PROFILE OF RANDOM TREES

Michael Drmota

Institute of Discrete Mathematics and Geometry

Vienna University of Technology

A 1040 Wien, Austria

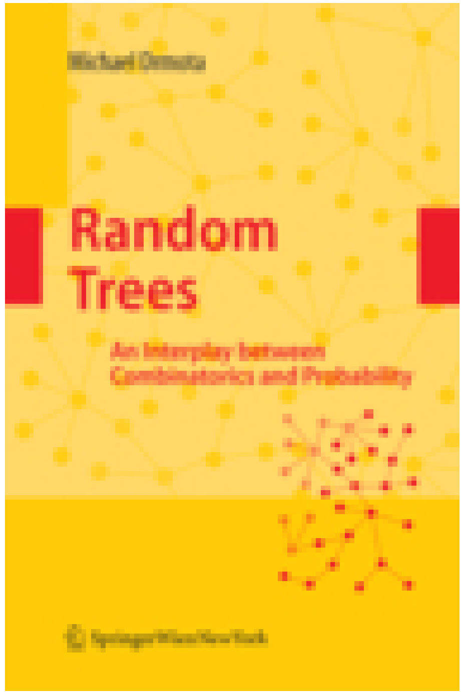
michael.drmota@tuwien.ac.at

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

LIPN Paris Nord, February 24, 2015

Contents

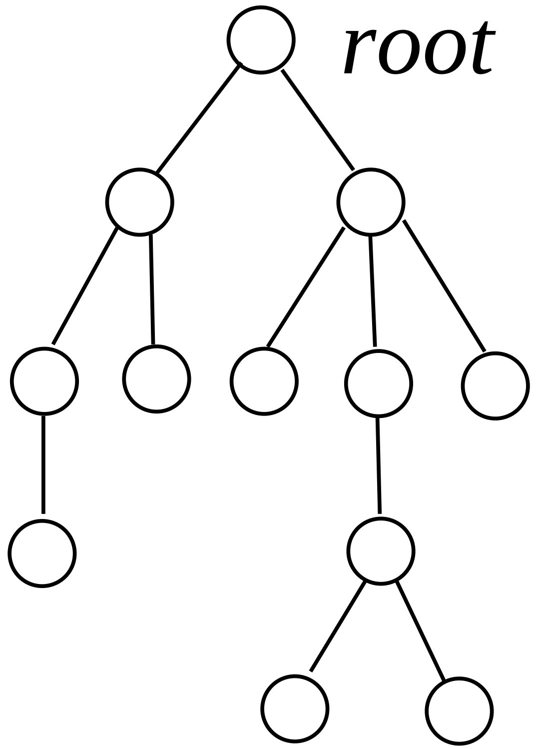
Book



Michael Drmota,
Random Trees, Springer, Wien-New York, 2009.

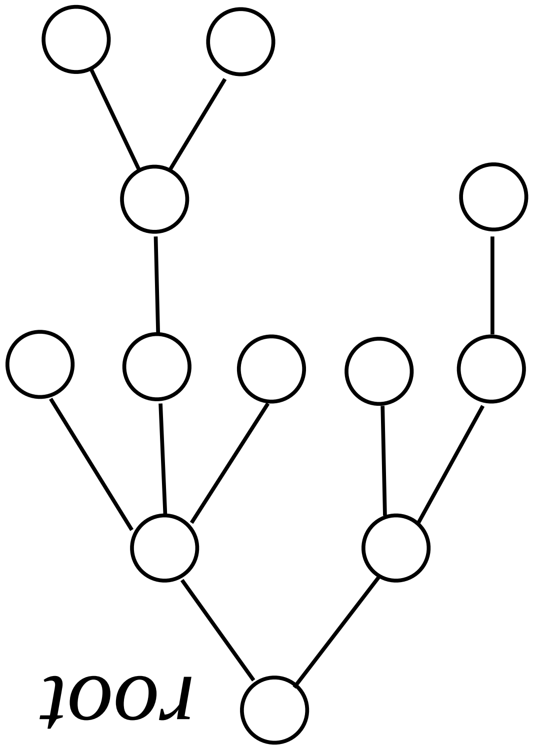
Profile of Trees

Rooted tree



Profile of Trees

Rooted tree



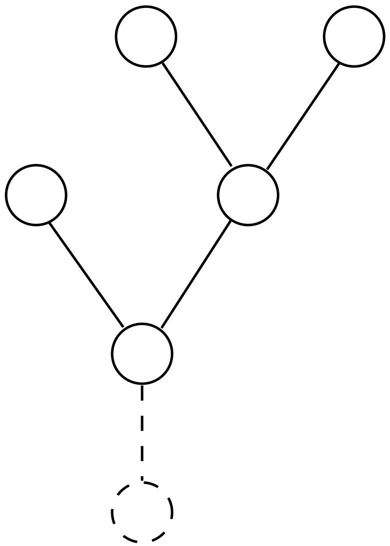
Profile of Trees

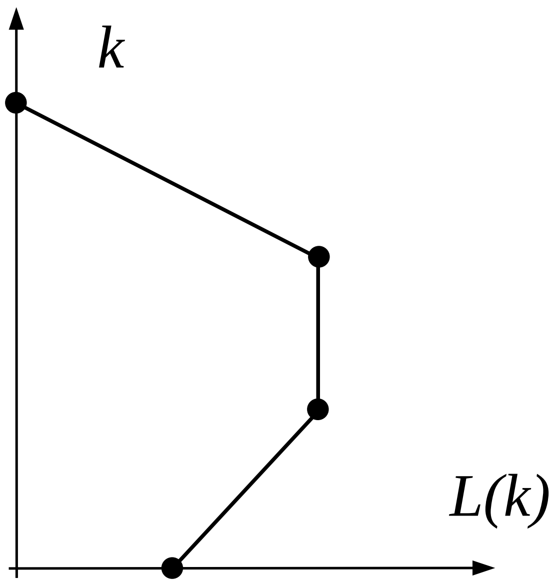
$I_T(k)$ ... number of nodes at distance k from the root

$(I_T(k))_{k\geq 0}$ ... profile of T

$(I_T(s), s \geq 0)$ ... linearly interpolated profile of T

$(I_{n,k})_{k\geq 0}$ ... profile in a **random tree** of size n





Profile of Trees

Parameters of interest:

- **Profile** $I_{n,k}$ (number of nodes at depth k)
- **Depth** of a random node: D_n
- **Internal path length**: L_n (sum of all distances to the root)
- **Height** H_n

Profile of Trees

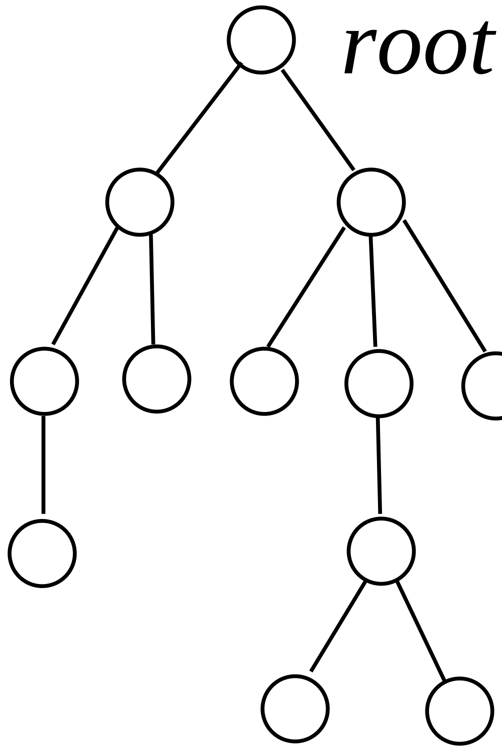
Relations to the profile $I_{n,k}$:

- $\mathbf{Pr}\{D_n = k\} = \dfrac{1}{n}\,\mathbf{E}\,I_{n,k}$
- $L_n = \sum_{k \ge 0} k I_{n,k}$
- $H_n = \max\{k \ge 0 : I_{n,k} > 0\}$
- The profile describes the **shape** of the tree.

Contents

Galton-Watson Trees

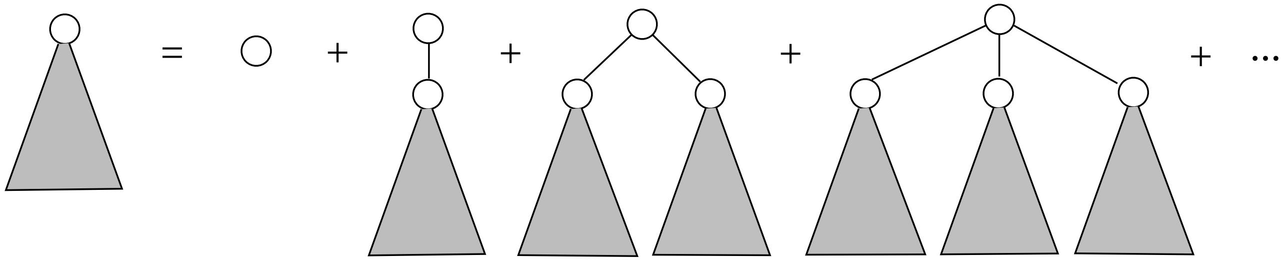
Catalan trees



rooted, ordered (or plane) tree

Galton-Watson Trees

Catalan trees. g_n = number of Catalan trees of size n; $G(x) = \sum_{n \geq 1} g_n x^n$



$$G(x) = x(1 + G(x) + G(x)^2 + \cdots) = \frac{x}{1 - G(x)}$$

$$G(x) = \frac{1 - \sqrt{1 - 4x}}{2} \quad \Longrightarrow \quad g_n = \frac{1}{n}\binom{2n-2}{n-1} \sim \frac{4^{n-1}}{\sqrt{\pi} \cdot n^{3/2}}$$

(Catalan numbers)

Galton-Watson Trees

Catalan trees with singularity analysis

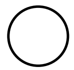
$$G(x) = \frac{1-\sqrt{1-4x}}{2} = \frac{1}{2} - \frac{1}{2}\sqrt{1-4x}$$

$$\Longrightarrow \quad g_n \sim -\frac{1}{2} \cdot \frac{4^n n^{-3/2}}{\Gamma(-\frac{1}{2})} = \frac{4^{n-1}}{\sqrt{\pi} \cdot n^{3/2}}$$

Galton-Watson Trees

Galton-Watson branching process

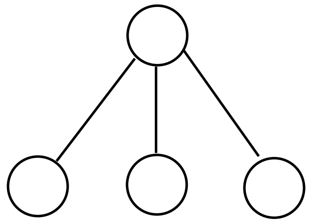
ξ ... offspring distribution, $\varphi_k = \mathbb{P}\{\xi = k\}$, $\varphi_0 > 0$



Galton-Watson Trees

Galton-Watson branching process

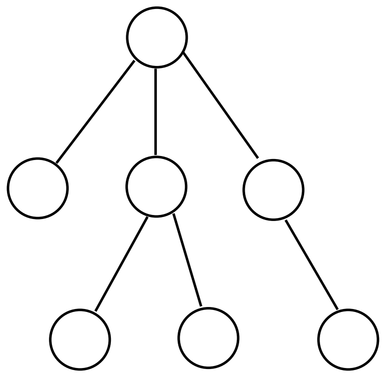
ξ ... offspring distribution, $\varphi_k = \mathbb{P}\{\xi = k\}$, $\varphi_0 > 0$



Galton-Watson Trees

Galton-Watson branching process

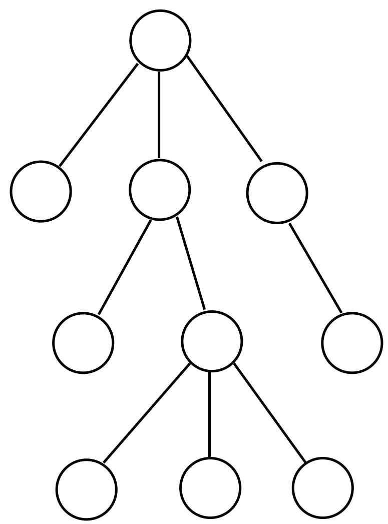
ξ ... offspring distribution, $\varphi_k = \mathbb{P}\{\xi = k\}$, $\varphi_0 > 0$



Galton-Watson Trees

Galton-Watson branching process

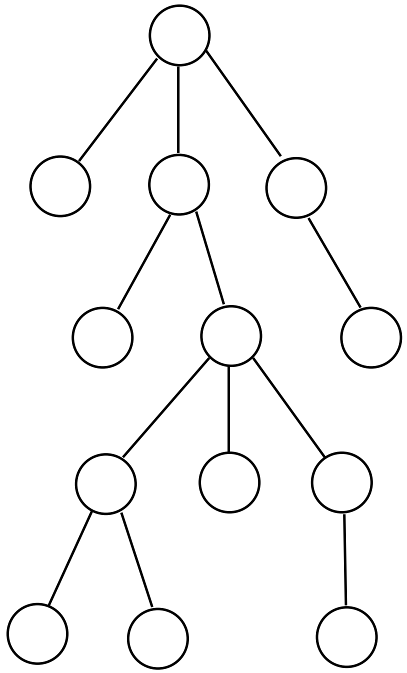
ξ ... offspring distribution, $\varphi_k = \mathbb{P}\{\xi = k\}$, $\varphi_0 > 0$



Galton-Watson Trees

Galton-Watson branching process

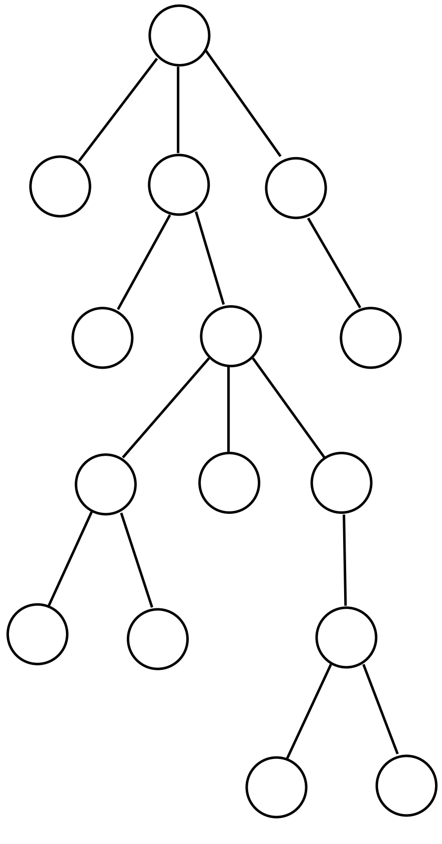
ξ ... offspring distribution, $\varphi_k = \mathbb{P}\{\xi = k\}$, $\varphi_0 > 0$



Galton-Watson Trees

Galton-Watson branching process

ξ ... offspring distribution, $\varphi_k = \mathbb{P}\{\xi = k\}$, $\varphi_0 > 0$



Galton-Watson Trees

Galton-Watson branching process. $(Z_k)_{k\geq 0}$

$Z_0 = 1$, and for $k \geq 1$

$$\boxed{Z_k = \sum_{j=1}^{Z_{k-1}} \xi_j^{(k)}},$$

where the $(\xi_j^{(k)})_{k,j}$ are iid random variables distributed as ξ.

Z_k ... number of nodes in k-th generation

$Z = Z_0 + Z_1 + Z_2 + \cdots$ ... total progeny

Galton-Watson Trees

Generating functions

$$y_n = \mathbb{P}\{Z = n\}, \qquad y(x) = \sum_{n\geq 1} y_n x^n$$

$$\Phi(w) = \mathbb{E}\, w^{\xi} = \sum_{k\geq 0} \varphi_k w^k$$

$$\Longrightarrow \quad \boxed{y(x) = x\,\Phi(y(x))}$$

Conditioned Galton-Watson tree

GW-branching process conditioned on the total progeny $Z = n$.

Galton-Watson Trees

Example. $\mathbb{P}\{\xi = k\} = 2^{-k-1}$, $\Phi(w) = 1/(2-w)$

$\Longrightarrow$ all trees of size n have the same probability

$\Longrightarrow$ conditioned GW-tree of size n is the same model as the **Catalan tree model** (with the uniform distribution on trees of size n)

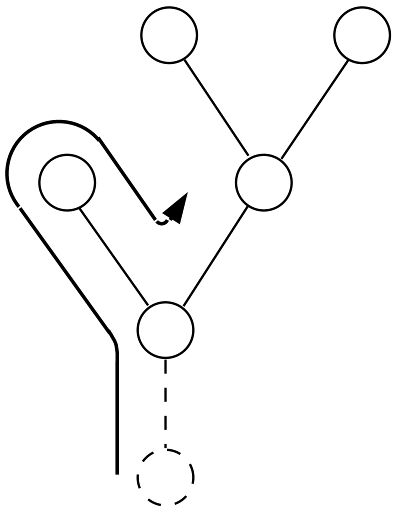
Example. $\Phi(w) = \frac{1}{2}(1+w)^2$: **binary trees** with n internal nodes.

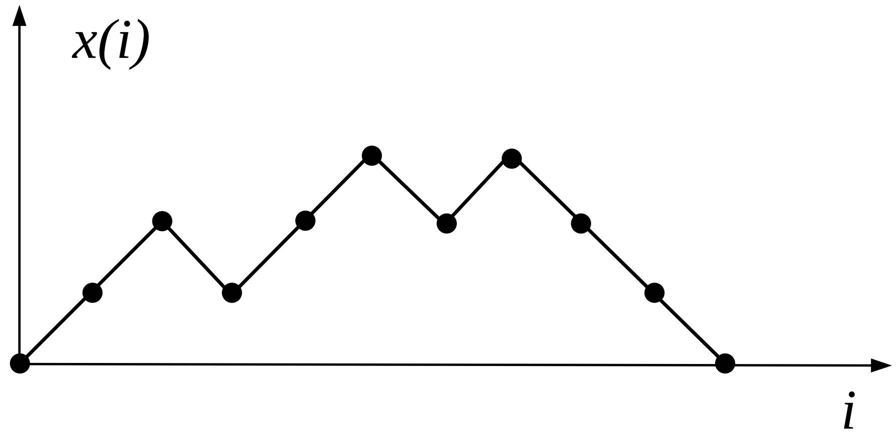
Example. $\Phi(w) = \frac{1}{3}(1+w+w^2)$: **Motzkin trees**

Example. $\Phi(w) = e^{w-1}$: **Cayley trees**

Galton-Watson Trees

Depth-First-Search – Rooted trees and discrete excursions





Bijection between

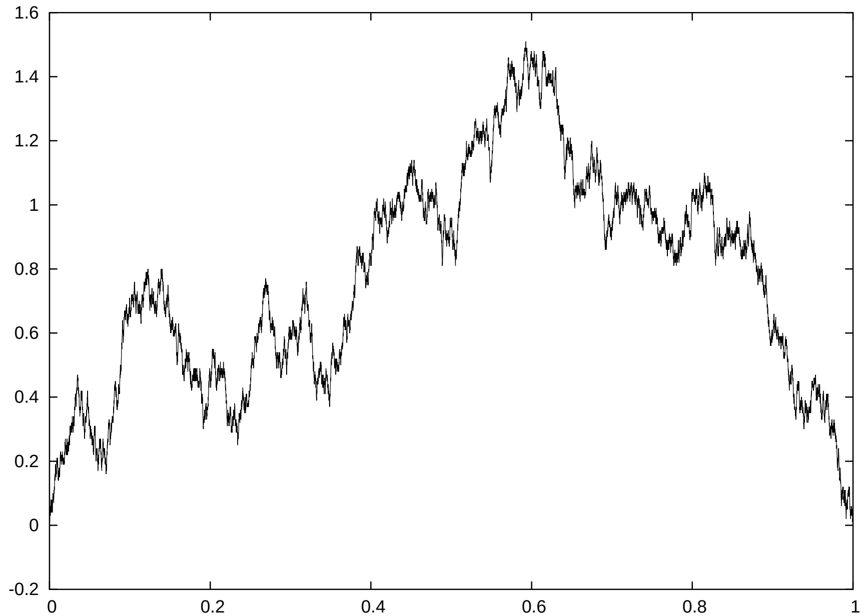
Catalan trees $\longleftrightarrow$ Dyck paths

random trees of size n $\longleftrightarrow$ **random** Dyck paths of length $2n$

Galton-Watson Trees

Depth-First-Search

Brownian excursion $(e(t), 0 \le t \le 1)$



Rescaled Brownian motion between 2 zeros.

Random function in $C[0,1]$.

Depth-First-Search

Kaigh's Theorem

$(X_n(t), 0 \le t \le 2n)$ ... random Dyck path of length $2n$.

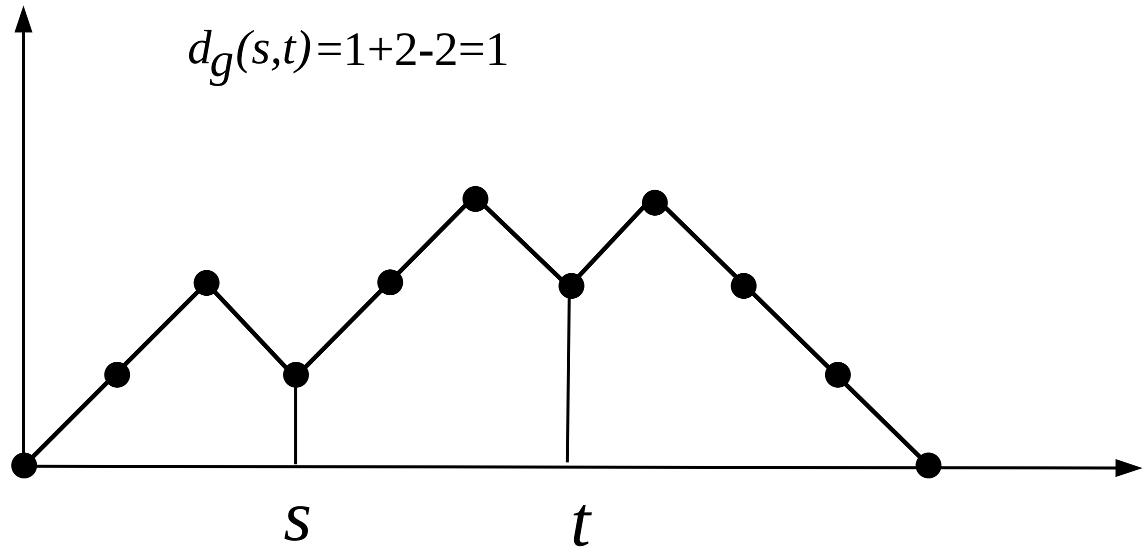
$$\Longrightarrow \quad \left(\frac{1}{\sqrt{2n}} X_n(2nt), 0 \le t \le 1\right) \xrightarrow{\mathrm{d}} (e(t), 0 \le t \le 1).$$

Remark. This theorem also holds for more general random walks with independent increments conditioned to be an excursion.

Galton-Watson Trees

$g : [0,1] \to [0,\infty)$ ... continuous, ≥ 0, $g(0) = g(1) = 0$

$$d_g(s,t) = g(s) + g(t) - 2 \inf_{\min\{s,t\} \leq u \leq \max\{s,t\}} g(u)$$



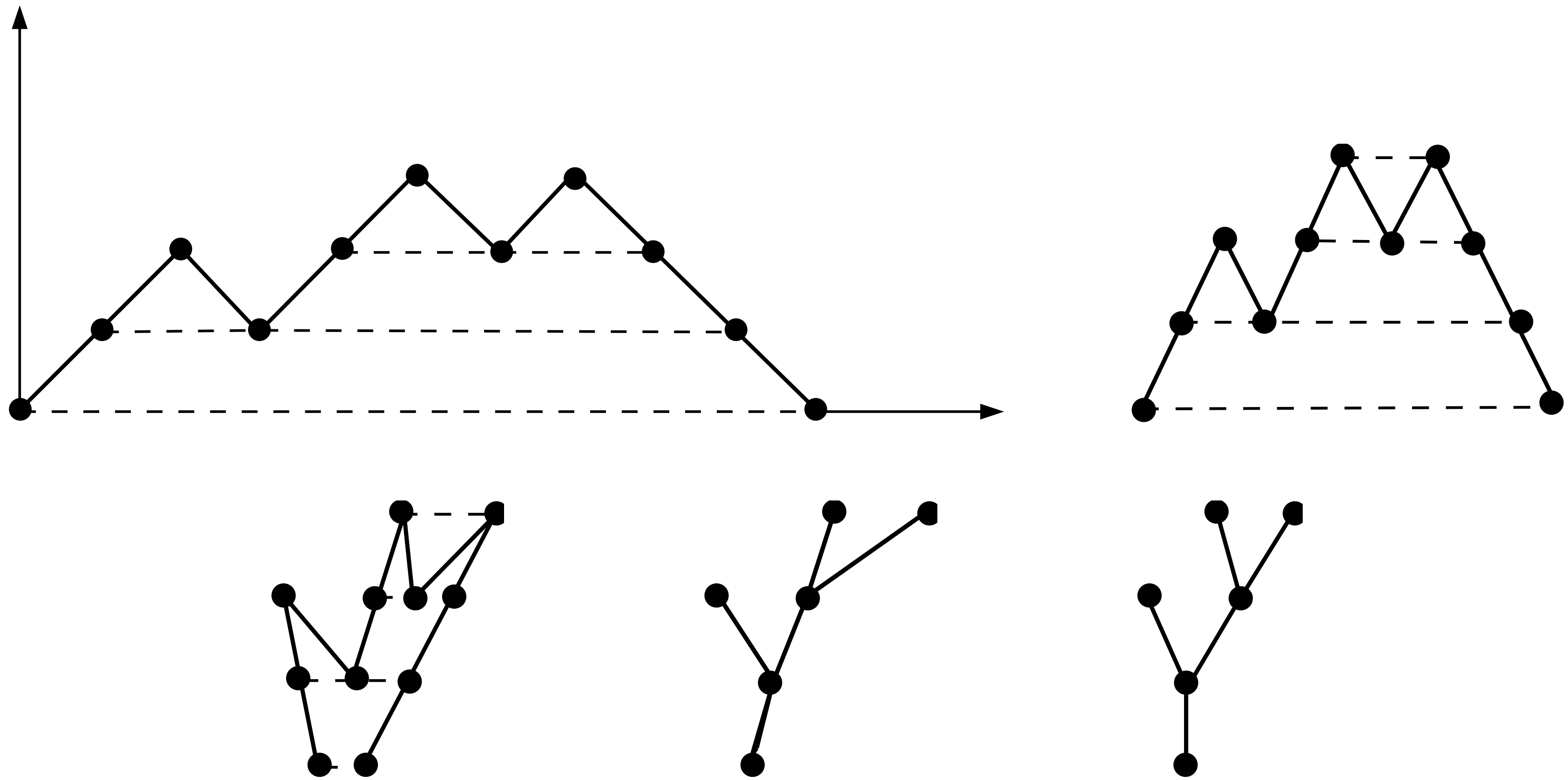
$$s \sim t \quad \Longleftrightarrow \quad d_g(s,t) = 0$$

$$\mathcal{T}_g = [0,1]/\sim$$

$\Longrightarrow$ $(\mathcal{T}_g, d_g)$ is a compact (so-called) real tree.

Galton-Watson Trees

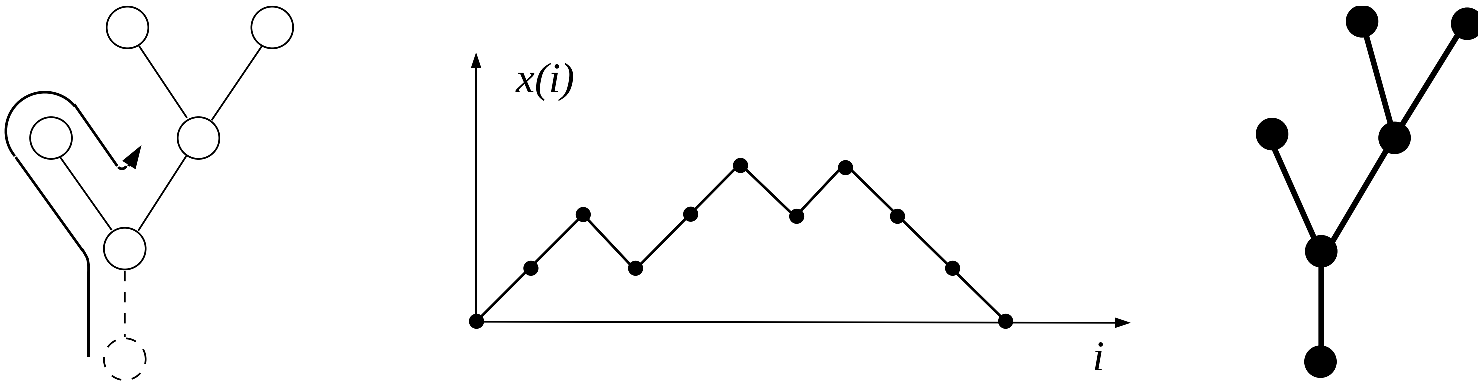
Construction of a real tree $\mathcal{T}_g$



The mapping $C[0,1] \to \mathbb{T}$, $g \mapsto \mathcal{T}_g$ is **continuous**.

Galton-Watson Trees

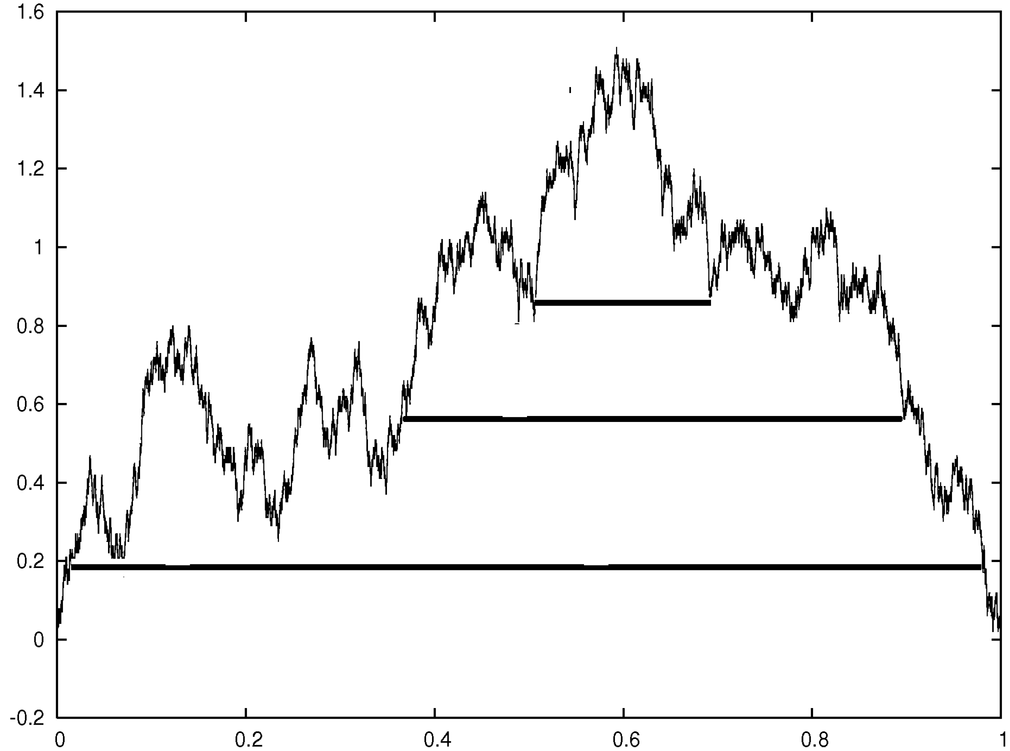
Catalan trees as real trees



T_n $\qquad$ $X_n = X_{T_n}$ $\qquad$ $\mathcal{T}_{X_n}$

Galton-Watson Trees

Continuum random tree $\mathcal{T}_{2e}$ (with Brownian excursion $e(t)$)



Galton-Watson Trees

Theorem

$(X_n(t), 0 \le t \le 2n)$ ... random Dyck paths of length $2n$
or the depth-first-search process of Catalan trees of size n.

$$\Longrightarrow \quad \boxed{\frac{1}{\sqrt{2n}}\mathcal{T}_{X_n} \xrightarrow{\text{d}} \mathcal{T}_{2e}}$$

In other words...

Scaled Catalan trees (interpreted as "real trees") converge weakly to the continuum random tree.

Galton-Watson Trees

General assumption: $\boxed{\mathbb{E}\,\xi = 1}$, $\boxed{0 < \mathbb{V}\mathrm{ar}\,\xi = \sigma^2 < \infty}$

Theorem (Aldous)

$X_n(t)$ ... depth-first-search of conditioned GW-trees of size n

$$\Longrightarrow \quad \boxed{\left(\frac{\sigma}{2\sqrt{n}} X_n(2nt), 0 \le t \le 1\right) \xrightarrow{\mathrm{d}} (e(t), 0 \le t \le 1)}.$$

Corollary

$$\boxed{\frac{\sigma}{\sqrt{n}} \mathcal{T}_{X_n} \xrightarrow{\mathrm{d}} \mathcal{T}_{2e}}$$

Galton-Watson Trees

Corollary H_n ... height of conditioned GW-trees of size n:

$$\Longrightarrow \quad \boxed{\frac{1}{\sqrt{n}} H_n \xrightarrow{\mathrm{d}} \frac{2}{\sigma} \max_{0 \le t \le 1} e(t)}$$

Remark. Distribution function of $\max_{0 \le t \le 1} e(t)$:

$$\mathbb{P}\{\max_{0 \le t \le 1} e(t) \le x\} = 1 - 2 \sum_{k=1}^{\infty} (4x^2k^2 - 1) e^{-2x^2k^2}$$

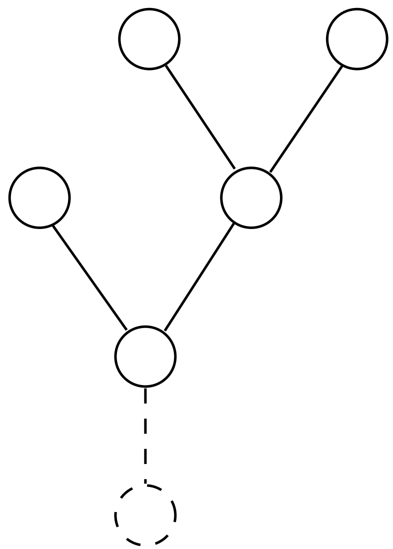
Galton-Watson Trees

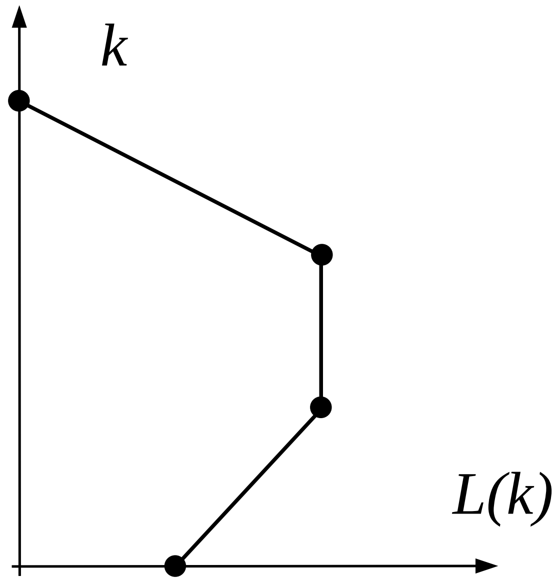
Profile

$I_T(k)$ ... number of nodes at distance k from the root

$(I_T(k))_{k\geq 0}$ ... profile of T

$(I_T(s), s \geq 0)$ ... linearly interpolated profile of T





Galton-Watson Trees

Value distribution

$$\mu_T = \frac{1}{|T|} \sum_{k \geq 0} I_T(k)\, \delta_k$$

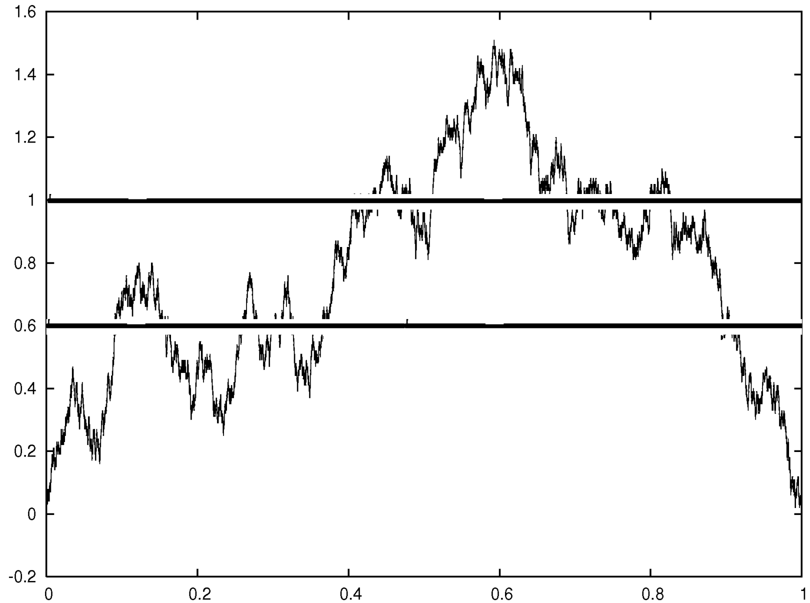
δ_x ... δ-distribution concentrated at x

Galton-Watson Trees

Occupation measure: random measure on $\mathbb{R}$

$$\mu(A) = \int_0^1 \mathbf{1}_A(e(t)\,dt$$

measure how long $e(t)$ stays in set A



Galton-Watson Trees

Theorem (Aldous)

$(I_{n,k}, k \ge 0)$ ... random profile of conditioned GW-trees of size n

$$\Longrightarrow \quad \boxed{\frac{1}{n}\sum_{k\ge 0} I_{n,k}\,\delta_{(\sigma/2)k/\sqrt{n}} \xrightarrow{\mathrm{d}} \mu}$$

Galton-Watson Trees

Local time of the Brownian excursion: random density of μ

$$\boxed{l(s) = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_0^1 \mathbf{1}_{[s,s+\varepsilon]}(e(t))\, dt}$$

Theorem (D.+Gittenberger)

$(l_n(s), s \ge 0)$ ... random profile of conditioned GW-trees of size n

$$\Longrightarrow \quad \boxed{\left(\frac{1}{\sqrt{n}} l_n(s\sqrt{n}),\ s \ge 0\right) \xrightarrow{\text{d}} \left(\frac{\sigma}{2} l\left(\frac{\sigma}{2}s\right),\ s \ge 0\right)}$$

Proof with asymptotics on generating functions (very involved)!!!

Galton-Watson Trees

Width

$$W = \max_{k \ge 0} L(k) = \max_{t \ge 0} L(t),$$

maximal number of nodes in a level.

Corollary

$$\boxed{\frac{1}{\sqrt{n}} W_n \xrightarrow{\text{d}} \frac{\sigma}{2} \sup_{0 \le t \le 1} l(t)}$$

Remark. $\sup_{t \ge 0} l(t) = 2 \sup_{0 \le t \le 1} e(t)$ (in distribution)

Contents

(Binary) Search Trees

Storing of data

4,6,3,5,1,8,2,7

(Binary) Search Trees

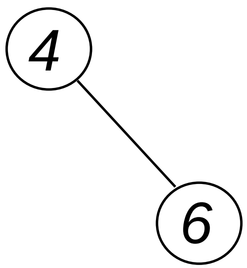
Storing of data

6,3,5,1,8,2,7

(Binary) Search Trees

Storing of data

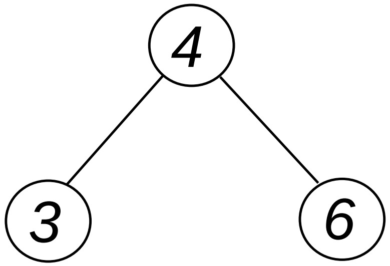
3,5,1,8,2,7



(Binary) Search Trees

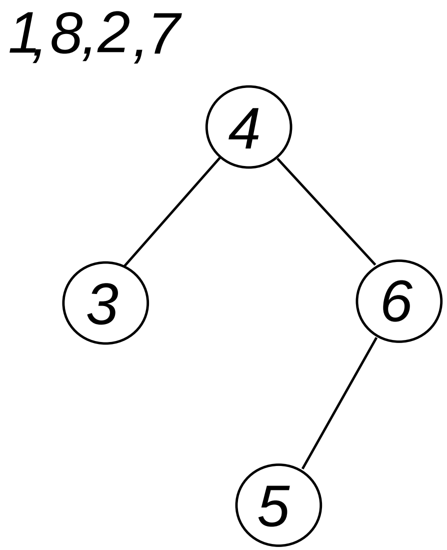
Storing of data

5,1,8,2,7



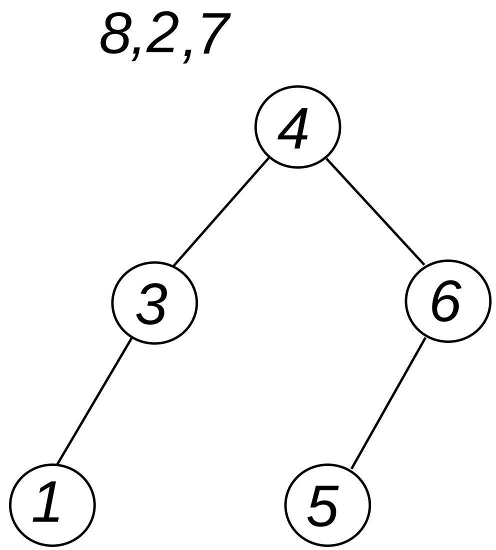
(Binary) Search Trees

Storing of data



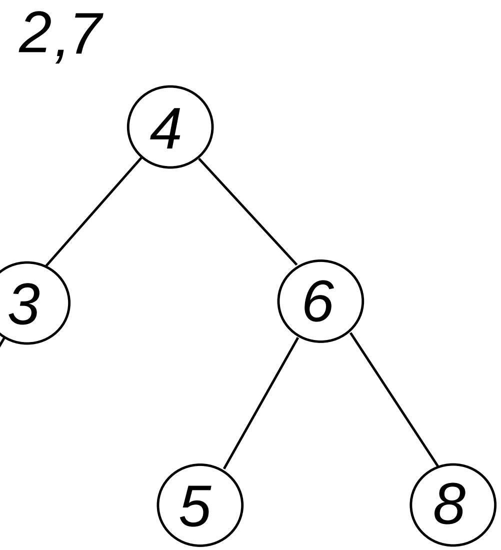
(Binary) Search Trees

Storing of data



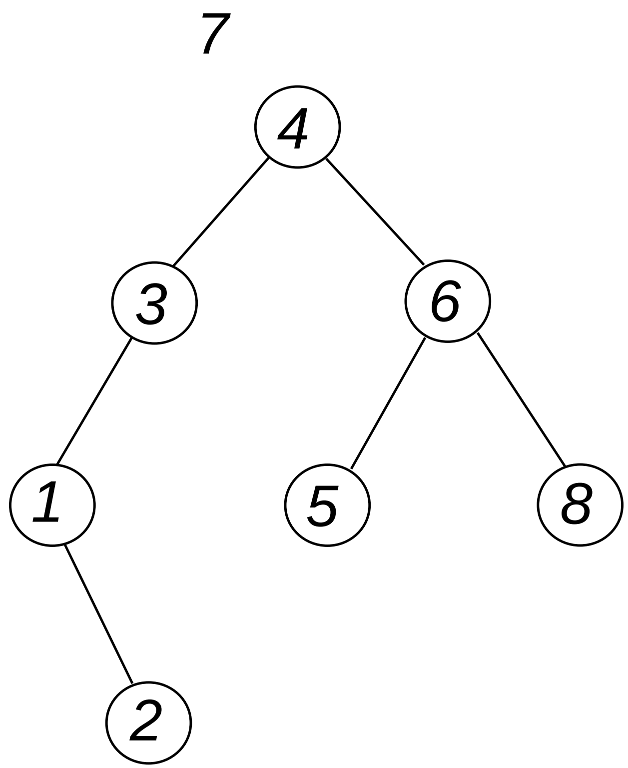
(Binary) Search Trees

Storing of data



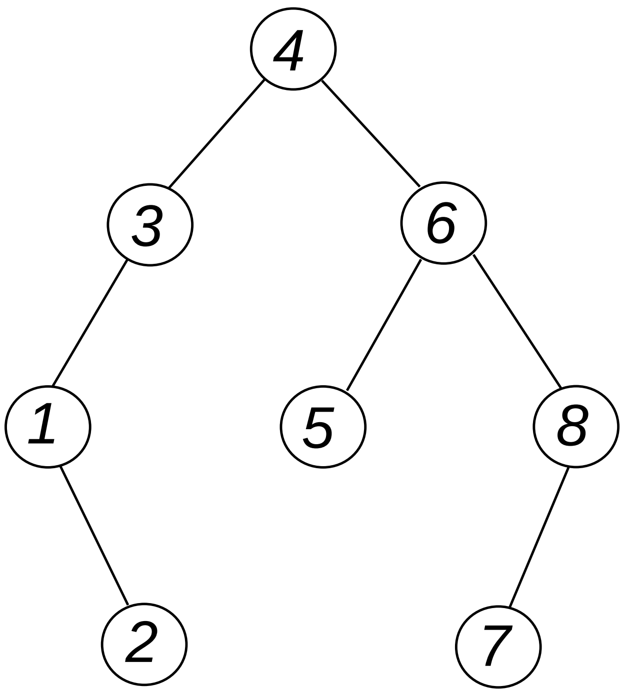
(Binary) Search Trees

Storing of data



(Binary) Search Trees

Storing of data



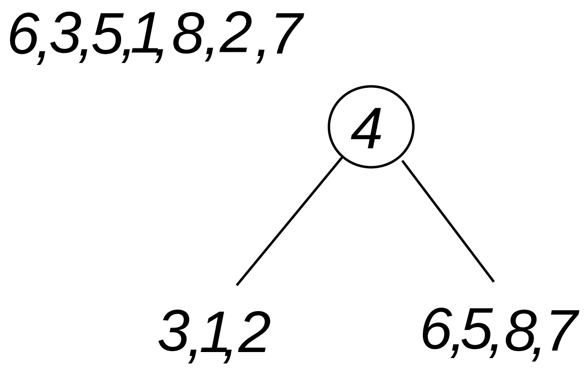
(Binary) Search Trees

Quicksort – Sorting of data

4,6,3,5,1,8,2,7

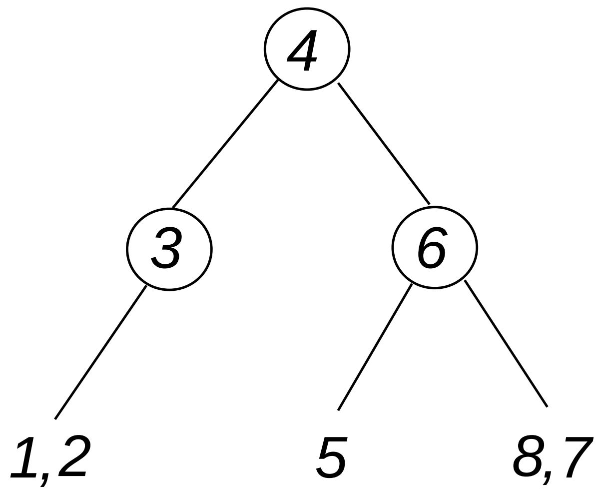
(Binary) Search Trees

Quicksort – Sorting of data



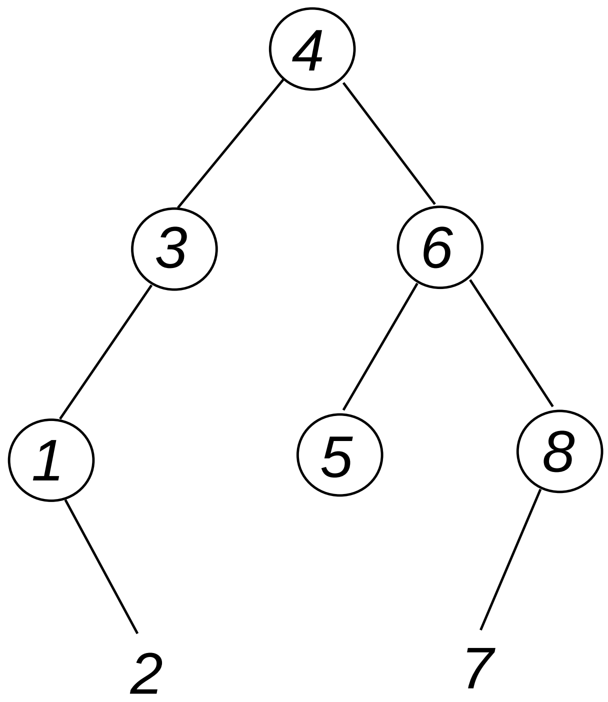
(Binary) Search Trees

Quicksort – Sorting of data



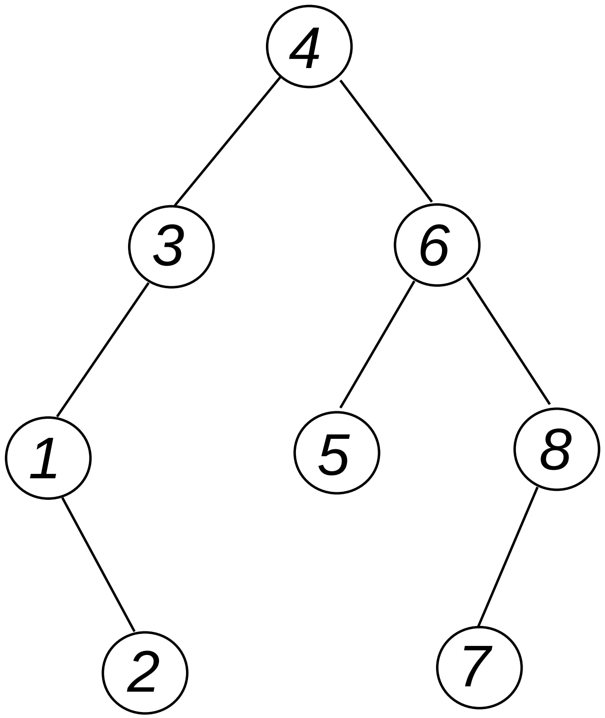
(Binary) Search Trees

Quicksort – Sorting of data



(Binary) Search Trees

Quicksort – Sorting of data



(Binary) Search Trees

Quicksort – Median of 3

4,6,3,5,1,8,2,7

(Binary) Search Trees

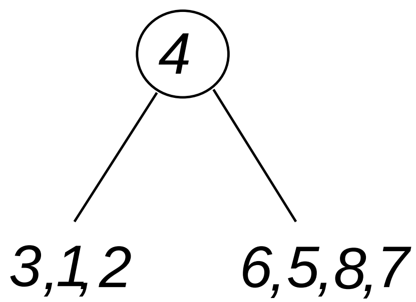
Quicksort – Median of 3

↓

4,6,3,5,1,8,2,7

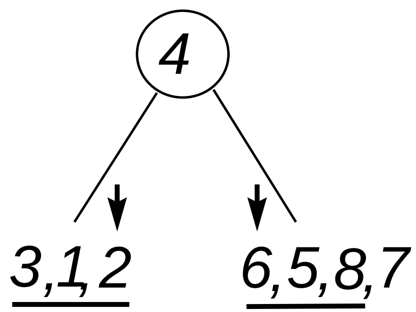
(Binary) Search Trees

Quicksort – Median of 3



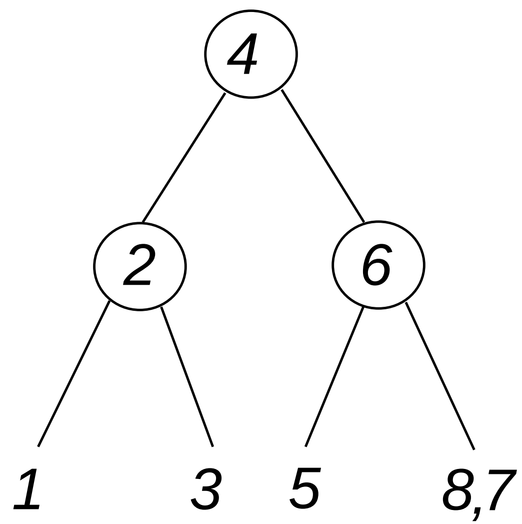
(Binary) Search Trees

Quicksort – Median of 3



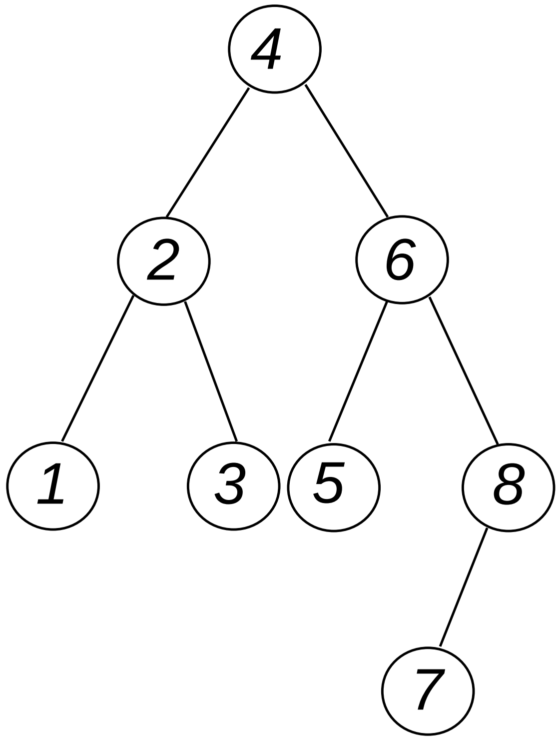
(Binary) Search Trees

Quicksort – Median of 3



(Binary) Search Trees

Quicksort – Median of 3



(Binary) Search Trees

Probabilistic Model

Every permutation on the data $\{1, 2, \ldots, n\}$ ist equally likely

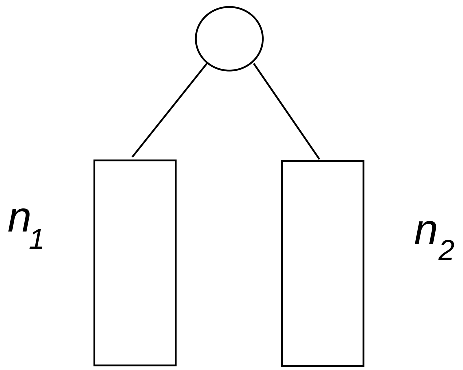
$\longrightarrow$ probability distribution on binary (m-ary) trees of size n

$\longrightarrow$ all tree parameters are **random variables**

(Binary) Search Trees

Probabilistic Model – Recursive structure

Subtrees have the same structure:
$(n = n_1 + n_2 + 1)$.



Splitting probabilities: p_{n_1,n_2}

Quicksort: $$p_{n_1,n_2} = \frac{1}{n}$$

Median of 3: $$p_{n_1,n_2} = \frac{n_1 n_2}{\binom{n}{3}}$$

Search Trees

General Model

$m \geq 2$, $t \geq 0$ ... given integers $\qquad$ n keys (data)

- If $n \geq m$, we **randomly select** $m-1$ **pivots** $\boxed{x_1 < x_2 < \cdots < x_{m-1}}$.

- The pivots are stored in the **root**.

- The remaining $n-m+1$ keys are divided into m **subsets** $I_1, \ldots, I_m$:

  $I_1 := \{x_i : x_i < x_1\},\ I_2 := \{x_i : x_1 < x_i < x_2\},\ \ldots, I_m := \{x_i : x_{m-1} < x_i\}.$

- Apply this procedure **recursively** to $I_1, I_2, \ldots, I_m$.

Search Trees

General Splitting Probabilities

$\mathbf{V}_n = (V_{n,1}, V_{n,2}, \ldots, V_{n,m})$.. random splitting vector

$V_{n,k} := |I_k|$ ... number of keys in the kth subset
(= the number of nodes in the kth subtree of the root)

$$V_{n,1} + V_{n,2} + \cdots + V_{n,m} = n - (m-1) = n + 1 - m$$

$$\boxed{\mathbb{P}\{\mathbf{V}_n = (n_1, \ldots, n_m)\} = \frac{\binom{n_1}{t} \cdots \binom{n_m}{t}}{\binom{n}{mt+m-1}}}$$

$$(n_1 + n_2 + \cdots + n_m = n - m + 1)$$

Quicksort: $m = 2$, $t = 0$ **Median of 3:** $m = 2$, $t = 1$

Search Trees

Recursive relation for the profile:

$$\boxed{I_{n,k} \stackrel{\mathrm{d}}{=} I^{(1)}_{V_{n,1},k-1} + I^{(2)}_{V_{n,2},k-1} + \cdots + I^{(m)}_{V_{n,m},k-1}}$$

$(I^{(j)}_{n,k})_{k\ge 0}$, $j = 1, \ldots, m$ ... independent copies of $X_{n,k}$

Search Trees

Expected Profile

$$F(\theta) := \frac{t!}{m(mt+m-1)!}(\theta+t)(\theta+t+1)\cdots(\theta+mt+m-2),$$

$\boxed{\lambda_1(z)}$, $\lambda_2(z)$, ..., $\lambda_{(m-1)(t+1)}(z)$ ... roots of $\boxed{F(\theta)=z}$:
$\boxed{\Re(\lambda_1(z))} \geq \Re(\lambda_2(z)) \geq \dots$.

$\beta(\alpha) > 0$ ... defined by $\boxed{\beta(\alpha)\lambda_1'(\beta(\alpha)) = \alpha}$.

$$\alpha_0 := \left(\frac{1}{t+1} + \frac{1}{t+2} + \cdots + \frac{1}{(t+1)m-1}\right)^{-1}$$

Search Trees

Expected Profile $\boxed{k = \alpha \log n}$

Theorem [D.+Janson+Neininger]

- $0 < \alpha = k/\log n < \alpha_0$:

$$\boxed{\mathbb{E}\, I_{n,k} \sim (m-1)m^k}.$$

- $\alpha = k/\log n > \alpha_0$:

$$\boxed{\mathbb{E}\, I_{n,k} \sim \frac{E(\beta(\alpha))n^{\lambda_1(\beta(\alpha))-\alpha\log(\beta(\alpha))-1}}{\sqrt{2\pi(\alpha+\beta(\alpha)^2\lambda_1''(\beta(\alpha)))\log n}}}$$

for some continuous function $E(z)$

Note: $m^k = n^{\alpha \log m}$

Search Trees

Expected Profile

$$\alpha_{\mathsf{max}} := \left(\frac{1}{t+2} + \frac{1}{t+3} + \cdots + \frac{1}{(t+1)m}\right)^{-1}$$

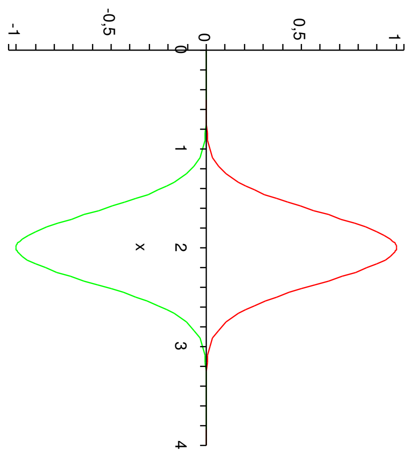
$$\boxed{\mathbb{E}\, I_{n,k} \sim \frac{n}{\sqrt{2\pi(\alpha_{\mathsf{max}} + \lambda_1''(1)) \log n}} \exp\left(-\frac{(k - \alpha_{\mathsf{max}} \log n)^2}{2(\alpha_{\mathsf{max}} + \lambda_1''(1)) \log n}\right)}$$

($\Longrightarrow$ CLT for depth D_n)

Search Trees

The average profile: $m = 2$, $t = 0$ (special case)

$$\boxed{\mathbf{E}\, I_{n,k} = \frac{n}{\sqrt{4\pi \log n}} \left(e^{-\frac{(k-2\log n)^2}{4\log n}} + \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right)\right)}.$$



Search Trees

Theorem 1 [D.+Janson+Neininger]

$m \geq 2$, $t \geq 0$ ... given integers
$(I_{n,k})_{k \geq 0}$ ... random profile

$I = \{\beta > 0 : 1 < \lambda_1(\beta^2) < 2\lambda_1(\beta) - 1\}$,
$I' = \{\beta \lambda_1'(\beta) : \beta \in I\}$

$\boxed{\beta(\alpha)\lambda_1'(\beta(\alpha)) = \alpha}$.

$\Longrightarrow$

$$\boxed{\left(\frac{I_{n,\lfloor \alpha \log n \rfloor}}{\mathbb{E}\, I_{n,\lfloor \alpha \log n \rfloor}}, \alpha \in I'\right) \xrightarrow{\text{d}} \Big(Y(\beta(\alpha)), \alpha \in I'\Big)}$$

in $D(I')$ (Skorohod topology).

Search Trees

Random analytic functions

$B \subseteq \mathbb{C}$, $(I \subseteq B)$ $\qquad$ $Y(z)$ ... random analytic function on B

$$\boxed{Y(z) \overset{\mathrm{d}}{=} zV_1^{\lambda_1(z)-1}Y^{(1)}(z) + zV_2^{\lambda_1(z)-1}Y^{(2)}(z) \cdots + zV_m^{\lambda_1(z)-1}Y^{(m)}(z)}$$

$Y^{(j)}(z)$ ... independent copies of $Y(z)$

$\mathbf{V} = (V_1, V_2, \ldots, V_m)$ ... random vector supported on the simplex $\Delta = \{(s_1, \ldots, s_m) : s_j \ge 0, s_1 + \cdots + s_m = 1\}$ with density

$$f(s_1, \ldots, s_m) = \frac{((t+1)m-1)!}{(t!)^m}(s_1 \cdots s_m)^t.$$

$\mathbf{V}, Y^{(1)}(z), \ldots, Y^{(m)}(z)$ ... independent.

Search Trees

Profile Polynomials

$$\boxed{W_n(z) := \sum_k I_{n,k} z^k}$$

$$I_{n,k} \stackrel{\mathrm{d}}{=} I^{(1)}_{V_{n,1},k-1} + I^{(2)}_{V_{n,2},k-1} + \cdots + I^{(m)}_{V_{n,m},k-1},$$

$\Longrightarrow$

$$\boxed{W_n(z) \stackrel{\mathrm{d}}{=} zW^{(1)}_{V_{n,1}}(z) + zW^{(2)}_{V_{n,2}}(z) + \cdots + zW^{(m)}_{V_{n,m}}(z) + m - 1}$$

for $n \geq m$

Search Trees

Profile Polynomials

Theorem 2 [D.+Janson+Neininger]

B ... complex region, $(1/m, \beta(\alpha_+)) \in B$,
$$\lambda_1(\beta(\alpha_+)) - \alpha_+ \log(\beta(\alpha_+)) - 1 = 0.$$

$\Longrightarrow$

$$\boxed{\left(\frac{W_n(z)}{\mathbb{E}\, W_n(z)}, z \in B\right) \xrightarrow{\mathrm{d}} (Y(z), z \in B)}$$

in $\mathcal{H}(B)$.

Remark Theorem 2 $\Longrightarrow$ Theorem 1

Search Trees

Profile Polynomials

$(I_{n,k})$ ... random profile

$\Longrightarrow W_n(z) := \sum_{k\ge 0} I_{n,k} z^k$ ... random analytic function

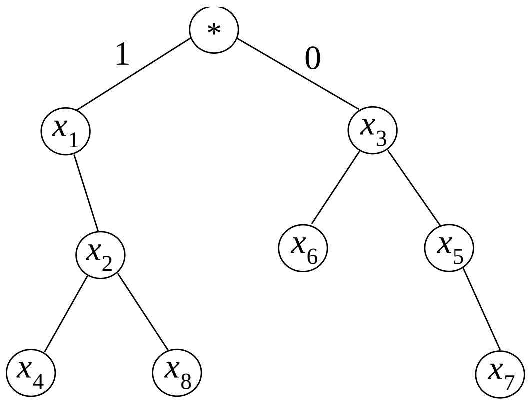
$\Longrightarrow \dfrac{W_n(z)}{\mathbb{E}\, W_n(z)}$ ... random analytic function

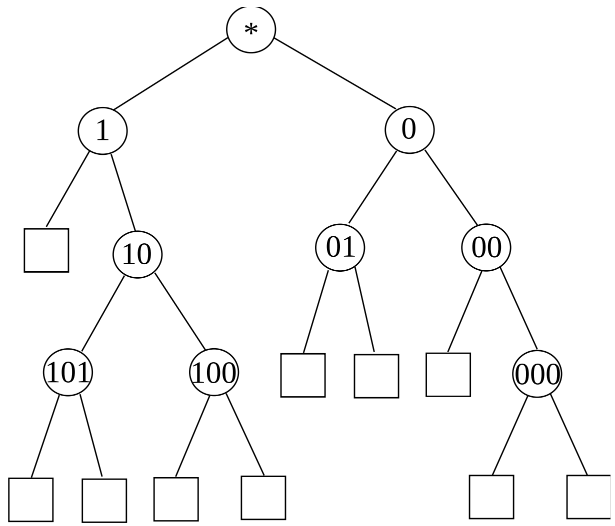
Contents

Digital Trees

Digital Search Trees

$$
\begin{aligned}
x_1 &= 110011\cdots \\
x_2 &= 100110\cdots \\
x_3 &= 010010\cdots \\
x_4 &= 101110\cdots \\
x_5 &= 000110\cdots \\
x_6 &= 010111\cdots \\
x_7 &= 000100\cdots \\
x_8 &= 100101\cdots .
\end{aligned}
$$





Digital Trees

Digital Search Trees

Bernoulli model

The input is a sequence of n independent and identically distributed random variables, each being composed of an infinite sequence of Bernoulli random variables with mean p, where $0 < p < 1$ is the probability of a 1 and $q = 1 - p$ is the probability of a 0.

Digital Trees

Profile

$B_{n,k}$ ... number of external nodes at level k after n insertions

$I_{n,k}$ ... number of internal nodes at level k after n insertions

Digital Trees

Expected Profile $p = q = \frac{1}{2}$

$$E_k(x) := \sum_{n \ge 0} \mathbb{E}\, B_{n,k} \frac{x^n}{n!}$$

$$E_k'(x) = 2e^{x/2} E_{k-1}\left(\frac{x}{2}\right)$$

$$E_0(x) = 1 \text{ and } E_k(0) = 0 \text{ for } k \ge 1$$

$$\Longrightarrow \quad \boxed{E_k(x) = 2^k e^x \sum_{m=0}^{k} \frac{(-1)^m 2^{-\binom{m}{2}}}{\gamma_m \gamma_{k-m}} e^{-x 2^{m-k}}}$$

$$\gamma_\ell = \prod_{j=1}^{\ell} \left(1 - \frac{1}{2^j}\right) \qquad (\ell > 0).$$

Digital Trees

Expected Profile

Theorem $p = q = \frac{1}{2}$

$$\Longrightarrow \quad \boxed{\mathbb{E}\, B_{n,k} = 2^k \sum_{m=0}^{k} \frac{(-1)^m 2^{-\binom{m}{2}}}{\gamma_m \gamma_{k-m}} \left(1 - \frac{1}{2^{k-m}}\right)^n}$$

$$F(z) = \sum_{m \ge 0} \frac{(-1)^m 2^{-\binom{m}{2}}}{\gamma \gamma_m} e^{-z 2^m},$$

$$\Longrightarrow \quad \boxed{\mathbb{E}\, B_{n,k} = 2^k F(n2^{-k}) + F'(n2^{-k}) + \mathcal{O}\left(n2^{-k}\right)}$$

Digital Trees

Variance of the Profile

Theorem [D.+Fuchs+Hwang+Neiniger] $p = q = \frac{1}{2}$

$$\mathrm{Var}(B_{n,k}) \begin{cases} \sim \dfrac{2^k}{Q_k}\left(1-2^{-k}\right)^n, & \text{if } n/2^k \to \infty; \\ = 2^k H(n/2^k) + \mathcal{O}(1), & \text{if } n/4^k \to 0, \end{cases}$$

$$H(x) \sim 2F(x), \qquad (x \to 0).$$

Remark

$\mathbb{E}(B_{n,k}) \to \infty$ iff $\mathrm{Var}(B_{n,k}) \to \infty$

Digital Trees

Central Limit Theorem for the Profile

Theorem [D.+Fuchs+Hwang+Neininger] $p = q = \frac{1}{2}$

If $\mathbb{E}(B_{n,k}) \to \infty$, we have

$$\boxed{\frac{B_{n,k} - \mathbb{E}(B_{n,k})}{\sqrt{\mathrm{Var}(B_{n,k})}} \xrightarrow{d} N(0,1)}.$$

Digital Trees

Theorem [D.+Szpankowski]

$\boxed{p \neq q}$, $\frac{1}{\log \frac{1}{p}} + \varepsilon \leq \frac{k}{\log n} \leq \frac{1}{\log \frac{1}{q}} - \varepsilon$ (for some $\varepsilon > 0$)

$$\boxed{\mathbb{E}\, B_{n,k} = G\left(\rho_{n,k}, \log_{p/q} p^k n\right) \frac{(p^{-\rho_{n,k}} + q^{-\rho_{n,k}})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}} \left(1 + O\left(k^{-1/2}\right)\right)}$$

$G(\rho, x)$ is a non-zero periodic function with period 1, $\rho_{n,k} = \rho(k/\log n)$.

$$\rho(\alpha) = \frac{1}{\log(p/q)} \log \frac{1 - \alpha \log(1/p)}{\alpha \log(1/q) - 1}.$$

$$\beta(\rho) = \frac{p^{-\rho} q^{-\rho} \log(p/q)^2}{(p^{-\rho} + q^{-\rho})^2},$$

Digital Trees

$$\alpha = \frac{k}{\log n}, \qquad \frac{1}{\log \frac{1}{p}} < \alpha < \frac{1}{\log \frac{1}{q}}$$

$$\boxed{\mathbb{E}\, B_{n,k} \approx \frac{n^{\kappa(\alpha)}}{\sqrt{\log n}}}$$

$$\kappa(\alpha) = \alpha \log \left(p^{-\rho(\alpha)} + q^{-\rho(\alpha)} \right) - \rho(\alpha)$$

$$\left[\rho(\alpha) = \frac{1}{\log(p/q)} \log \frac{1 - \alpha \log(1/p)}{\alpha \log(1/q) - 1} \right]$$

Digital Trees

Generating Functions for External Profile

$$P_{n,k}(u) = \mathbb{E}\, u^{B_{n,k}} = \sum_{\ell \geq 0} \mathbb{P}\{B_{n,k} = \ell\}\, u^{\ell}$$

$$\Longrightarrow \quad \boxed{P_{n+1,k+1}(u) = \sum_{\ell=0}^{n} \binom{n}{\ell} p^{\ell} q^{n-\ell} P_{n,\ell}(u) P_{n,n-\ell}(u)}$$

$$G_k(x,u) = \sum_{n \geq 0} P_{n,k}(u) \frac{x^n}{n!}$$

$$\Longrightarrow \quad \boxed{\frac{\partial}{\partial x} G_k(x,u) = G_{k-1}(px,u) G_{k-1}(qx,u)}, \qquad (k \geq 1),$$

$$G_0(x,u) = u + e^x - 1 \text{ and } G_k(0,u) = 1 \ (k \geq 1)$$

Digital Trees

Generating Functions for Internal Profile

$$G_k^{[I]}(x,u) = \sum_{n\ge 0} \mathbb{E}\, u^{I_{n,k}} \frac{x^n}{n!}$$

$$\Longrightarrow \quad \boxed{\frac{\partial}{\partial x} G_k^{[I]}(x,u) = G_{k-1}^{[I]}(px,u)G_{k-1}^{[I]}(qx,u)}, \qquad (k \ge 1),$$

$$G_0^{[I]}(x,u) = 1 + u(e^x - 1) \text{ and } G_k^{[I]}(0,u) = 1 \ (k \ge 1)$$

The analysis of the internal profile is very similar to that of the external one and will not be discussed.

Digital Trees

Generating Functions for External Profile

$$E_k(x) = \sum_{n\ge 0} \mathbb{E}\, B_{n,k} \frac{x^n}{n!} = \left[\frac{\partial G_k(x,u)}{\partial u}\right]_{u=1}$$

$$\boxed{E_k'(x) = e^{qx} E_{k-1}(px) + e^{px} E_{k-1}(qx)}$$

$E_0(x) = 1$ and $E_k(0) = 0$ $(k \ge 1)$

$$\Delta_k(x) := e^{-x} E_k(x)$$

$$\boxed{\Delta_k'(x) + \Delta_k(x) = \Delta_{k-1}(px) + \Delta_{k-1}(qx)}$$

Digital Trees

Generating Functions for External Profile

$$
\begin{aligned}
E_0(x) &= 1,\\
E_1(x) &= \frac{e^{(1-p)x}}{1-p} - \frac{1}{1-p} + \frac{e^{(1-q)x}}{1-q} - \frac{1}{1-q},\\
E_2(x) &= \frac{e^{(1-p^2)x}-1}{(1-p)(1-p^2)} - \frac{e^{(1-p)x}-1}{(1-p)^2} + \frac{e^{(1-pq)x}-1}{(1-q)(1-pq)} - \frac{e^{(1-p)x}-1}{(1-p)(1-q)}\\
&\quad + \frac{e^{(1-pq)x}-1}{(1-p)(1-pq)} - \frac{e^{(1-q)x}-1}{(1-p)(1-q)} + \frac{e^{(1-q^2)x}-1}{(1-q)(1-q^2)} - \frac{e^{(1-q)x}-1}{(1-q)^2}
\end{aligned}
$$

Digital Trees

Mellin transform for $\boxed{\Delta_k(x) := e^{-x} E_k(x)}$

$$\Delta_k^*(s) = \int_0^\infty \Delta_k(x) x^{s-1}\, dx.$$

$$\boxed{\Delta_k^*(s) - (s-1)\Delta_k^*(s-1) = p^{-s}\Delta_{k-1}^*(s) + q^{-s}\Delta_{k-1}^*(s)}$$

Inverse Mellin transform

$$\Delta_k(x) = \frac{1}{2\pi i}\int_{\rho-i\infty}^{\rho+i\infty} \Delta_k^*(s) x^{-s}\, ds$$

Digital Trees

"Simplified version"

Original version

$$E_k'(x) = e^{qx}E_{k-1}(px) + e^{px}E_{k-1}(qx)$$

"simplified" to

$$\boxed{E_k(x) = e^{qx}E_{k-1}(px) + e^{px}E_{k-1}(qx)}$$

$\Delta_k(x) = e^{-x}E_k(x),\ \Delta_k^*(s) = \int_0^\infty \Delta_k(x)x^{s-1}\,dx$

$$\Delta_k^*(s) = (p^{-s} + q^{-s})\Delta_{k-1}^*(s)$$

$$\Longrightarrow \quad \boxed{\Delta_k^*(s) = \Gamma(s)(p^{-s} + q^{-s})^k}$$

Digital Trees

"Simplified version"

Inverse Mellin transform for $x = n$

$$\begin{aligned}\Delta_k(n) &= \frac{1}{2\pi i}\int_{\rho-i\infty}^{\rho+i\infty} \Delta_k^*(s) n^{-s}\, ds \\ &= \frac{1}{2\pi i}\int_{\rho-i\infty}^{\rho+i\infty} \Gamma(s)\,(p^{-s}+q^{-s})^k n^{-s}\, ds\end{aligned}$$

$$(p^{-s}+q^{-s})^k n^{-s} = e^{k\log(p^{-s}+q^{-s}) - s\log n}$$

Saddle point:

$$\frac{\partial}{\partial s}\Big(k\log(p^{-s}+q^{-s}) - s\log n\Big) = 0$$

Digital Trees

"Simplified version"

... infinitely many saddle points for on the line $\Re(s) = \rho_{n,k} = \rho(k/\log n)$:

$$\Longrightarrow \quad \boxed{s_j = \rho_{n,k} + \frac{2\pi i j}{\log \frac{p}{q}}}$$

... with usual saddle point analysis:

$$\Longrightarrow \quad \boxed{\Delta_k(n) \sim G\left(\rho_{n,k}, \log_{p/q} p^k n\right) \frac{(p^{-\rho_{n,k}} + q^{-\rho_{n,k}})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}}}$$

Recall: $\mathbb{E}\, B_{n,k} \sim \Delta_k(n)$ (by the Poisson heuristics)

Digital Trees

Mellin transform (for the original problem)

$$\Delta_k^*(s) = \Gamma(s) F_k(s),$$

$$\boxed{F_k(s) - F_k(s-1) = (p^{-s} + q^{-s}) F_{k-1}(s)}$$

$$F_0(x) = 1,$$

$$F_1(x) = \frac{p^{-s}}{1-p} - \frac{1}{1-p} + \frac{q^{-s}}{1-q} - \frac{1}{1-q},$$

$$F_2(x) = \frac{p^{-2s}-1}{(1-p)(1-p^2)} - \frac{p^{-s}-1}{(1-p)^2} + \frac{p^{-s}q^{-s}-1}{(1-q)(1-pq)} - \frac{p^{-s}-1}{(1-p)(1-q)}$$
$$+ \frac{p^{-s}q^{-s}-1}{(1-p)(1-pq)} - \frac{q^{-s}-1}{(1-p)(1-q)} + \frac{q^{-2s}-1}{(1-q)(1-q^2)} - \frac{q^{-s}-1}{(1-q)^2}$$

Digital Trees

Remark.

The Mellin transform $\Delta_k^*(s)$ exists for $\Re(s) > -k$

$$\Delta_k^*(s) = \Gamma(s) F_k(s)$$

$$\Longrightarrow \quad \boxed{F_k(0) = 0} \quad (k > 0)$$

Digital Trees

A linear operator

Set $\boxed{T(s) = p^{-s} + q^{-s}}$ and define

$$\boxed{\mathbf{A}[f](s) = \sum_{j\ge 0} f(s-j)T(s-j)}$$

Furthermore set $\boxed{R_k(s) = \mathbf{A}^k[1](s)}$:

$$
\begin{aligned}
R_0(s) &= 1, \\
R_1(s) &= \frac{p^{-s}}{1-p} + \frac{q^{-s}}{1-q}, \\
R_2(s) &= \frac{p^{-2s}}{(1-p)(1-p^2)} + \frac{p^{-s}q^{-s}}{(1-p)(1-pq)} \\
&\quad + \frac{p^{-s}q^{-s}}{(1-q)(1-pq)} + \frac{q^{-2s}}{(1-q)(1-q^2)}.
\end{aligned}
$$

Digital Trees

Lemma 1

$$\boxed{F_k(s) = \mathbf{A}[F_{k-1}](s) - \mathbf{A}[F_{k-1}](0)}$$

$$\boxed{\sum_{k\geq 0} F_k(s) w^k = \frac{\sum_{\ell\geq 0} R_\ell(s) w^\ell}{\sum_{\ell\geq 0} R_\ell(0) w^\ell}}$$

Digital Trees

Proof

One has to show

$$\sum_{\ell=0}^{k} F_\ell(s) R_{k-\ell}(0) = R_k(s), \qquad (k \geq 0),$$

or equivalently

$$F_k(s) = R_k(s) - \sum_{\ell=0}^{k-1} F_\ell(s) R_{k-\ell}(0), \qquad (k \geq 0).$$

Digital Trees

Proof

The case $k = 0$ is obvious.

General induction step:

$$\begin{aligned} F_{k+1}(s) &= \mathbf{A}[F_k](s) - \mathbf{A}[F_k](0) \\ &= \mathbf{A}[R_k](s) - \mathbf{A}[R_k](0) \\ &\quad - \sum_{\ell=0}^{k-1} (\mathbf{A}[F_\ell](s) - \mathbf{A}[F_\ell](0)) R_{k-\ell}(0) \\ &= R_{k+1}(s) - R_{k+1}(0) - \sum_{\ell=0}^{k-1} F_{\ell+1}(s) R_{k-\ell}(0) \\ &= R_{k+1}(s) - \sum_{\ell=0}^{k} F_\ell(s) R_{k+1-\ell}(0). \end{aligned}$$

Digital Trees

$$g(w,s) = \sum_{\ell \geq 0} R_\ell(s) w^\ell$$

$$\boxed{g(w,s) = 1 + w\mathbf{A}[g(w,\cdot)](s) = 1 + \sum_{j \geq 0} g(w,s-j)T(s-j)}$$

Digital Trees

Lemma 2

$$\boxed{g(w,s) = \frac{h(w,s)}{1 - wT(s)}}$$

with

$$h(w,s) = 1 + \sum_{j\ge 1} h(w,s-j)\frac{wT(s-j)}{1-wT(s-j)}.$$

which is analytic for $w = T(s)$.

Digital Trees

Corollary

$$\boxed{F_k(s) = f(s)T(s)^k \left(1 + O\left(e^{-\eta k}\right)\right)}$$

$\longrightarrow F_k(s)$ behaves as in the "simplified" case. Hence, the inverse Mellin transform (with infinitely many saddle points) works and the Poisson heuristics applies as well. *QED*

Thank You!