}\, D_n = \log n + O(1), \qquad \mathbb{V}\, D_n = \log n + O(1).$$

Profile of Recursive Trees

Lemma [Dondajewski+Szymánski]

$$\mathbb{E}\, X_{n,k} = [u^k]\binom{n+u-1}{n-1} = \frac{|s_{n,k}|}{(n-1)!}$$

$s_{n,k}$ ... **Stirling numbers** of the first kind
$|s_{n,k}|$ ... number of permutations of $\{1,\ldots n\}$ with k cycles:

$$\sum_{k=0}^{n} s_{n,k}u^k = u(u-1)\cdots(u-n+1)$$
$$\sum_{k=0}^{n} |s_{n,k}|u^k = u(u+1)\cdots(u+n-1)$$

Stirling Numbers

$$s_{n+1,k} = s_{n,k-1} - n s_{n,k}, \qquad |s_{n+1,k}| = |s_{n,k-1}| + n|s_{n,k}|$$

$s_{n,k}$	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$
$n=0$	1								
$n=1$	0	1							
$n=2$	0	-1	1						
$n=3$	0	2	-3	1					
$n=4$	0	-6	11	-6	1				
$n=5$	0	24	-50	35	-10	1			
$n=6$	0	-120	274	-225	85	-15	1		
$n=7$	0	720	-1764	1624	-735	175	-21	1	
$n=8$	0	-5040	13068	-13132	6769	-1960	322	-28	1

Profile of Recursive Trees

Remark

$$\binom{n+\alpha-1}{n} = (-1)^n \binom{-\alpha}{n} = \frac{n^{\alpha-1}}{\Gamma(\alpha)} \left(1 + O\left(\frac{1}{n}\right)\right)$$

Corollary

$$|s_{n,k}| = \frac{(n-1)!(\log n)^k}{k!\Gamma(\frac{k}{\log n}+1)} \left(1 + O\left(\frac{1}{n}\right)\right)$$
$$\sim \frac{n!}{\sqrt{2\pi \log n}} e^{-\frac{(k-\log n)^2}{2\log n}}$$

Proof:

$$\binom{n+u-1}{n-1} \sim \frac{n^u}{\Gamma(u+1)} \quad \Longrightarrow \quad [u^k]\binom{n+u-1}{n-1} \sim \frac{(\log n)^k}{k!\Gamma(\frac{k}{\log n}+1)}.$$

Profile of Recursive Trees

Corollary

$$\mathbb{P}\{D_n = k\} = \frac{\mathbb{E}\, X_{n-1,k-1}}{n-1} \sim \frac{1}{\sqrt{2\pi \log n}} e^{-\frac{(k-\log n)^2}{2\log n}}$$

This implies the **central limit theorem** for D_n.

Cycles in Permutations

$|s_{n,k}|$ = number of permutations of $\{1, \ldots n\}$ with k cycles

Corollary

C_n ... random number of **cycles in permutations**

$$\boxed{\frac{C_n - \log n}{\sqrt{\log n}} \to N(0,1)}$$

Corollary

R_n ... **root degree** of random recursive trees

$$\boxed{\frac{R_n - \log n}{\sqrt{\log n}} \to N(0,1)}$$

Profile of Recursive Trees

Profile polynomial

$$\boxed{W_n(u) = \sum_{k \geq 0} X_{n,k}\, u^k}$$

Lemma. The normalized profile polynomial

$$M_n(u) = \frac{W_n(u)}{\mathbb{E}\, W_n(u)}$$

is a **martingale** (with respect to the natural filtration related to the tree evolution process).

Profile of Recursive Trees

Theorem [Chauvin+Drmota+Jabbour for binary search trees]

$$\left(\frac{W_n(u)}{\mathbb{E}\,W_n(u)}, u \in B\right) \to (M(u), u \in B)$$

for a suitable domain $B \subseteq \mathbb{C}$.

Remarks

- $(M(u), u \in B)$ stochastic process of **random analytic functions**.

- Fixed point equation:

$$\boxed{M(u) \equiv uU^u M^{(1)}(u) + (1-U)^u M^{(2)}(u)},$$

  where $M^{(1)}(u)$ and $M^{(2)}(u)$ are independent copies of $M(u)$, U is uniform in $[0,1]$ and $(U, M^{(1)}(u), M^{(2)}(u))$ are independent.

Profile of Recursive Trees

Theorem [Chauvin+Drmota+Jabbour for binary search trees]

$$\left(\frac{X_{n,\lfloor \alpha \log n \rfloor}}{\mathbb{E}\, X_{n,\lfloor \alpha \log n \rfloor}}, \alpha \in I\right) \to (M(\alpha), \alpha \in I)\,.$$

Idea

$$\begin{aligned} X_{n,k} &= [u^k] W_n(u) \\ &= [u^k] M_n(u) \cdot \mathbb{E}\, W_n(u) \\ &\sim [u^k] M(u) \cdot \mathbb{E}\, W_n(u) \cdot \\ &\sim M(\alpha)[u^k] \mathbb{E}\, W_n(u) = M(\alpha) \mathbb{E}\, X_{n,k}. \end{aligned}$$

$\alpha = k/\log n$ ... saddle point the function $n^u u^{-k}$.

Path Length

Remark

$$M_n'(1) = \frac{I_n - \mathbb{E}\, I_n}{n}$$

Corollary

$$\boxed{\frac{I_n - \mathbb{E}\, I_n}{n} \longrightarrow M'(1)}$$

The random variable $M'(1)$ is not normal. Note also that $\mathbb{E}\, I_n \sim n \log n$.

Leaves in Recursive Trees

Theorem [Najock+Heyde]

L_n ... number of leaves in a random recursive tree of size n

$$\mathbb{P}\{L_n = k\} = \frac{1}{(n-1)!} \left\langle {n-1 \atop k-1} \right\rangle .$$

$\left\langle {n \atop k} \right\rangle$ ... Eulerian numbers

Remark

$\left\langle {n-1 \atop k-1} \right\rangle =$ number or recursive trees of size n with k leaves.

Eulerian Numbers

$$\left\langle {n \atop k} \right\rangle = k \left\langle {n-1 \atop k} \right\rangle + (n-k) \left\langle {n-1 \atop k-1} \right\rangle$$

$\left\langle {n \atop k} \right\rangle$	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$
$n=0$	1								
$n=1$	1	1							
$n=2$	1	4	1						
$n=3$	1	11	11	1					
$n=4$	1	26	66	26	1				
$n=5$	1	57	302	302	57	1			
$n=6$	1	120	1191	4216	1191	120	1		
$n=7$	1	247	4293	15619	15619	4293	247	1	
$n=8$	1	502	14608	88234	156190	88234	14608	502	1

Leaves in Recursive Trees

Corollary

$$\boxed{\frac{L_n - \frac{n}{2}}{\sqrt{\frac{7}{12}n}} \to N(0,1)}$$

$$\mathbb{E}\, L_n = \frac{n}{2}, \qquad \mathbb{E}\,(L_n)^2 = \frac{1}{12}(3(n-1)^2 + 13(n-1) + 14),$$

Leaves in Recursive Trees

Generating functions

$\ell_{n,k}$ ... number or recursive trees of size n with k leaves.

$$y(x,u) = \sum_{n,k} \ell_{n,k} \cdot u^k \cdot \frac{x^n}{n!} = \sum_{n,k} \mathbb{P}\{L_n = k\} \cdot u^k \cdot \frac{x^n}{n}$$

$$\frac{\partial y(x,u)}{\partial x} = u + e^{y(x,u)} - 1$$

$$\boxed{y(x,u) = (x-1)(u-1) + \log\left(\frac{u-1}{1-e^{(x-1)(u-1)}}\right)}$$

Degree Distribution

$\ell_{n,k}^{(d)}$ ... number or r.t.'s of size n with k nodes of outdegree d.

$L_n^{(d)}$ ... number of nodes of outdegree d in a random r.t. of size n:

$$\mathbb{P}\{L_n^{(d)} = k\} = \frac{\ell_{n,k}^{(d)}}{(n-1)!}$$

$\overline{D}_n$ ... **degree of a random node** in a random r.t. of size n

$$\mathbb{P}\{\overline{D}_n = d\} = \frac{1}{n}\,\mathbb{E}\,L_n^{(d)}$$

Theorem

$$\boxed{\mathbb{P}\{\overline{D}_n = d\} = \frac{1}{2^{d+1}} + O\left(\frac{1}{n^2}\frac{(2\log n)^d}{d!}\right)}$$

Degree Distribution

Generating functions

$$y(x,u) = \sum_{n,k} \ell_{n,k}^{(d)} \cdot u^k \cdot \frac{x^n}{n!} = \sum_{n,k} \mathbb{P}\{L_n^{(d)} = k\} \cdot u^k \cdot \frac{x^n}{n}$$

$$\boxed{\frac{\partial y(x,u)}{\partial x} = e^{y(x,u)} + (u-1)\frac{y(x,u)^d}{d!}}$$

$$Y(x) = \left.\frac{\partial y(x,u)}{\partial u}\right|_{u=1} = \sum_{n\ge 0} \mathbb{E}\, L_n^{(d)} \cdot \frac{x^n}{n} = \sum_{n\ge 0} \mathbb{P}\{\overline{D}_n = d\} \cdot x^n$$

$$\boxed{Y'(x) = \frac{1}{1-x}Y(x) + \frac{1}{d!}\left(\log\frac{1}{1-x}\right)^d}$$

Degree Distribution

$$Y'(x) = \frac{1}{1-x} Y(x) + \frac{1}{d!}\left(\log \frac{1}{1-x}\right)^d$$

$$\Longrightarrow \quad Y(x) = \frac{1}{2^{d+1}} \frac{1}{1-x} + (x-1)\sum_{j=0}^{d} \frac{1}{j!\,2^{d+1-j}}\left(\log \frac{1}{1-x}\right)^j$$

$$\Longrightarrow \quad \mathbb{P}\{\overline{D}_n = d\} = \frac{1}{2^{d+1}} + O\left(\frac{1}{n^2}\frac{(2\log n)^d}{d!}\right)$$

uniformly for $d \le (2-\varepsilon)\log n$.

Corollary

$$\mathbb{P}\{\overline{D}_n > d\} = \frac{1}{2^{d+1}} + O\left(\frac{1}{n^2}\frac{(2\log n)^d}{d!}\right)$$

Maximum Degree

Theorem [Szymánski, Pittel]

Δ_n ... **maximum node degree** in random r.t.'s.

$$\boxed{\mathbb{E}\,\Delta_n \sim \log_2 n}$$

Remark. This degree is much larger than the expected root degree with is about $\log n$.

First Moment Method

X ... discrete random variable on **non-negative integers**.

$$\Longrightarrow \quad \boxed{\mathbb{P}\{X > 0\} \leq \min\{1, \mathbb{E}\, X\}}.$$

Proof

$$\mathbb{E}\, X = \sum_{k \geq 0} k\, \mathbb{P}\{X = k\} \geq \sum_{k \geq 1} \mathbb{P}\{X = k\} = \mathbb{P}\{X > 0\}.$$

Maximum Degree

Upper bound: first moment method

X_d ... number of nodes of degree $> d$:

$$\boxed{\mathbb{E}\, X_d = n\mathbb{P}\{\overline{D}_n > d\}} \qquad \boxed{\Delta_n > d \iff X_d > 0}$$

$$\begin{aligned}
\Longrightarrow \qquad \mathbb{E}\,\Delta_n &= \sum_{d\ge 0} \mathbb{P}\{\Delta_n > d\} \\
&= \sum_{d\ge 0} \mathbb{P}\{X_d > 0\} \\
&\le \sum_{d\ge 0} \min\{1, \mathbb{E}\, X_d\} \\
&\le \sum_{d\le \log_2 n} 1 + n \sum_{d>\log_2 n} \mathbb{P}\{\overline{D}_n > d\} \\
&= \log_2 n + O(1).
\end{aligned}$$

Second Moment Method

Theorem

X ... non-negative random variable with bounded second moment

$$\implies \boxed{\mathbb{P}\{X > 0\} \geq \frac{(\mathbb{E}\, X)^2}{\mathbb{E}\,(X^2)}}.$$

Proof

$$\mathbb{E}\, X = \mathbb{E}\left(X \cdot \mathbf{1}_{[X>0]}\right) \leq \sqrt{\mathbb{E}\,(X^2)}\sqrt{\mathbb{E}\,(\mathbf{1}_{[X>0]}^2)} = \sqrt{\mathbb{E}\,(X^2)}\sqrt{\mathbb{P}\{X > 0\}}$$

Remark In order to apply the second moment method to obtain a lower bound for $\mathbb{E}\,\Delta_n$ one needs estimates for $\mathbb{E}\,(X_d)^2$ which can be derived in a similar fashion as above.

Maximum Degree

Theorem [Goh+Schmutz]

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

Remark. The limiting behaviour of Δ_n is related the the exteme value (= Gumbel) distribution ($F(t) = e^{-e^{-t}}$).

The distribution of Δ_n is extremely concentrated around $d \approx \log_2 n$.

The proof is an analytic "tour de force".

Height of Recursive Trees

Height H_n

Theorem [Devroye 1987, Pittel 1994]

$$\boxed{\frac{H_n}{\log n} \longrightarrow e \qquad (a.s.)}$$

Height Distribution

$F(z)$ solution of

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

Recursive sequence of generating functions:

$$\boxed{y'_{k+1}(x) = e^{y_k(x)}},\ \ y_0(x) = 0,\ \ y_k(0) = 0.$$

Theorem [Drmota]

$$\mathbb{E}\,H_n = e\log n + O\Big(\sqrt{\log n}\,(\log\log n)\Big).$$

$$\boxed{\mathbf{P}\{H_n \le k\} = F(n/y'_k(1)) + o(1)}$$

$$\mathbf{P}\{|H_n - \mathbb{E}\,H_n| \ge \eta\} \ll e^{-c\eta} \quad (c > 0)$$

Height Distribution

$y_{n,k}$ ... number of r.t.'s of size n and height $\le k$:

$\mathbb{P}\{H_n \le k\} = y_{n,k}/(n-1)!$

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

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

$$Y_k(x) = y'_k(x) = \sum_{n\ge 0} \mathbb{P}\{H_{n+1} \le k\} x^n$$

$$\boxed{Y'_{k+1}(x) = Y_{k+1}(x) Y_k(x)}$$

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

Height Distribution

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

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

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

Height Distribution

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

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

- For every positive integer ℓ and for every real number $k > 0$ the difference

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

has exactly one zero (**"Intersection Property"**).

Height Distribution

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

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

and asymptotically we have

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

Height Distribution

Remark:

The functions

$$\overline{y}_k(x) = \int_0^x \overline{Y}_k(t)\,dt = \log \overline{Y}_{k+1}(x)$$

satisfy the recurrence

$$\boxed{\overline{y}_{k+1}(x) = e^{\overline{y}_k(x)}}$$

Height Distribution

Proof idea

- $Y_k(x)$ is approximated by the *auxiliary function* $\overline{Y}_{e_k}(x)$:

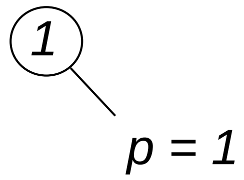
$$Y_k(1) = \overline{Y}_{e_k}(1) \quad \Longleftrightarrow \quad e_k = e \cdot \log Y_k(\rho) \sim k.$$

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

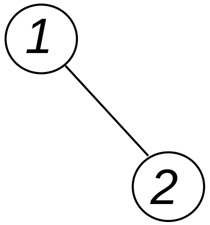
$$\Longrightarrow \quad \boxed{\mathbf{P}\{H_n \le k\} \approx \overline{Y}_{n,e_k} = F\Big(n/Y_k(1)\Big) + o(1)}$$

Plane Oriented Trees

Plane Oriented Trees

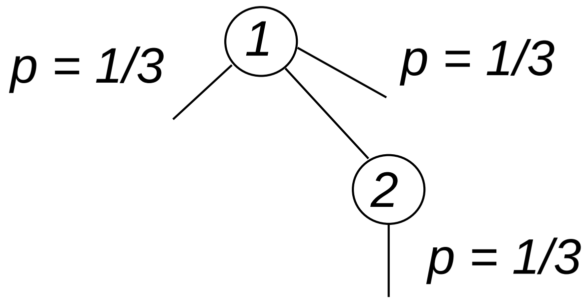


Plane Oriented Trees

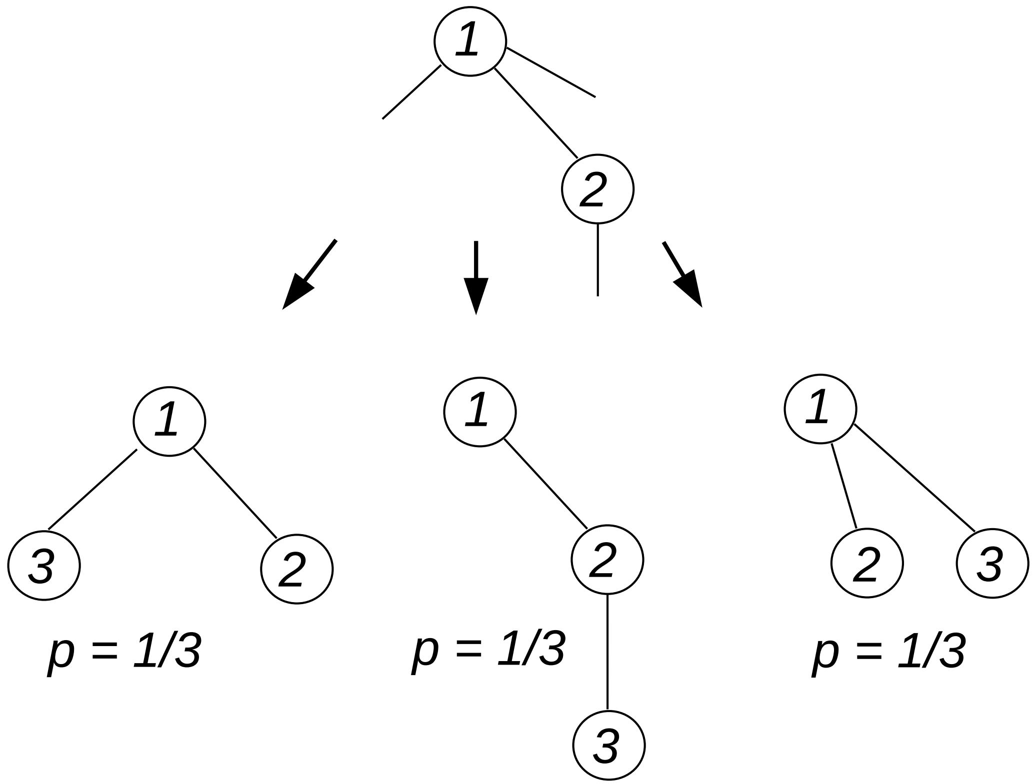


$p = 1$

Plane Oriented Trees



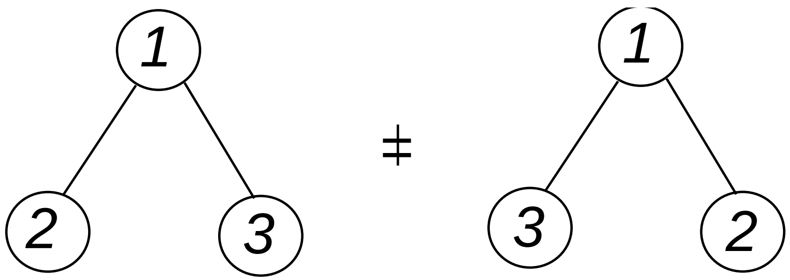
Plane Oriented Trees



p = 1/3 p = 1/3 p = 1/3

Plane Oriented Trees

Remark: left-to-right order is relevant



Plane Oriented Trees

Number of Plane Oriented Trees:

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

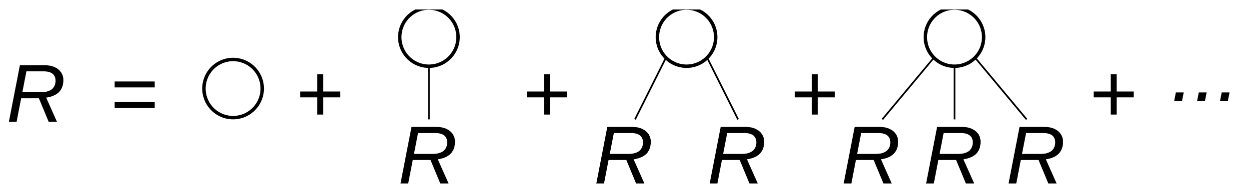
The node with label j has exactly $2j-3$ possibilities to be inserted $\Longrightarrow y_n = 1 \cdot 3 \cdots (2n-3)$.

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 $1/(2n-3)!!$.

Plane Oriented Trees

Depth D_n of the n-th node

$$\mathbb{E}\, D_n = H_{2n-1} - \frac{1}{2} H_{n-1} = \frac{1}{2} \log n + O(1)$$

$$\begin{aligned} \mathbb{V}\, D_n &= H_{2n-1} - \frac{1}{2} H_{n-1} - H^{(2)}_{2n-1} + \frac{1}{4} H^{(2)}_{n-1} \\ &= \frac{1}{2} \log n + O(1) \end{aligned}$$

Central limit theorem:

$$\boxed{\frac{D_n - \frac{1}{2}\log n}{\sqrt{\frac{1}{2}\log n}} \to N(0,1)}$$

Plane Oriented Trees

Number L_n of leaves

$$\mathbb{E}\, L_n = \frac{2n-1}{3}$$

$$\mathbb{V}\, L_n = \frac{n}{9} - \frac{1}{18} - \frac{1}{6(2n-1)}$$

Central limit theorem:

$$\boxed{\frac{L_n - \frac{2}{3}n}{\sqrt{\frac{n}{9}}} \to N(0,1)}$$

Plane Oriented Trees

Distribution of out-degrees

$\overline{D}_n$ ... **degree of a random node** in a random p.o.r.t. of size n

$$\mathbb{P}\{\overline{D}_n = d\} = \frac{4}{(d+1)(d+2)(d+3)} + o(1)$$

Remark. $\frac{4}{(d+1)(d+2)(d+3)} \sim 4\, d^{-3}$ as $d \to \infty$.

Plane Oriented Trees

Root degree R_n

$$\mathbb{P}\{R_n = k\} = \frac{(2n-3-k)!}{2^{n-1-k}(n-1-k)!} \sim \sqrt{\frac{2}{\pi n}} e^{-k^2/(4n)}$$

$$\mathbb{E}\, R_n = \sqrt{\pi n} + O(1)$$

Plane Oriented Trees

Height H_n

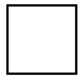
[Pittel 1994]

$$\boxed{\frac{H_n}{\log n} \to \frac{1}{2s} = 1.79556\ldots \qquad (a.s.)}$$

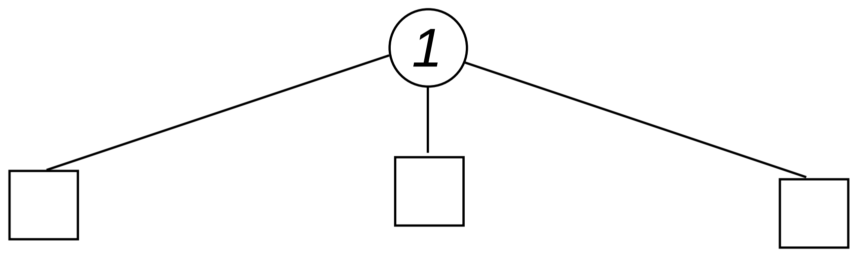
where $s = 0.27846\ldots$ is the positive solution of $se^{s+1} = 1$.

Precise results (as above) are also available ([Drmota]).

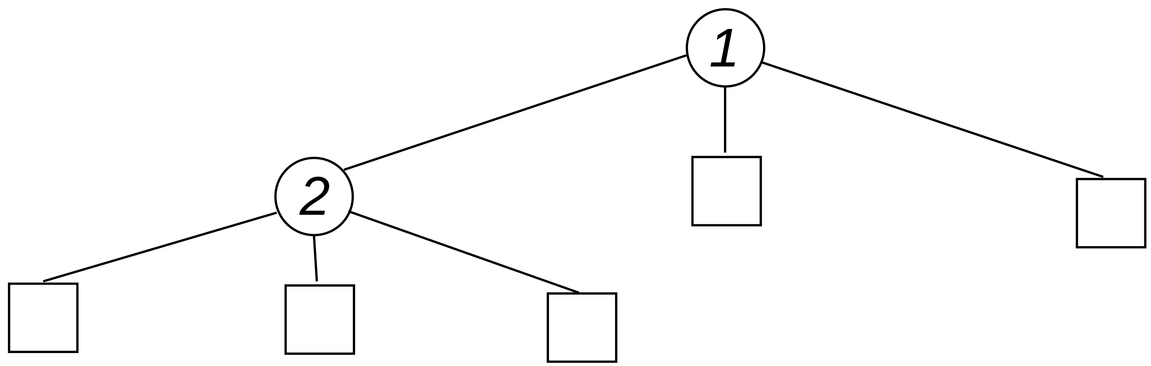
D-ary Recursive Trees



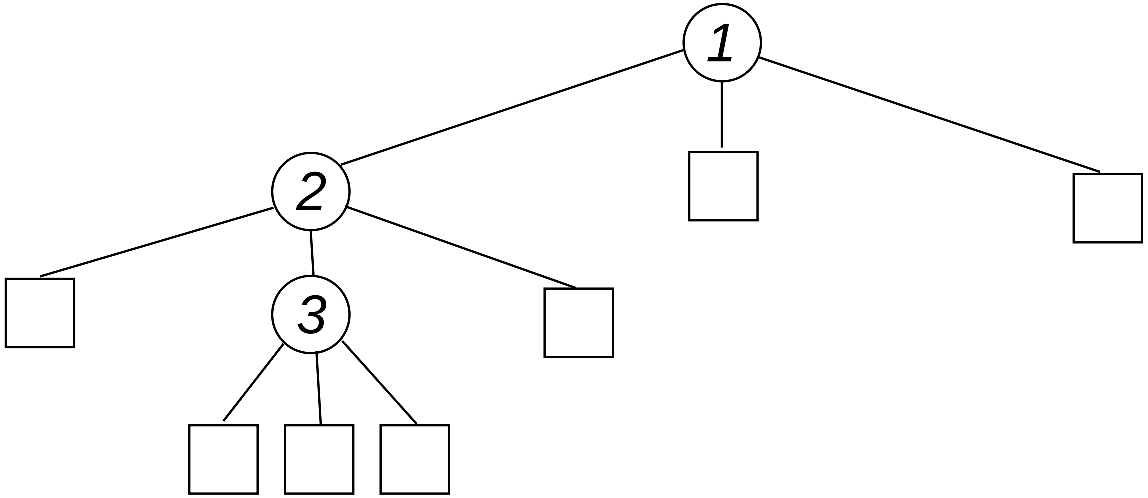
D-ary Recursive Trees



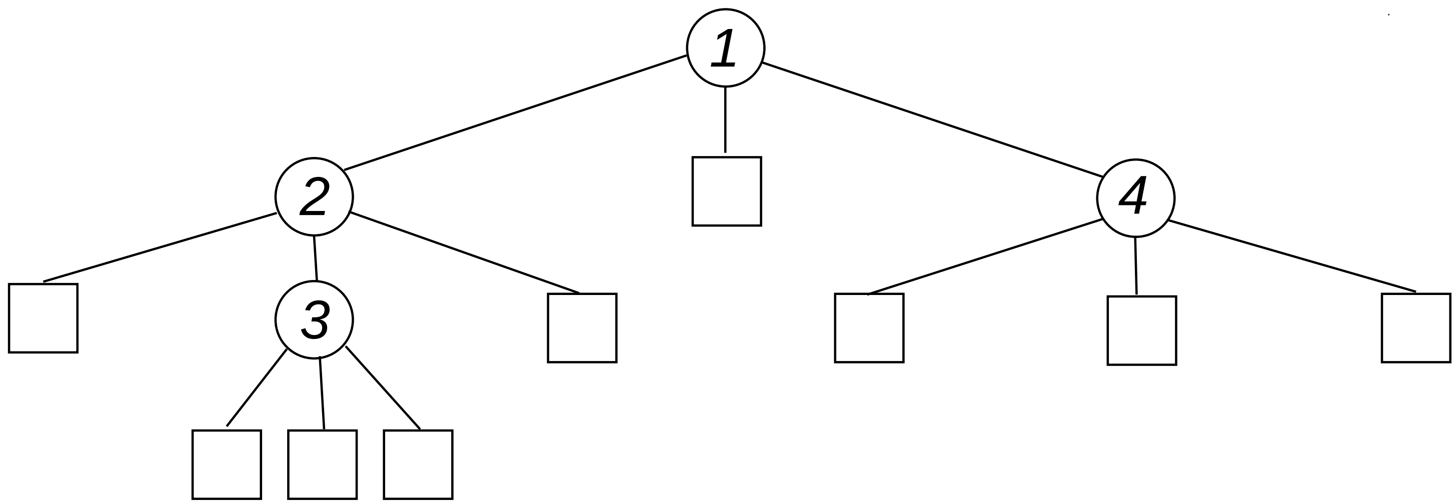
D-ary Recursive Trees



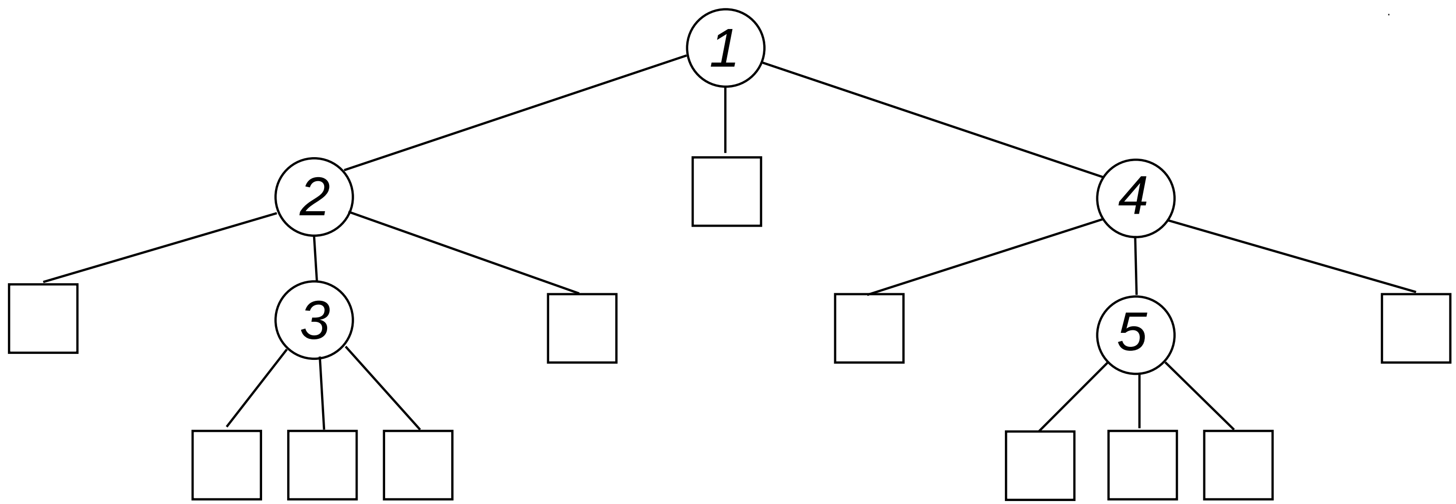
D-ary Recursive Trees



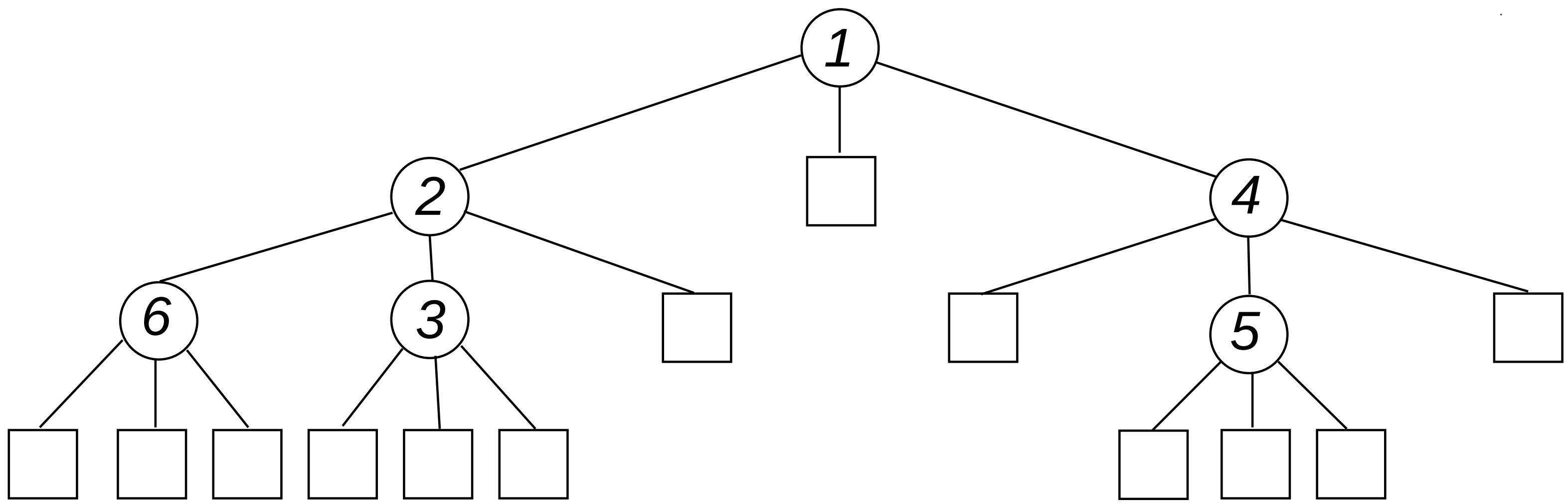
D-ary Recursive Trees



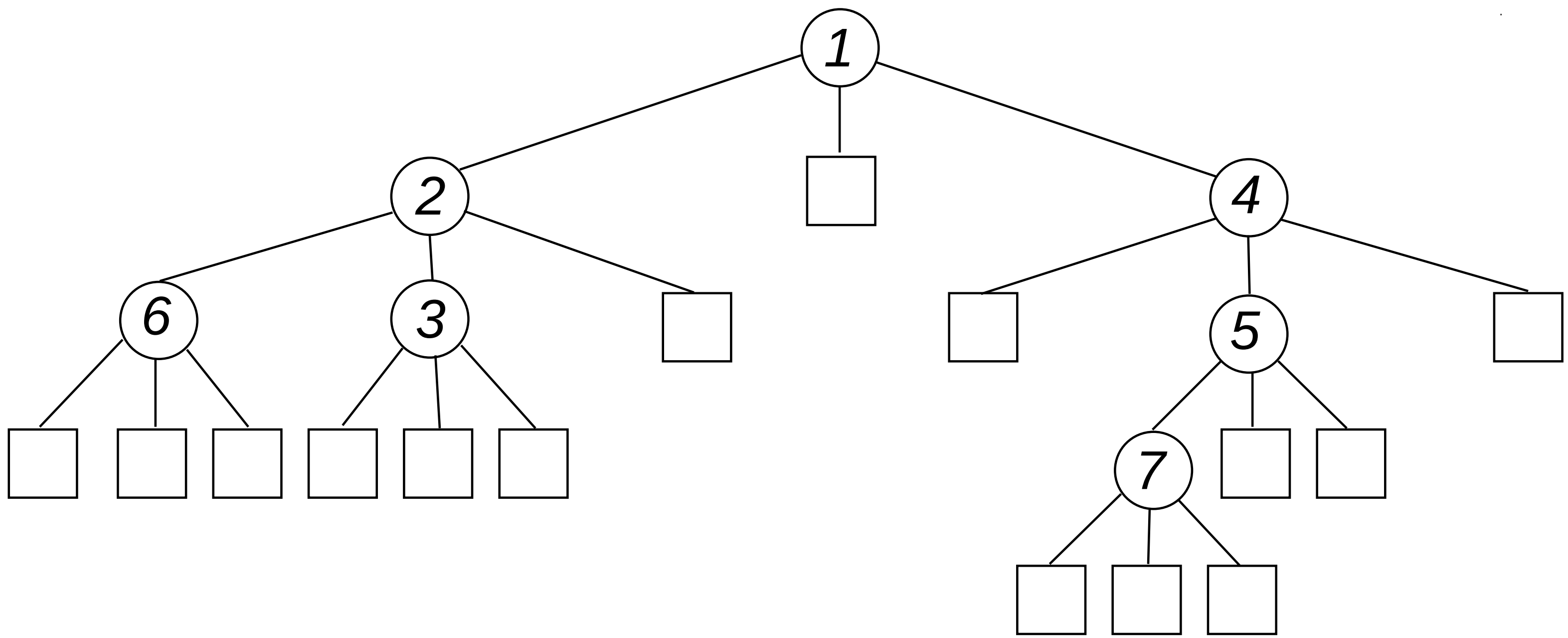
D-ary Recursive Trees



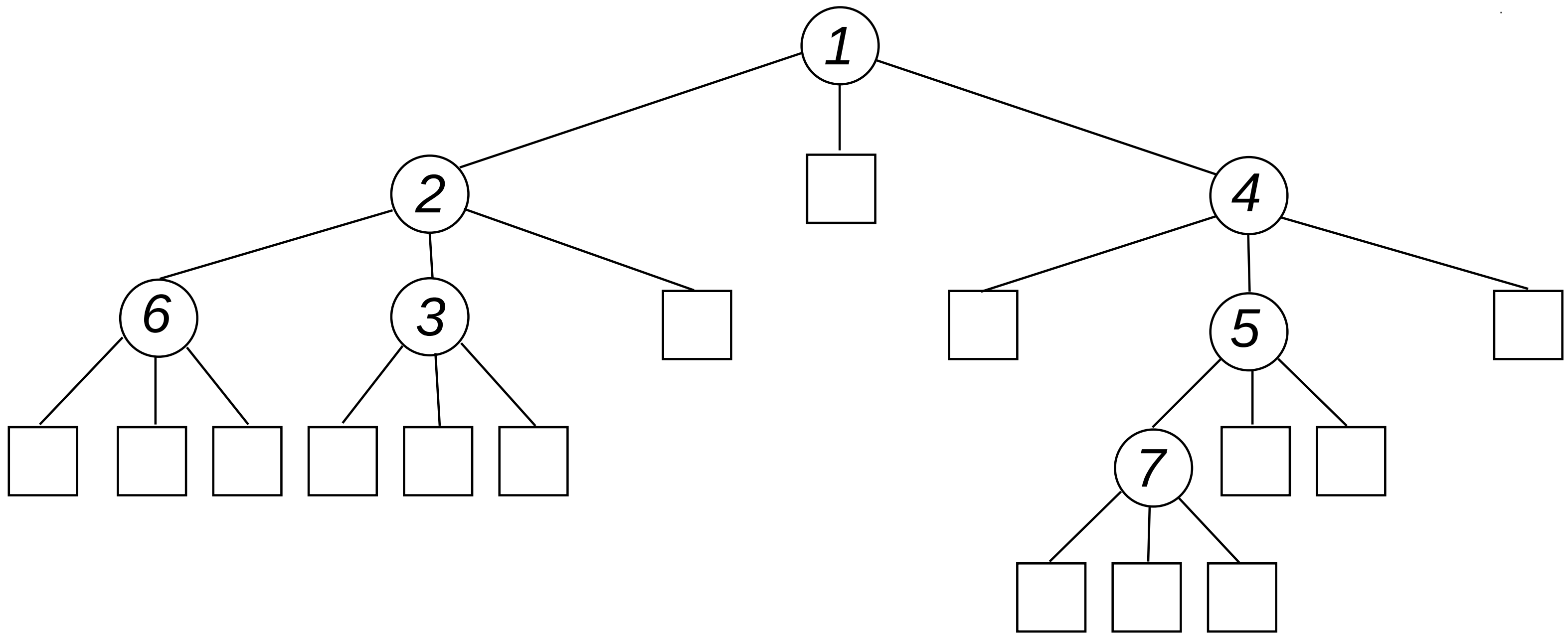
D-ary Recursive Trees

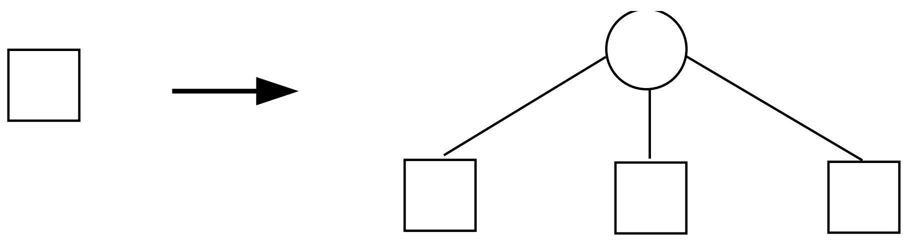


D-ary Recursive Trees

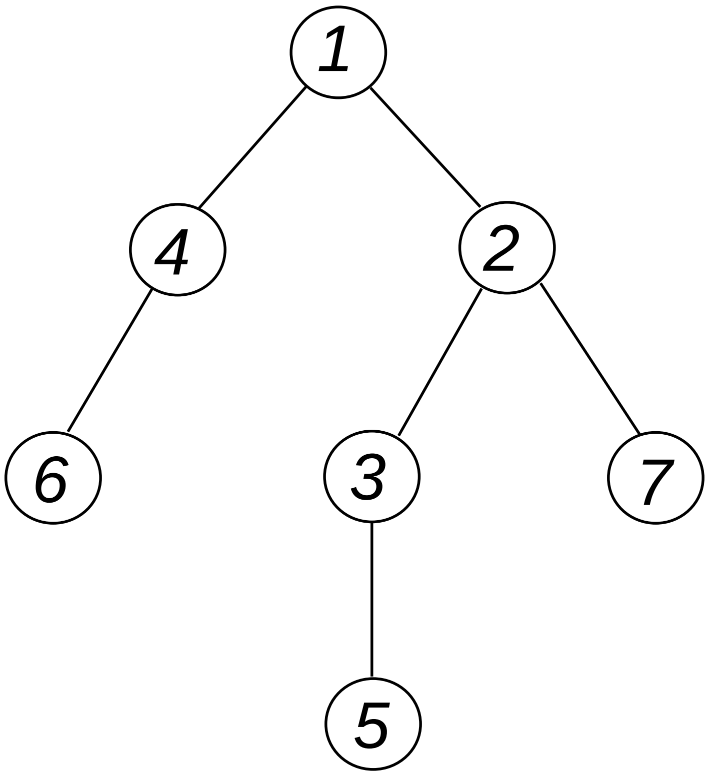


D-ary Recursive Trees

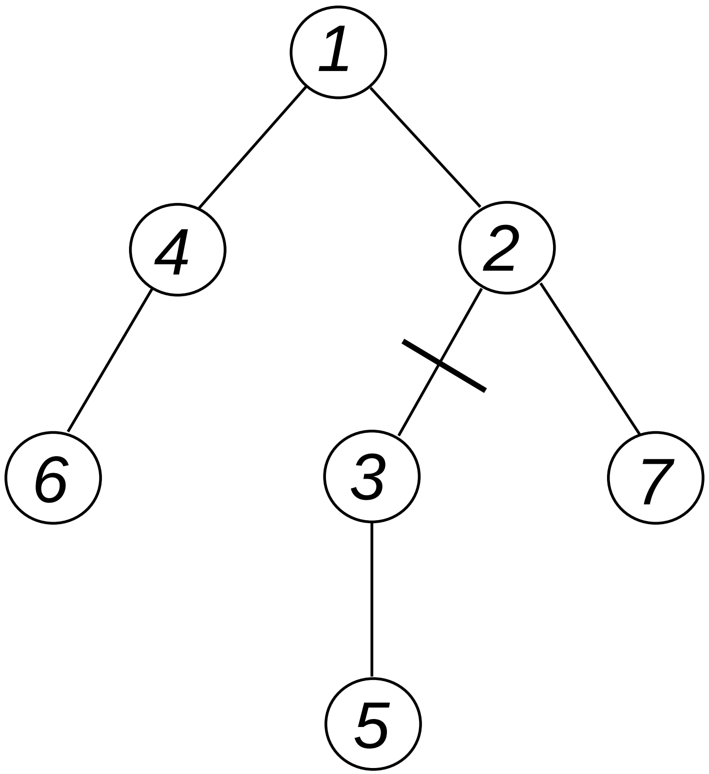




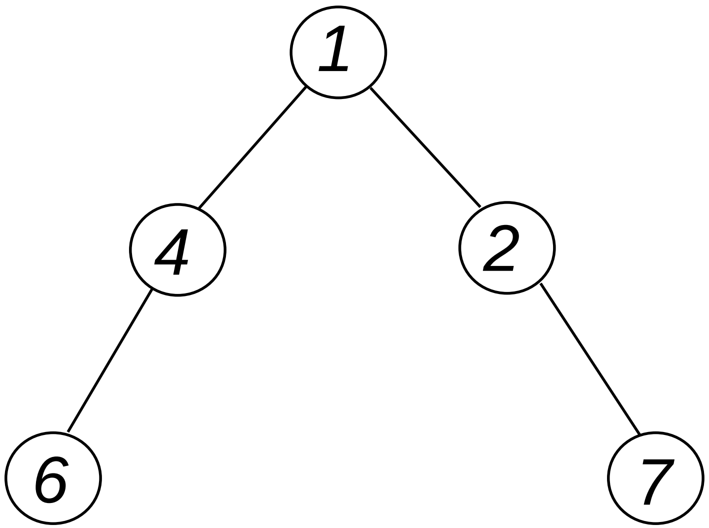
Cutting down Recursive Trees



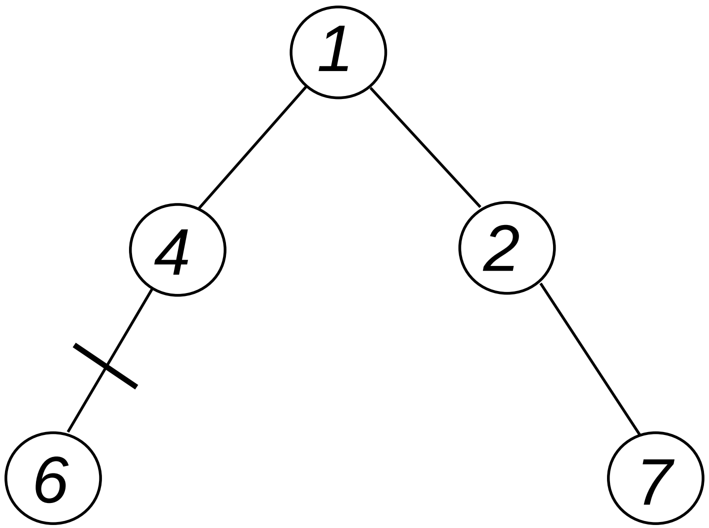
Cutting down Recursive Trees



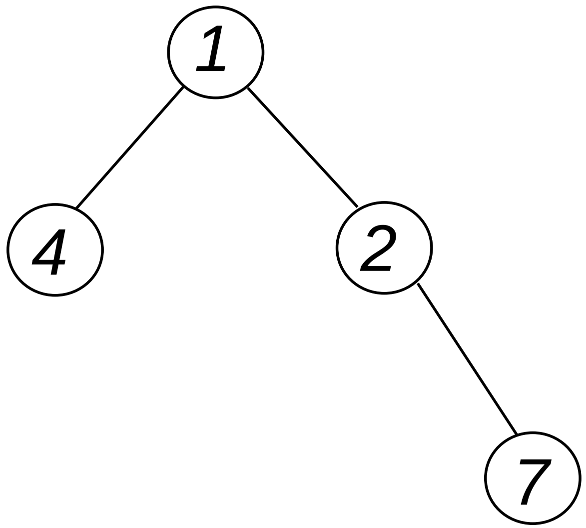
Cutting down Recursive Trees



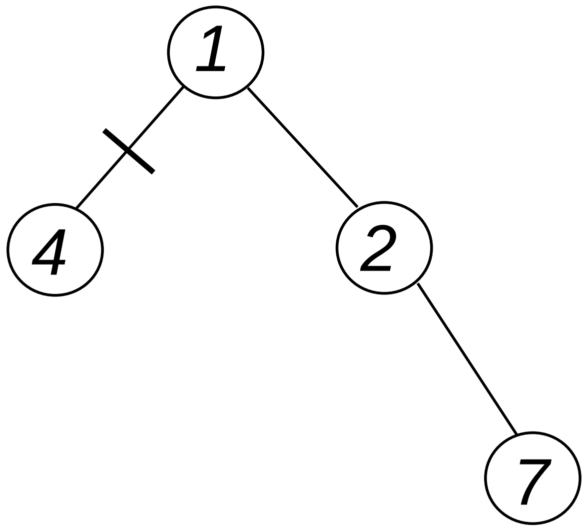
Cutting down Recursive Trees



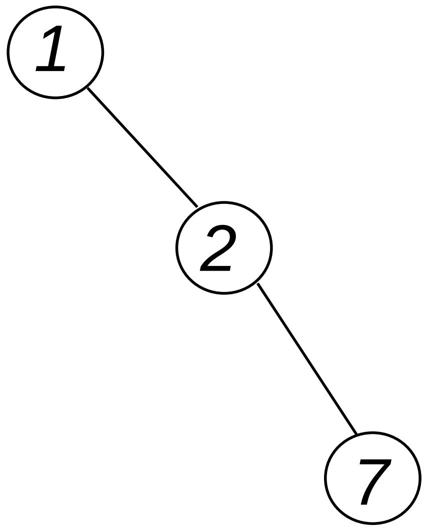
Cutting down Recursive Trees



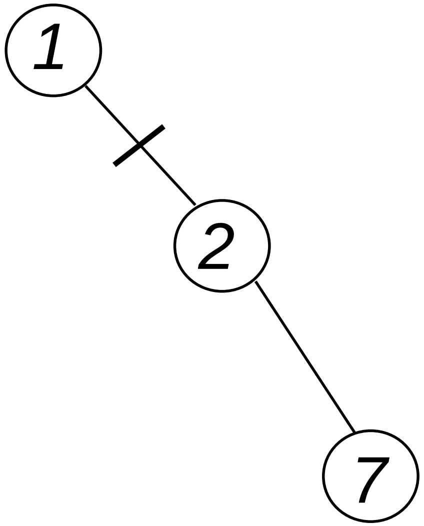
Cutting down Recursive Trees



Cutting down Recursive Trees



Cutting down Recursive Trees



Cutting down Recursive Trees

Cutting down Recursive Trees

X_n ... number of random cuts to cut down a random r.t. of size n.

$X_0 = X_1 = 0$,

$$\boxed{X_n \equiv X_{I_n} + 1} \qquad (n \ge 2),$$

where I_n is a discrete random variable with

$$\mathbb{P}\{I_n = k\} = \frac{1}{(n-k)(n-k+1)} \frac{n}{n-1} \qquad (0 \le k < n)$$

that is independent of $(X_0, X_1, \ldots, X_{n-1})$.

Cutting down Recursive Trees

Lemma

The probability to that the remaining tree has size $= k$ if we cut a random edge in a random recursive tree of size n equals

$$\frac{1}{(n-k)(n-k+1)}\,\frac{n}{n-1}$$

Proof

$$\begin{aligned}\frac{(k-1)!(n-k-1)!}{(n-1)(n-1)!}\sum_{j=1}^{k}\binom{n-j}{n-k} &= \frac{(k-1)!(n-k-1)!}{(n-1)(n-1)!}\binom{n}{n-k+1} \\ &= \frac{1}{(n-k)(n-k+1)}\,\frac{n}{n-1}\end{aligned}$$

Cutting down Recursive Trees

Theorem [Drmota+Iksander+Möhle+Rösler]

$$\boxed{\frac{X_n - \frac{n}{\log n} - \frac{n \log\log n}{(\log n)^2}}{\frac{n}{(\log n)^2}} \to Y},$$

where Y is a stable random variable with characteristic function

$$\mathbb{E}\, e^{i\lambda Y} = e^{i\lambda \log|\lambda| - \frac{\pi}{2}|\lambda|}.$$

$$\mathbb{E}\, X_n = \frac{n}{\log n} + O\left(\frac{n}{\log^2 n}\right), \quad \mathbb{V}\, X_n = \frac{n^2}{2\log^3 n} + O\left(\frac{n^2}{\log^4 n}\right)$$

Cutting down Recursive Trees

Stable distributions

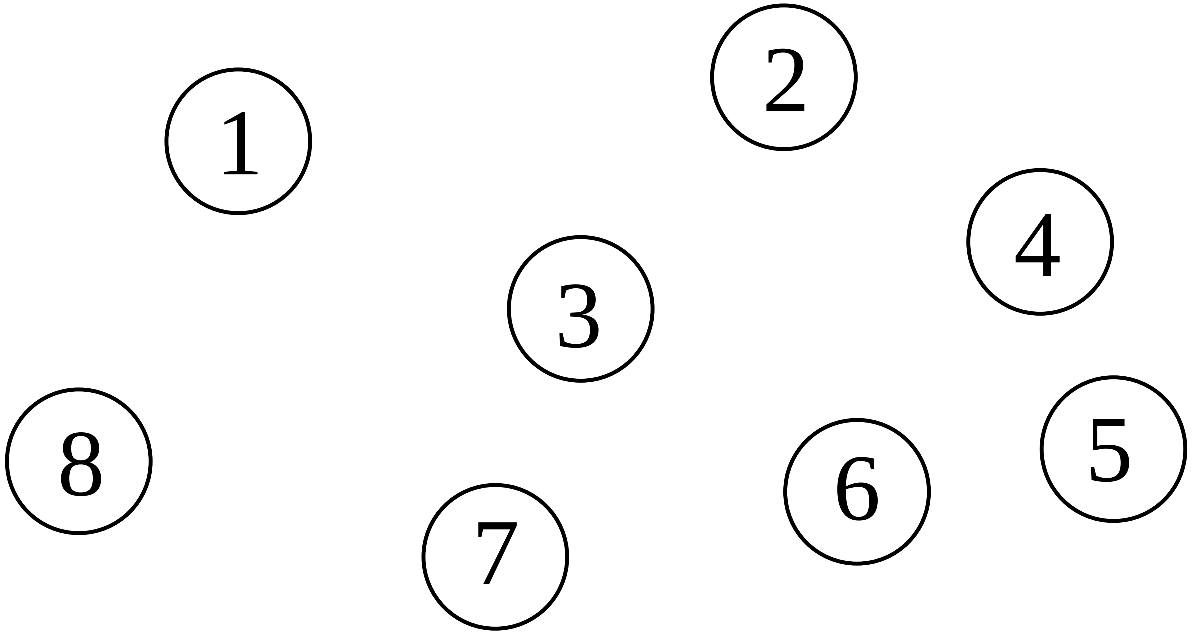
The distribution of random variable X is **stable**, if for all real a, b and independent copies X_1, X_2 of X there exists c, d with

$$aX_1 + bX_2 \equiv cX + d$$

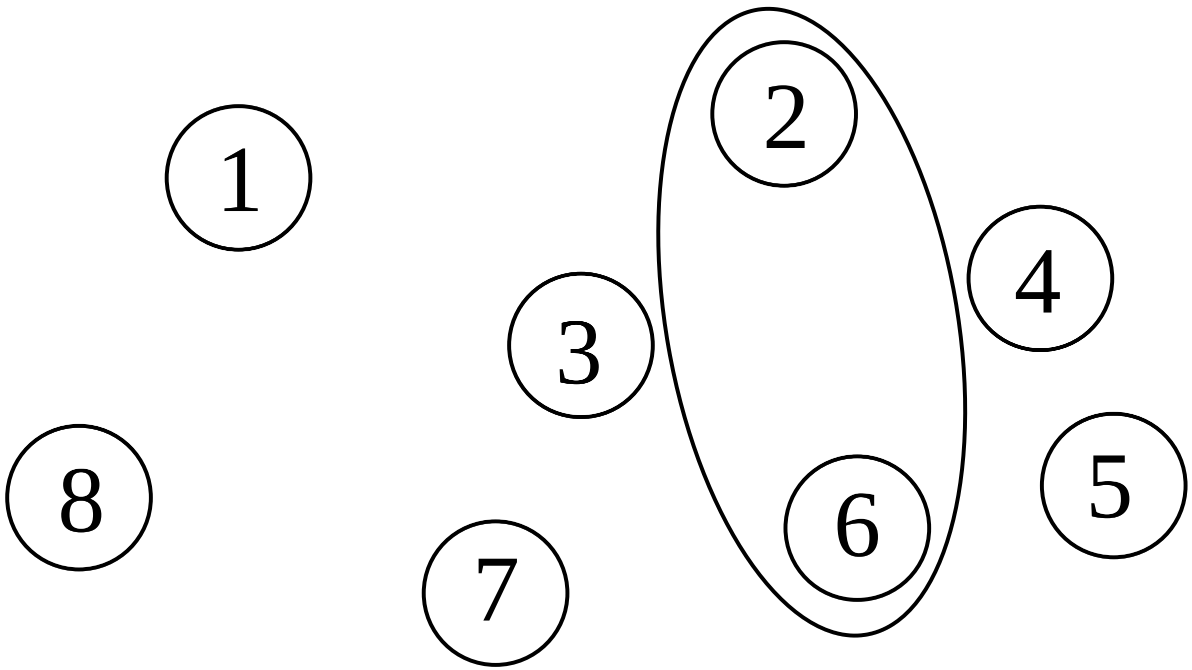
Examples: normal distribution, Cauchy distribution, Levy distribution

All stable distributions can be characterized in term of the characteristic function $\mathbb{E}\, e^{i\lambda X}$.

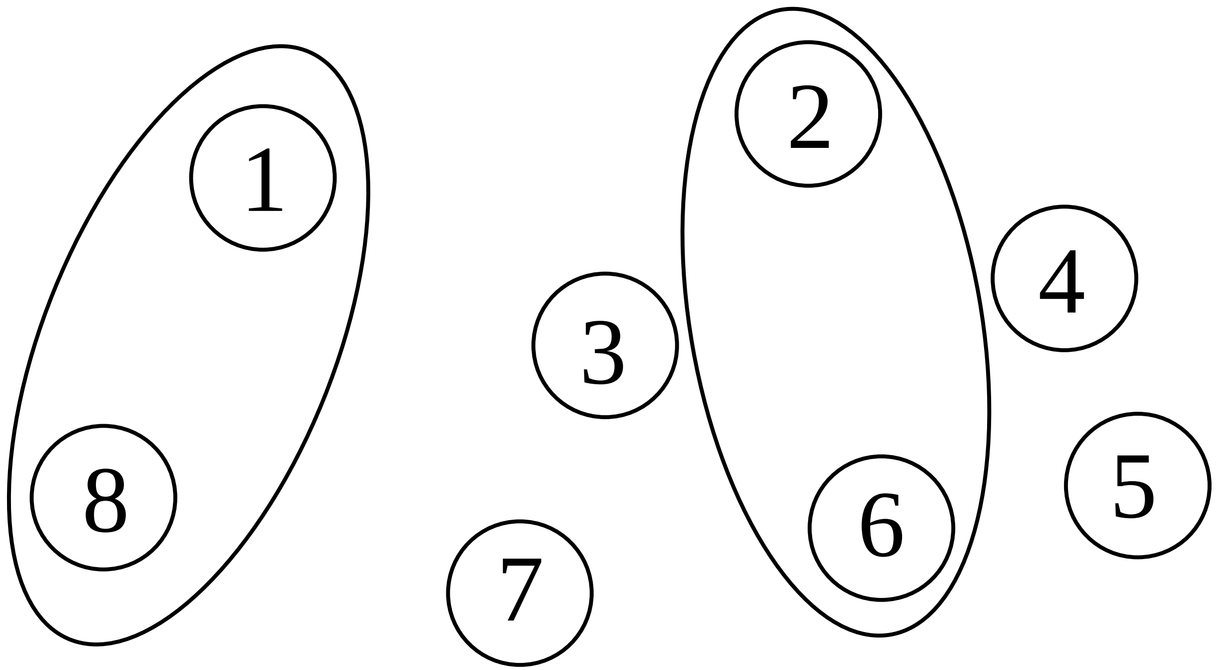
Coalescent Process



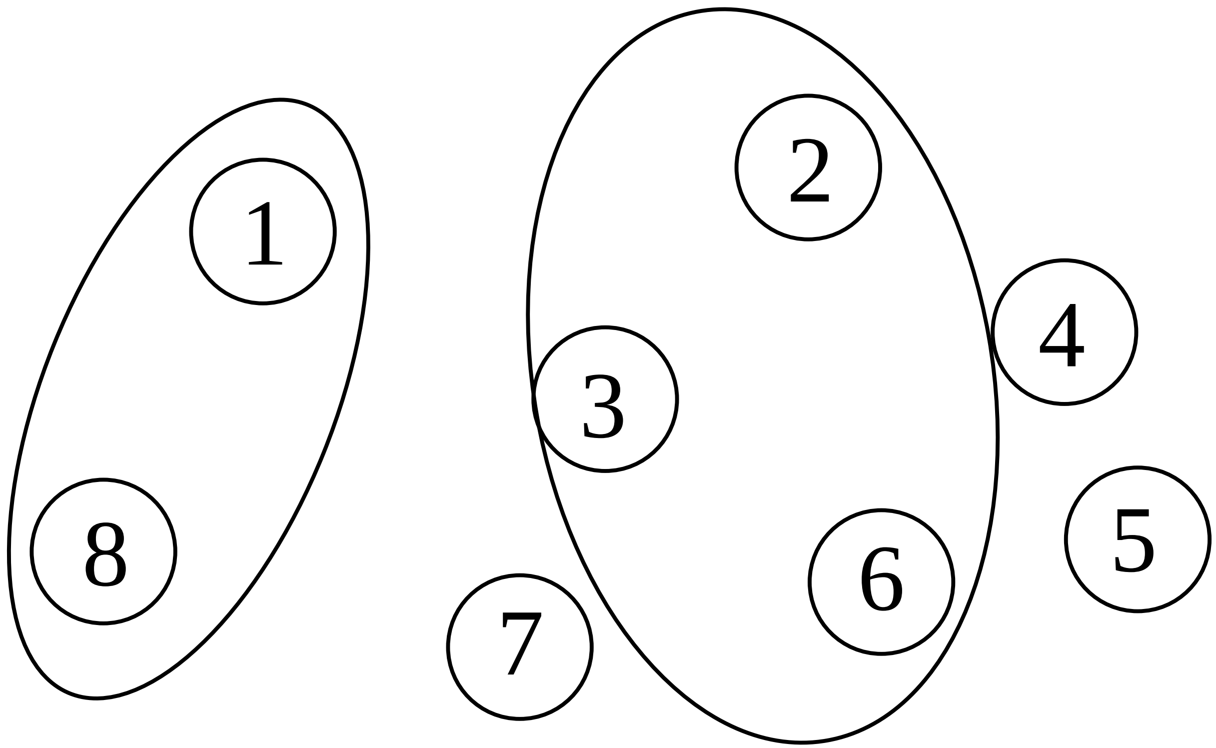
Coalescent Process



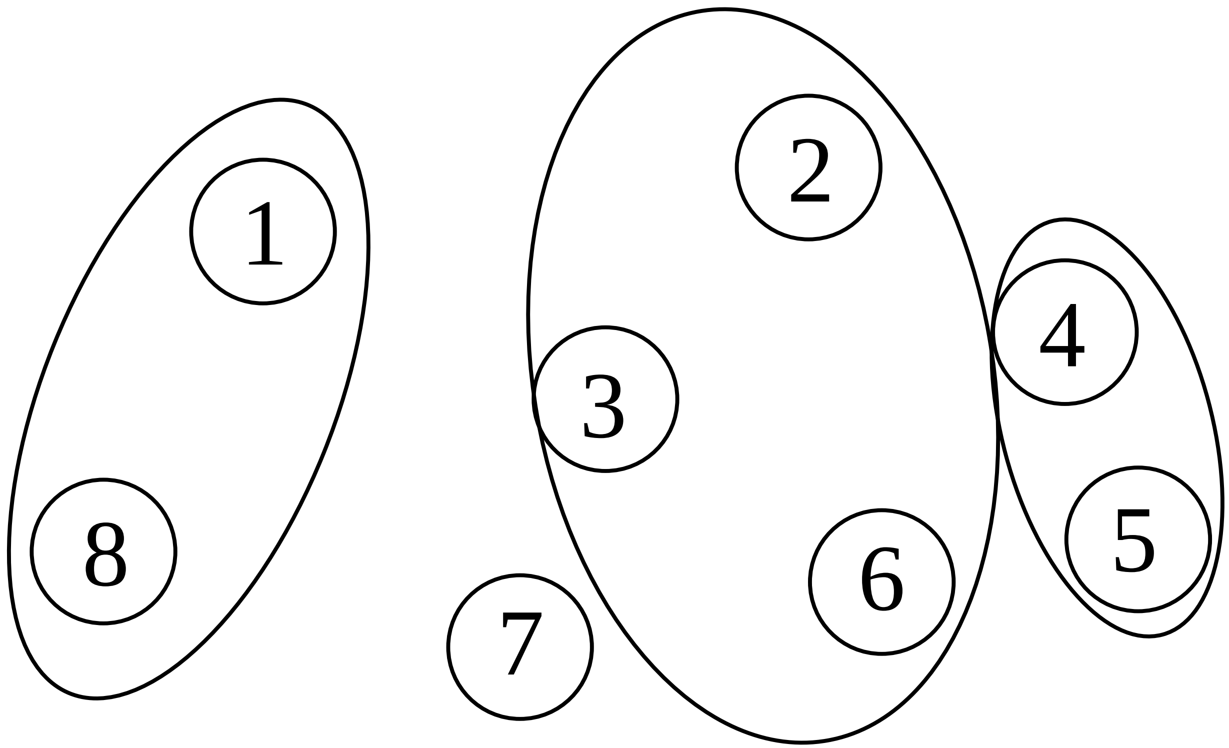
Coalescent Process



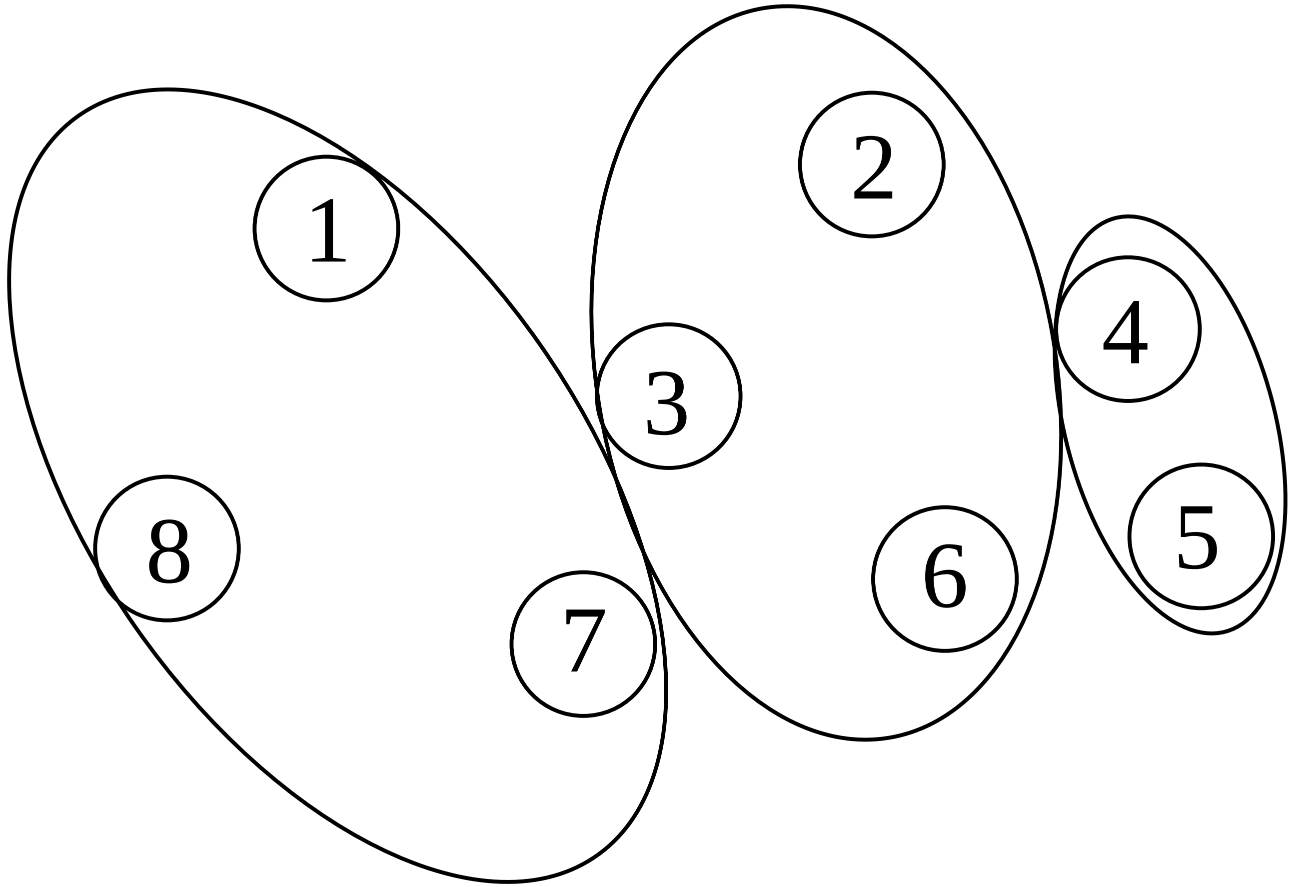
Coalescent Process



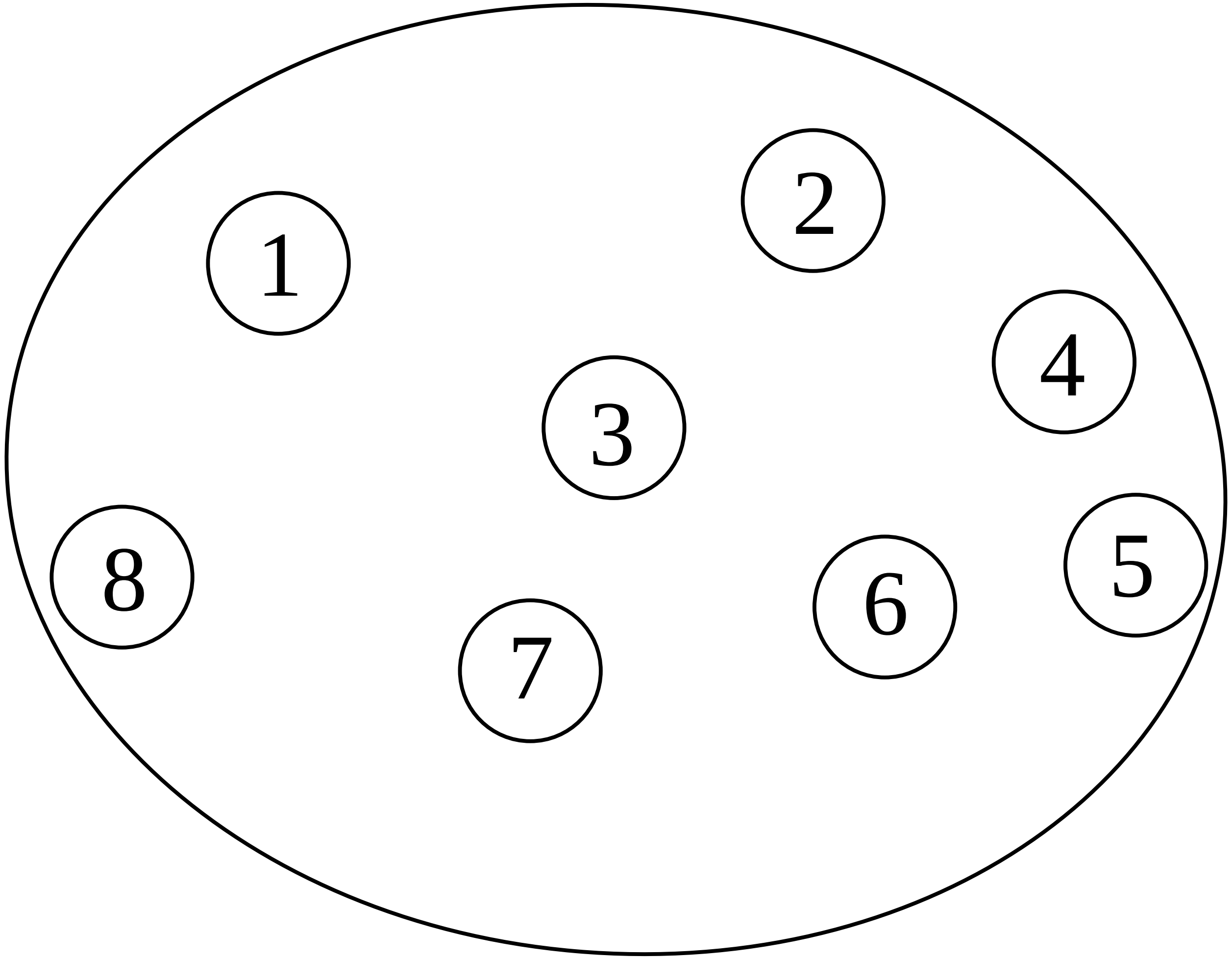
Coalescent Process



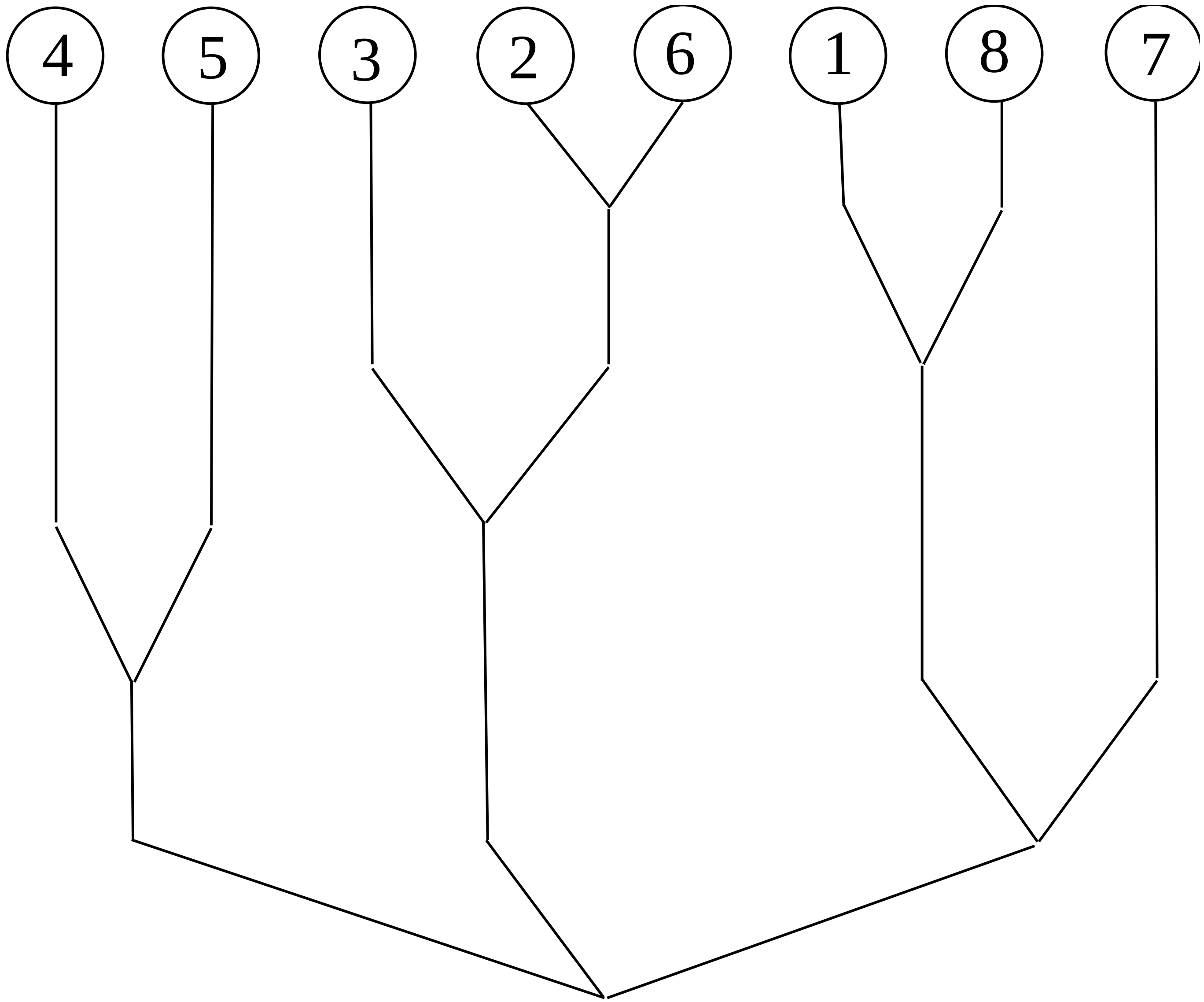
Coalescent Process



Coalescent Process



Coalescent Process



Coalescent Process

Stochastic model

Let Λ be a measure on $[0,1]$.

- Continuous Markov process of partitions of $\{1,2\ldots,n\}$,
  Initial partition: $\{\{1\},\{2\},\ldots,\{n\}\}$.

- If ξ and η are two partitions with a resp. b equivalence classes, where $b-a+1$ classes of ξ are merged to obtain η.
  Then the rate $q_{\xi,\eta}$ that ξ merges to η is

$$q_{\xi,\eta} = \begin{cases} \int_{[0,1]}(1-(1-x)^b - bx(1-x)^{b-1})x^{-2}d\Lambda(x) & \text{if } \xi = \eta, \\ \int_{[0,1]} x^{b-a-1}(1-x)^{a-1}d\Lambda(x) & \text{if } \xi \neq \eta. \end{cases}$$

Coalescent Process

Kingman-coalescent

$$\Lambda = \delta_0$$

Bolthausen-Sznitman-coalescent

$$\Lambda = \mathrm{univ}[0,1]$$

Coalescent Process

Remark

The process of number of classes is also a Markov process with rates

$$g_{ba} = \binom{b}{a-1} \int_{[0,1]} x^{b-a-1}(1-x)^{a-1} d\Lambda(x)$$

$(1 \leq a < b \leq n)$

Coalescent Process

Bolthausen-Sznitman-coalescent

X_n ... **number of collisions** until there is a single block:

$X_0 = X_1 = 0$,

$$\boxed{X_n \equiv X_{I_n} + 1} \qquad (n \ge 2),$$

where I_n is a discrete random variable with

$$\mathbb{P}\{I_n = k\} = \frac{1}{(n-k)(n-k+1)} \frac{n}{n-1} \qquad (0 \le k < n)$$

that is independent of $(X_0, X_1, \ldots, X_{n-1})$.

Cutting down Recursive Trees

Lemma

$$f(s,t) = \sum_{n\ge 1} \mathbb{E}\, s^{X_n}\, t^{n-1}$$

satisfies the partial differential equation

$$\boxed{\frac{\partial f(s,t)}{\partial t}\left(1 - t + \frac{t}{\log(1-t)}\left(1-\frac{1}{s}\right)\right) = f(s,t)}$$

with initial condition $f(s,0) = 1$.

Cutting down Recursive Trees

Expected Value

$$g(t) := \frac{\partial f(s,t)}{\partial s}\bigg|_{s=1} = \sum \mathbb{E}\, X_n\, t^{n-1}$$

$$\Longrightarrow \quad \boxed{g'(t) - \frac{g(t)}{1-t} = \frac{t}{(1-t)^3 \log \frac{1}{1-t}}}$$

$$\Longrightarrow \quad g(t) = \frac{1}{(1-t)^2 \log \frac{1}{1-t}} - \frac{\log\log \frac{1}{1-t}}{1-t} + O\left(\frac{1}{(1-t)^2 \log^2 \frac{1}{1-t}}\right)$$

$$\Longrightarrow \quad \boxed{\mathbb{E}\, X_n = \frac{n}{\log n} + O\left(\frac{n}{\log^2 n}\right)}.$$

Thank You!