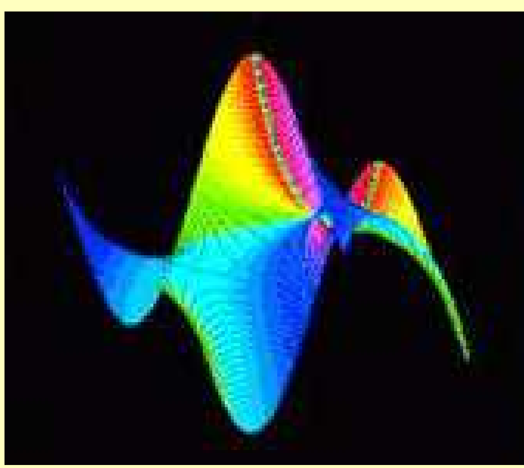
PROFILE OF RANDOM RECURSIVE TREES AND RANDOM BINARY SEARCH TREES

Hsien-Kuei Hwang

Academia Sinica, Taiwan

(joint with M. Drmota, M. Fuchs, R. Neininger)

April 26, 2004




PROFILE OF TREES

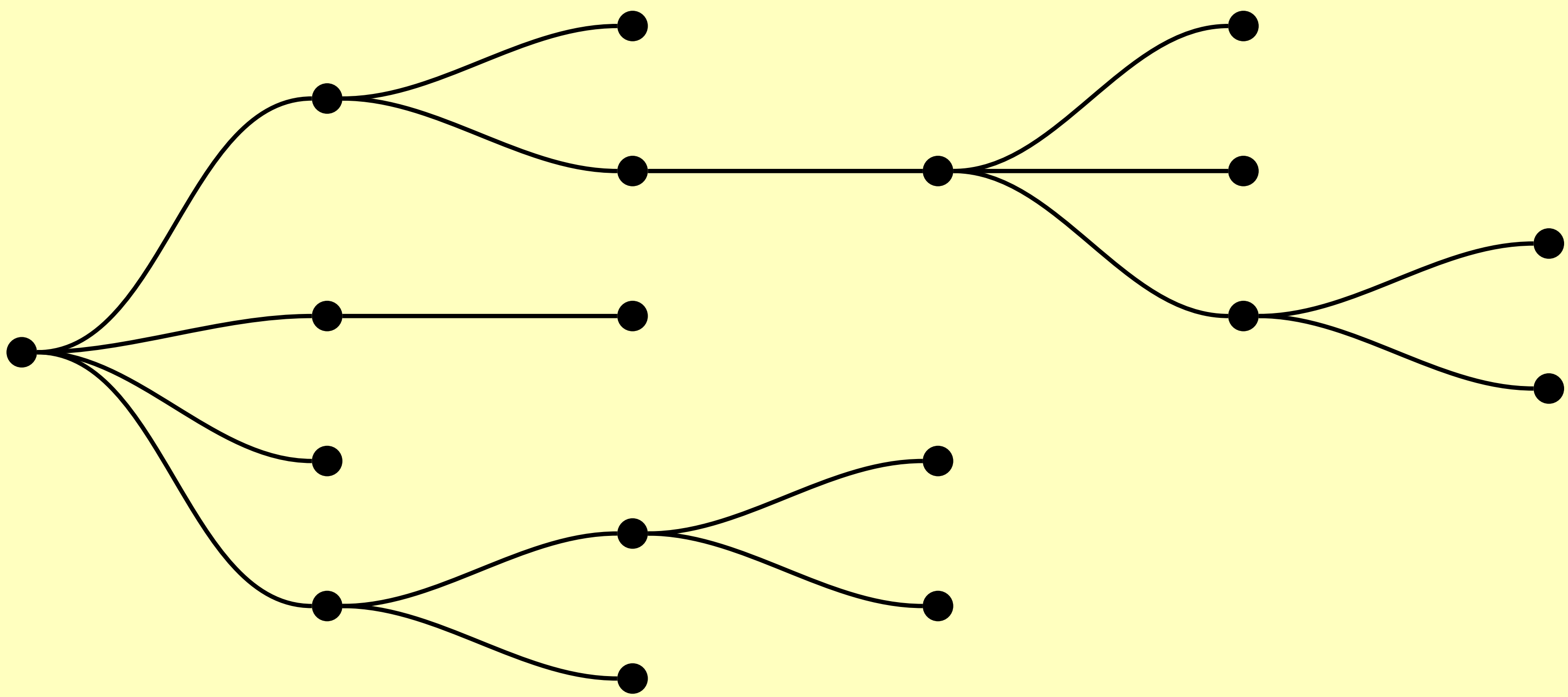


Figure 1: Profile $= \{$1,4,5,3,3,2$\}$

PROFILE: MOTIVATIONS

➠ **fine, informative shape characteristic**

PROFILE: MOTIVATIONS

- fine, informative shape characteristic
- descendants by generation in branching process

PROFILE: MOTIVATIONS

- fine, informative shape characteristic
- descendants by generation in branching process
- related to path length, depth, height, width, etc.

PROFILE: MOTIVATIONS

- fine, informative shape characteristic
- descendants by generation in branching process
- related to path length, depth, height, width, etc.
- breadth-first search

PROFILE: MOTIVATIONS

- fine, informative shape characteristic
- descendants by generation in branching process
- related to path length, depth, height, width, etc.
- breadth-first search
- compression algorithms (Jacquet, Szpankowski)

PROFILE: MOTIVATIONS

- fine, informative shape characteristic
- descendants by generation in branching process
- related to path length, depth, height, width, etc.
- breadth-first search
- compression algorithms (Jacquet, Szpankowski)
- generation of random trees (Devroye, Robson)

PROFILE: MOTIVATIONS

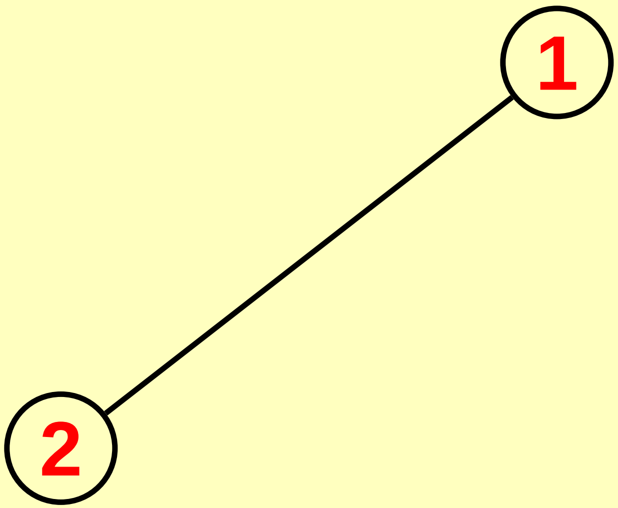
- fine, informative shape characteristic
- descendants by generation in branching process
- related to path length, depth, height, width, etc.
- breadth-first search
- compression algorithms (Jacquet, Szpankowski)
- generation of random trees (Devroye, Robson)
- level-wise analysis of quicksort (Chern, H.)

PROFILE OF RANDOM TREES: A RICH SOURCE OF INTRIGUING PHENOMENA

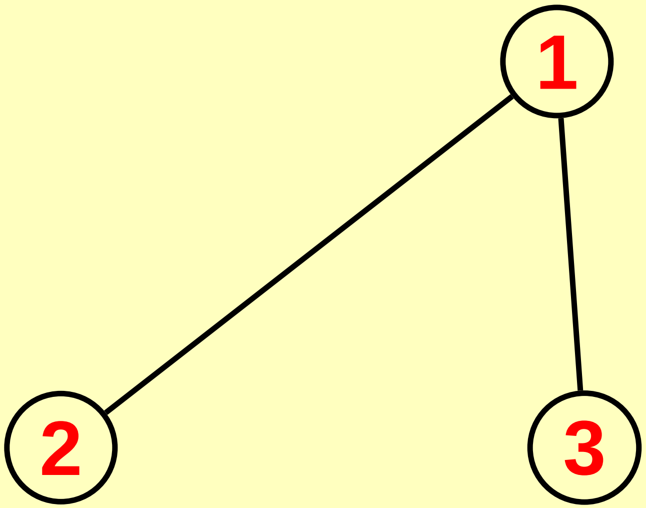
RANDOM RECURSIVE TREES

1

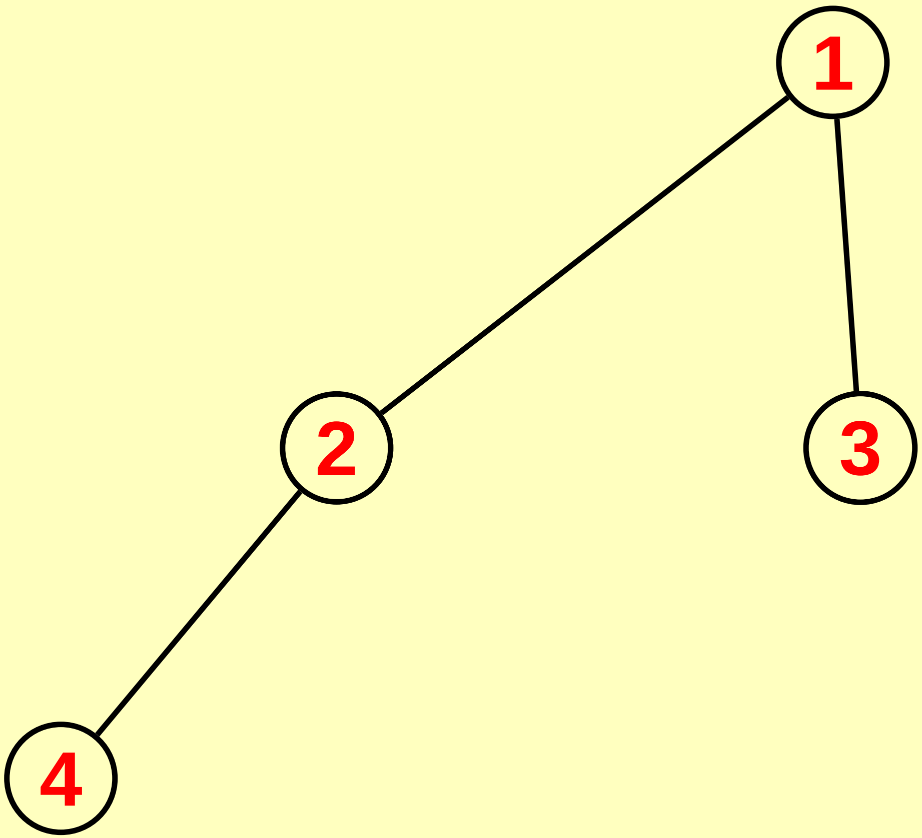
RANDOM RECURSIVE TREES



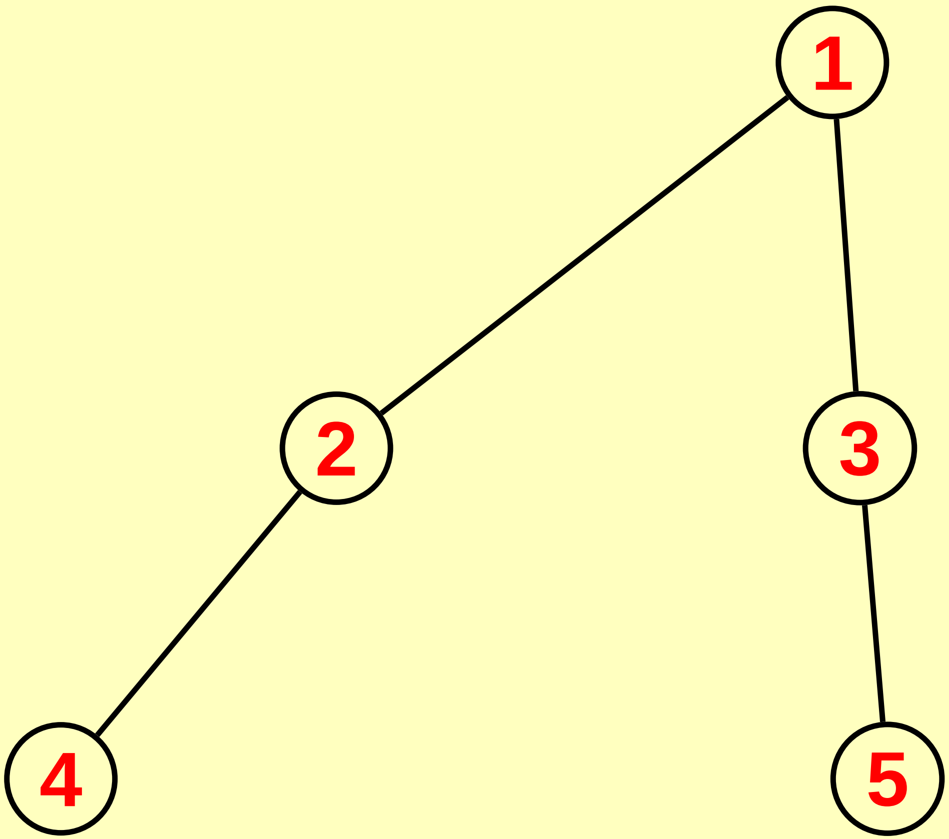
RANDOM RECURSIVE TREES



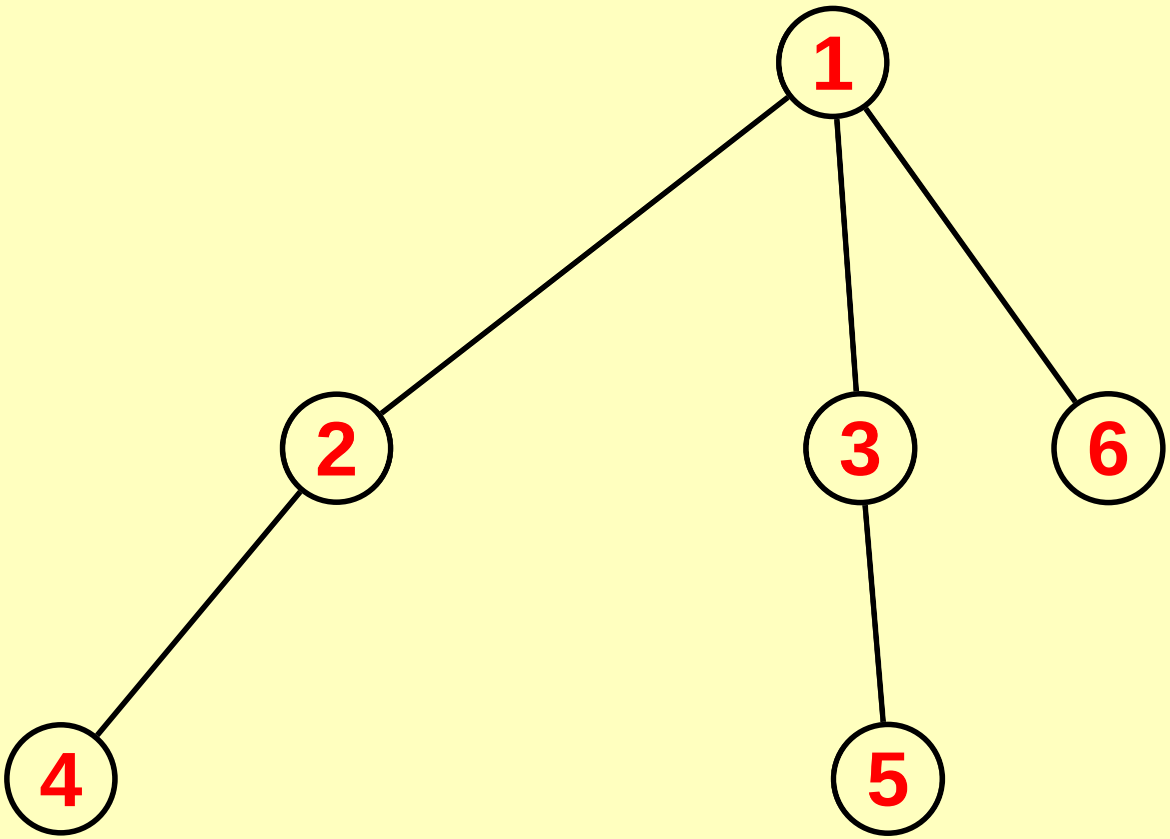
RANDOM RECURSIVE TREES



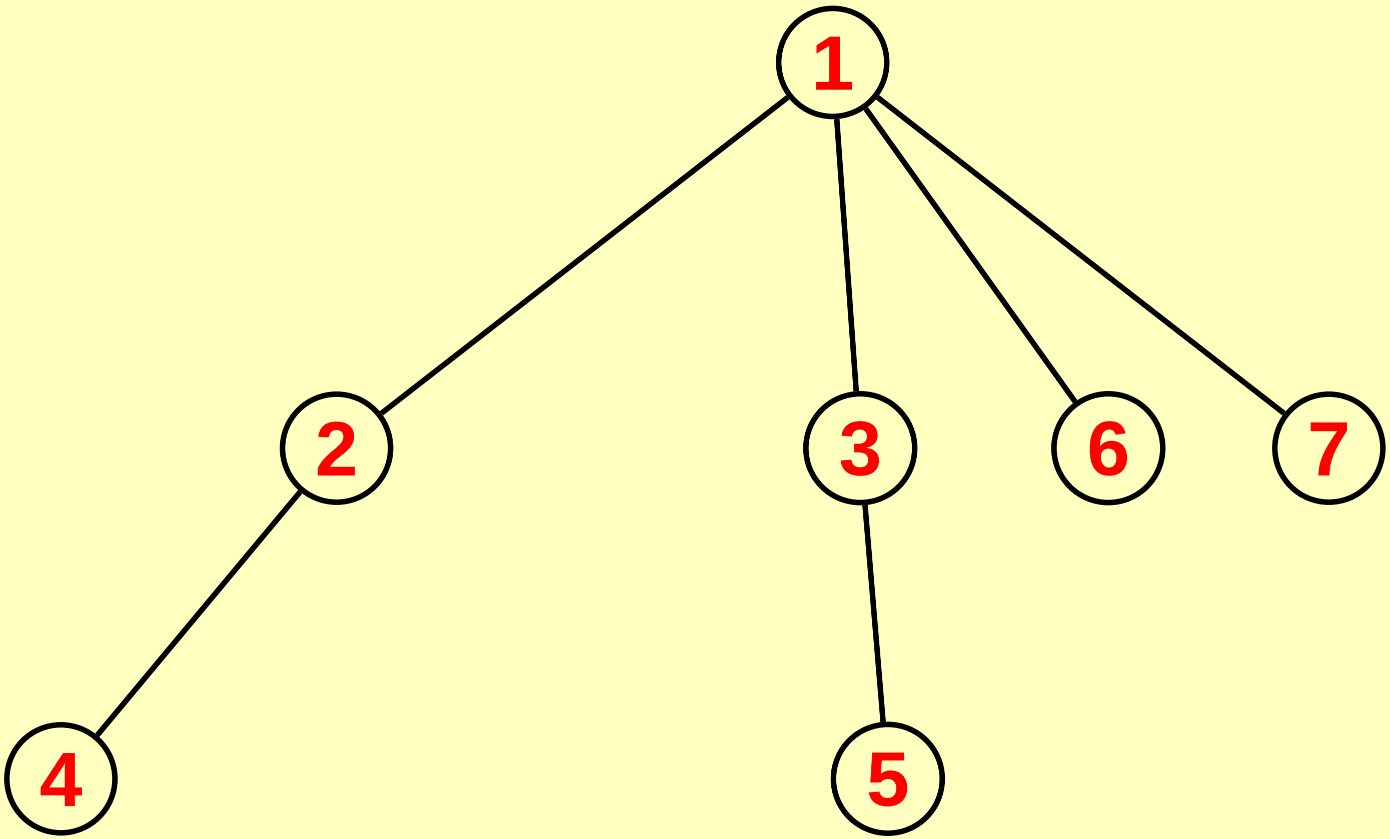
RANDOM RECURSIVE TREES



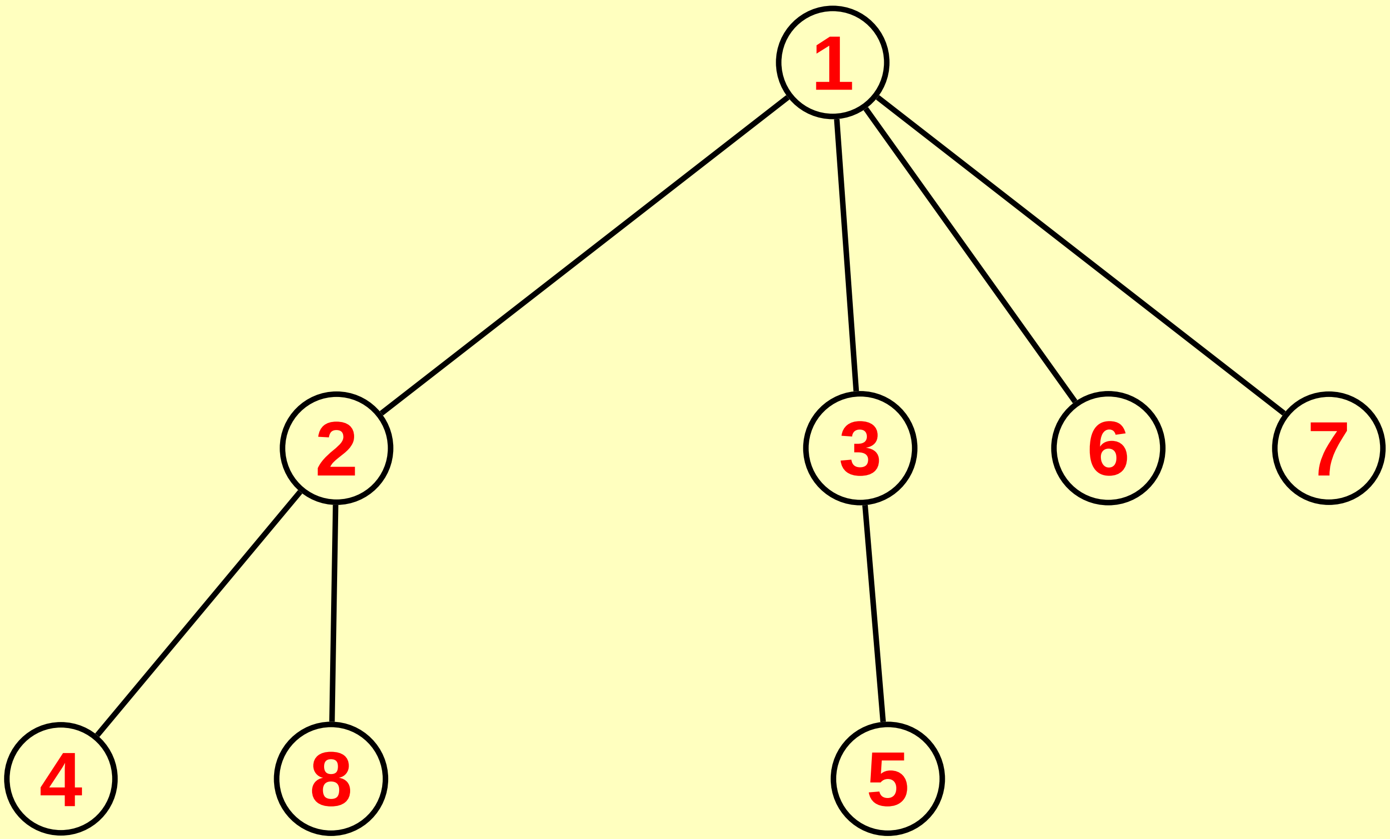
RANDOM RECURSIVE TREES



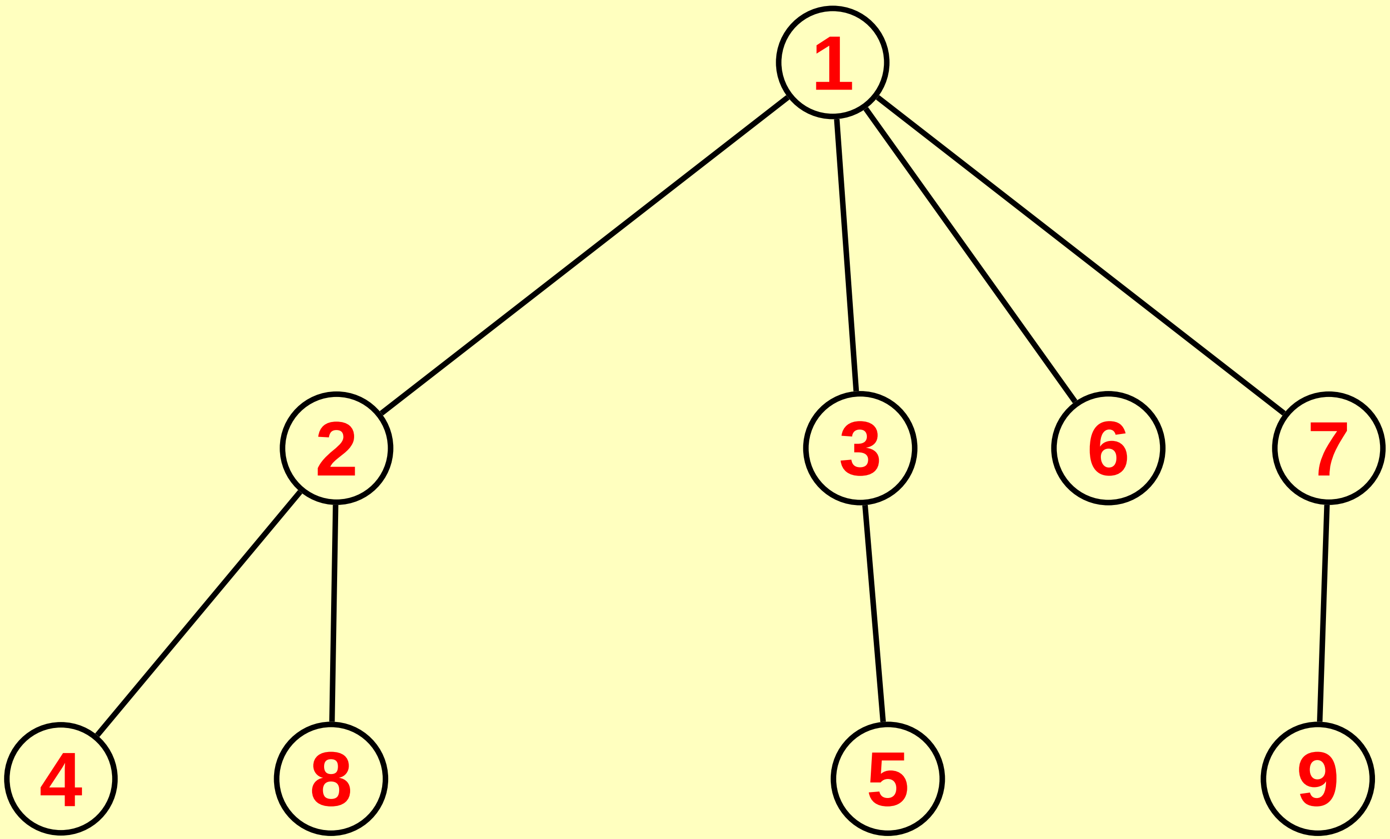
RANDOM RECURSIVE TREES



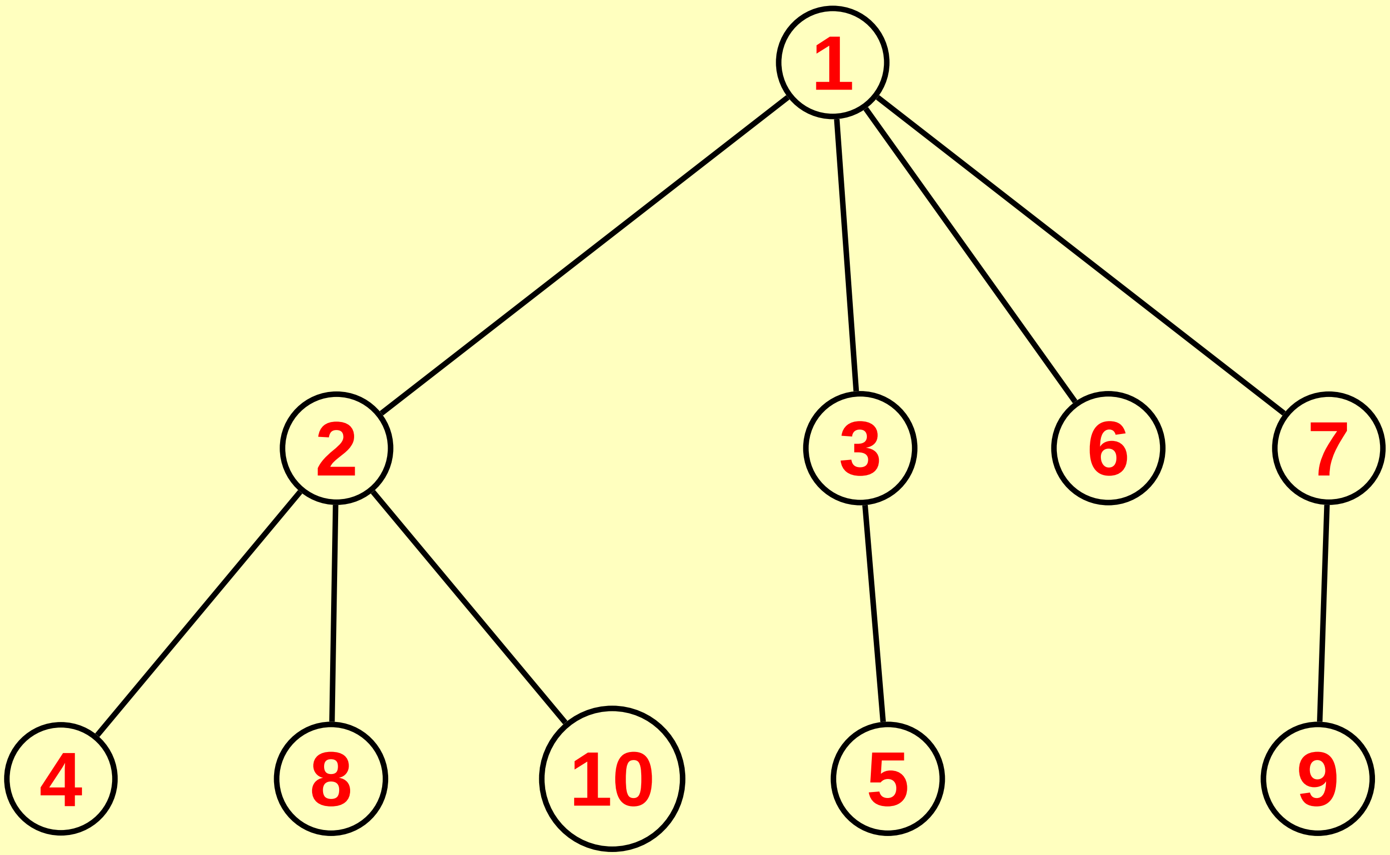
RANDOM RECURSIVE TREES



RANDOM RECURSIVE TREES



RANDOM RECURSIVE TREES



USEFULNESS OF RANDOM RECURSIVE TREES

Simple probability models for

- **system generation (Na, Rapoport, 1970)**
- **spread of contamination of organisms (Meir, Moon, 1974)**
- **pyramid scheme (Bhattacharya, Gastwirth, 1984)**
- **stemma construction of philology (Najock, Heyde, 1982)**
- **Internet interface map (Janic et al., 2002)**
- **stochastic growth of networks (Chan et al., 2003).**
- **Internet (van Mieghem et al., 2001; Devroye, McDiarmid, Reed, 2002)**
- **statistical physics (Tetzlaff, 2002)**

USEFULNESS OF RANDOM RECURSIVE TREES

Simple probability models for

- system generation (Na, Rapoport, 1970)
- spread of contamination of organisms (Meir, Moon, 1974)
- pyramid scheme (Bhattacharya, Gastwirth, 1984)
- stemma construction of philology (Najock, Heyde, 1982)
- Internet interface map (Janic et al., 2002)
- stochastic growth of networks (Chan et al., 2003).
- Internet (van Mieghem et al., 2001; Devroye, McDiarmid, Reed, 2002)
- statistical physics (Tetzlaff, 2002)

They also appeared in Hopf algebra under the name of "heap-ordered trees"; see Grossman, Larsen (1989).

PROFILE OF RANDOM RECURSIVE TREES

$X_{n,k} := \#$ **of nodes at distance** k **from the root in a random recursive tree of** n **keys.**

PROFILE OF RANDOM RECURSIVE TREES

$X_{n,k} := \#$ **of nodes at distance** k **from the root in a random recursive tree of** n **keys.**

Main questions: **mean, variance, limit distribution of** $X_{n,k}$ **for all possible values of** k**?**

PROFILE OF RANDOM RECURSIVE TREES

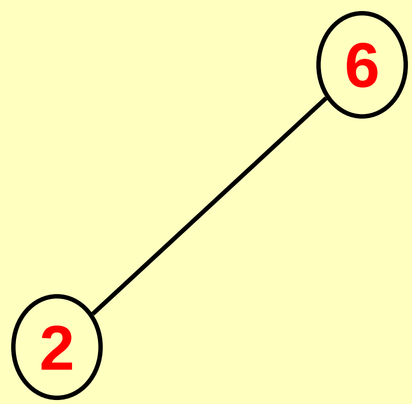
$X_{n,k} := \#$ **of nodes at distance** k **from the root in a random recursive tree of** n **keys.**

Main questions: **mean, variance, limit distribution of** $X_{n,k}$ **for all possible values of** k**?**

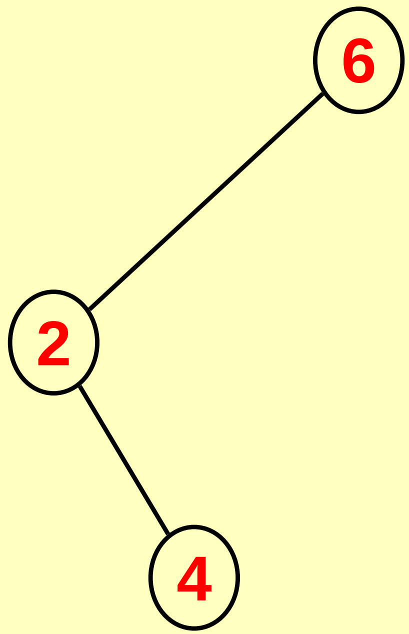
Main range: $k \le K \log n$

THE BINARY SEARCH TREE CONSTRUCTED FROM $\{$6,2,4,8,7,1,5,3,10,9$\}$

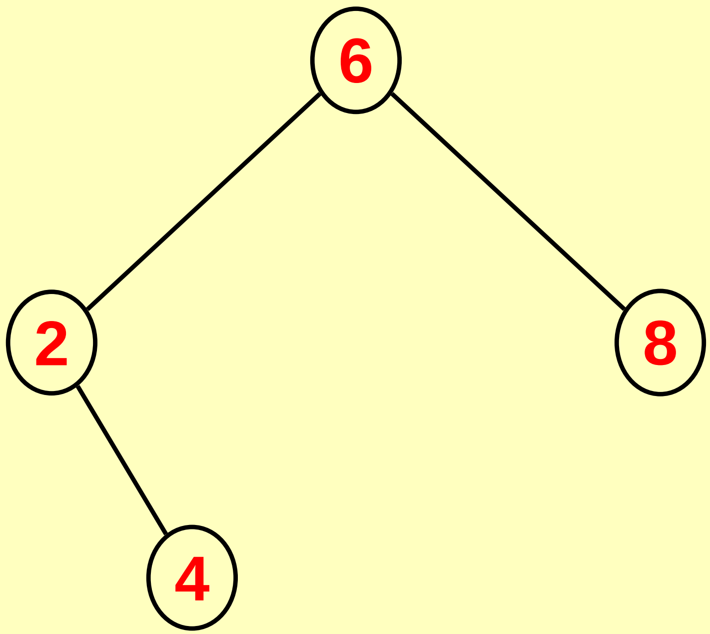
THE BINARY SEARCH TREE CONSTRUCTED FROM {6,2,4,8,7,1,5,3,10,9}



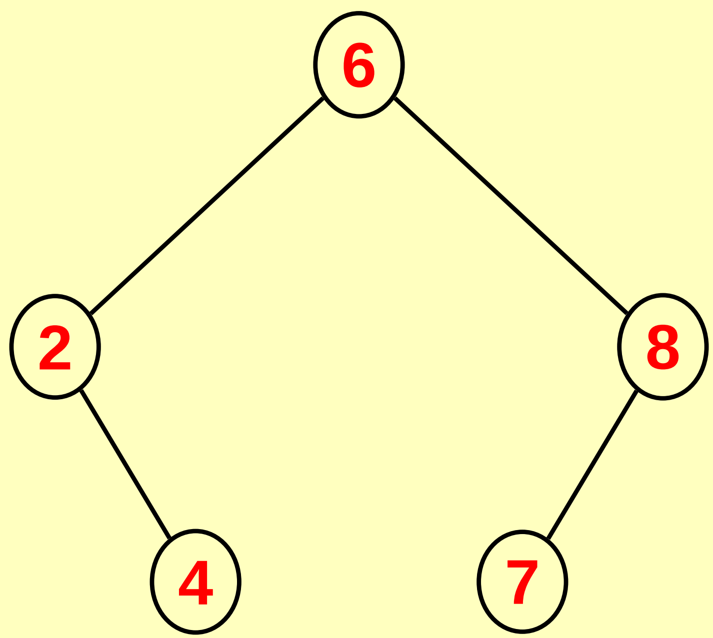
THE BINARY SEARCH TREE CONSTRUCTED FROM $\{$6,2,4,8,7,1,5,3,10,9$\}$



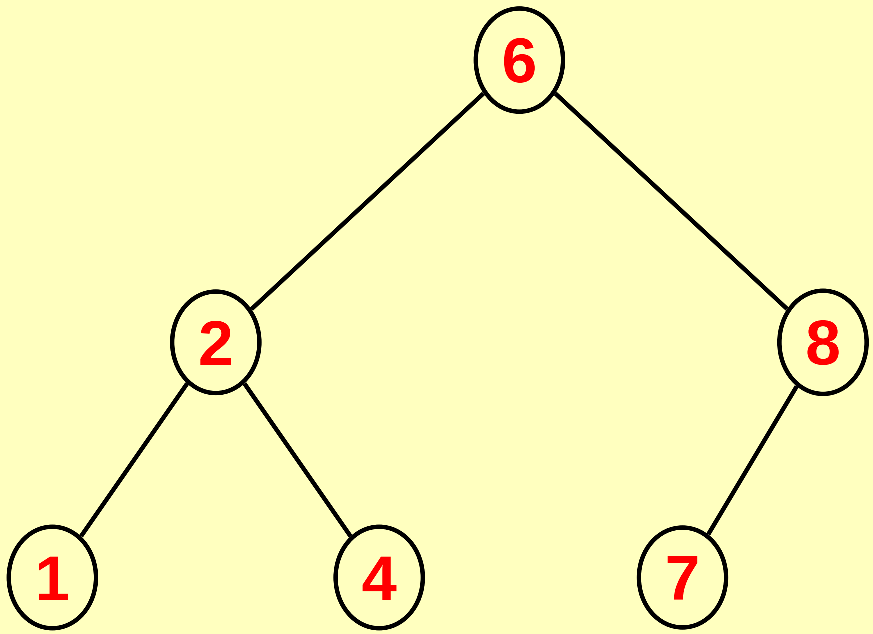
THE BINARY SEARCH TREE CONSTRUCTED FROM $\{$6,2,4,8,7,1,5,3,10,9$\}$



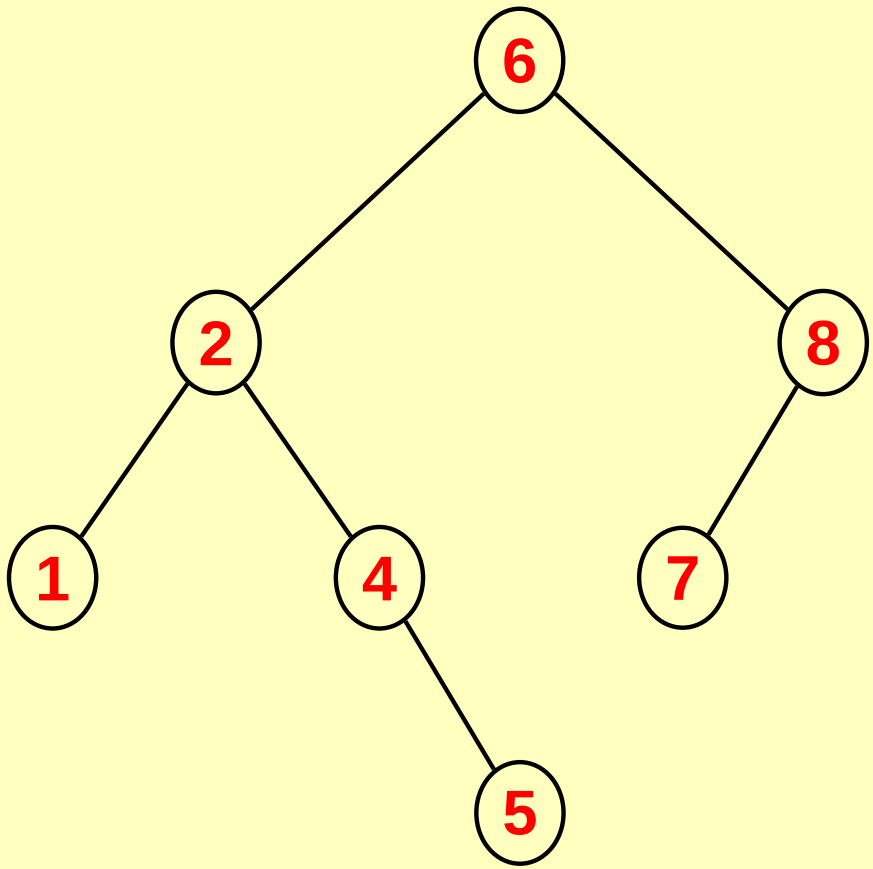
THE BINARY SEARCH TREE CONSTRUCTED FROM $\{$6,2,4,8,7,1,5,3,10,9$\}$



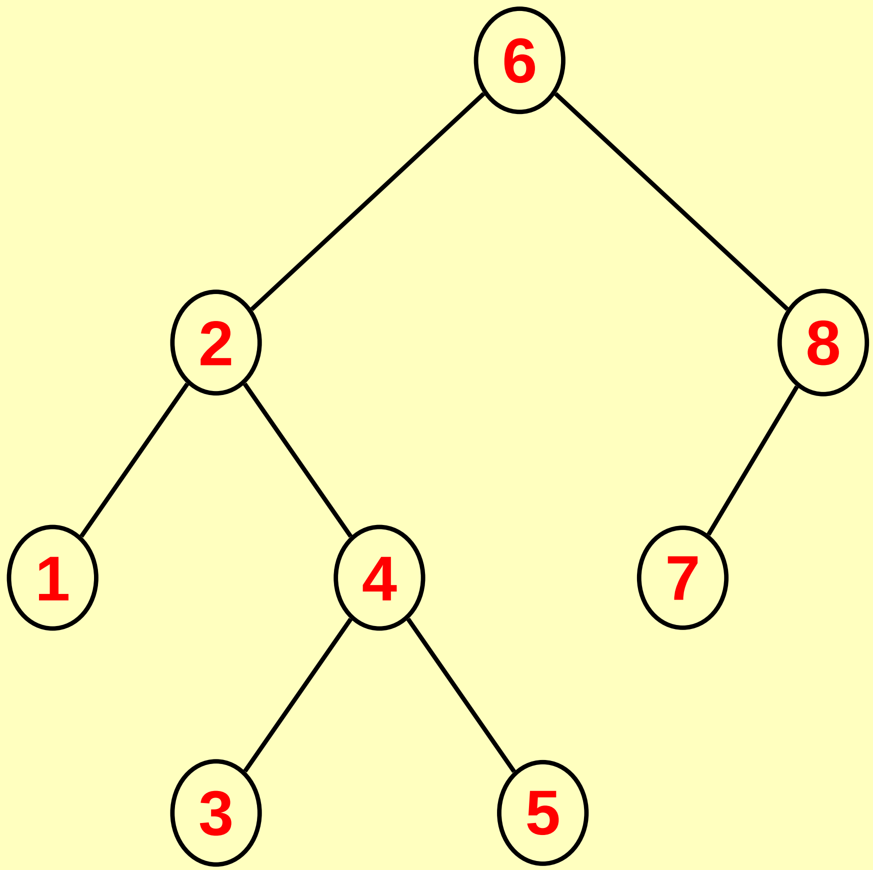
THE BINARY SEARCH TREE CONSTRUCTED FROM $\{$6,2,4,8,7,1,5,3,10,9$\}$



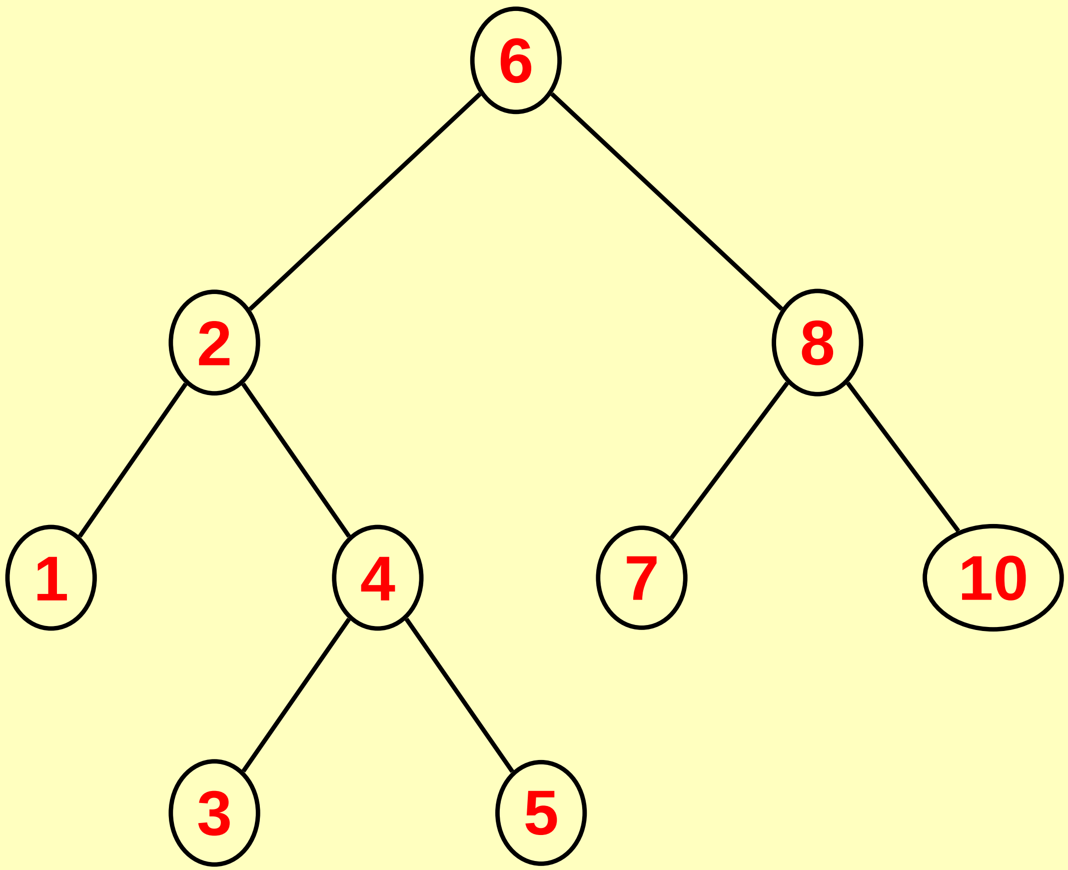
THE BINARY SEARCH TREE CONSTRUCTED FROM $\{$6,2,4,8,7,1,5,3,10,9$\}$



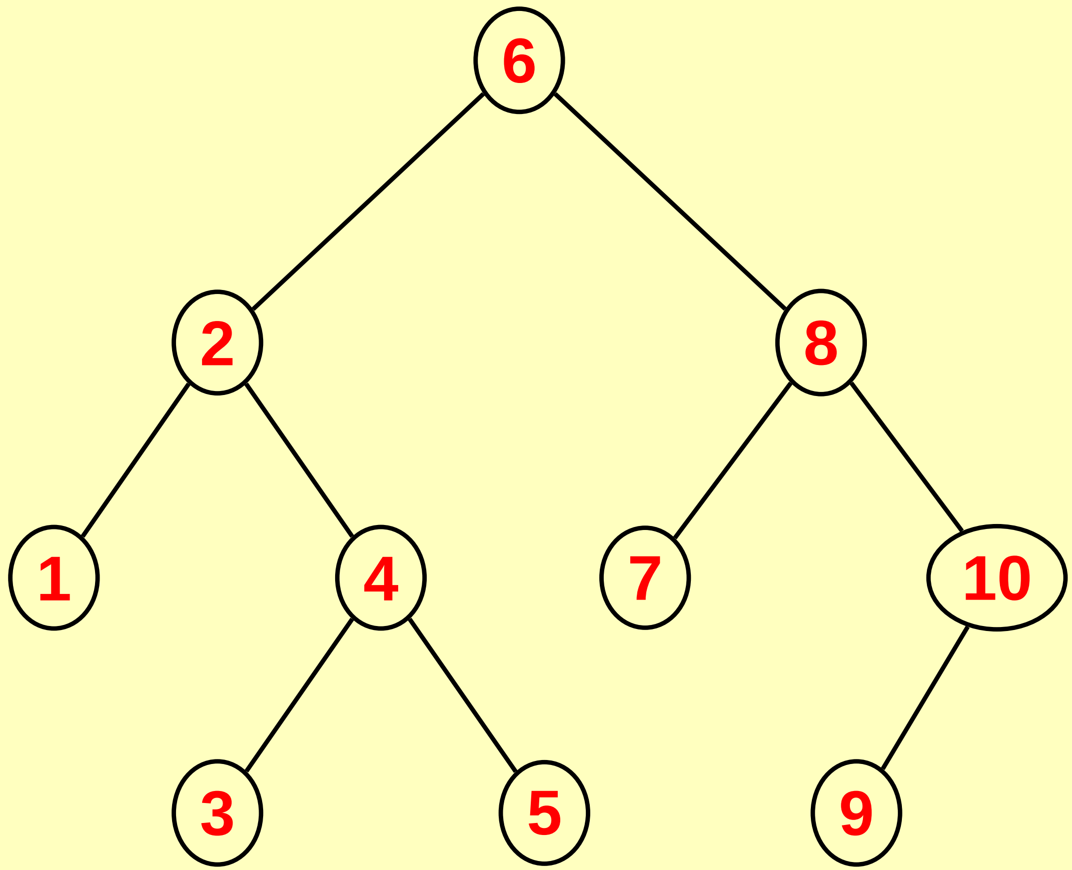
THE BINARY SEARCH TREE CONSTRUCTED FROM $\{$6,2,4,8,7,1,5,3,10,9$\}$



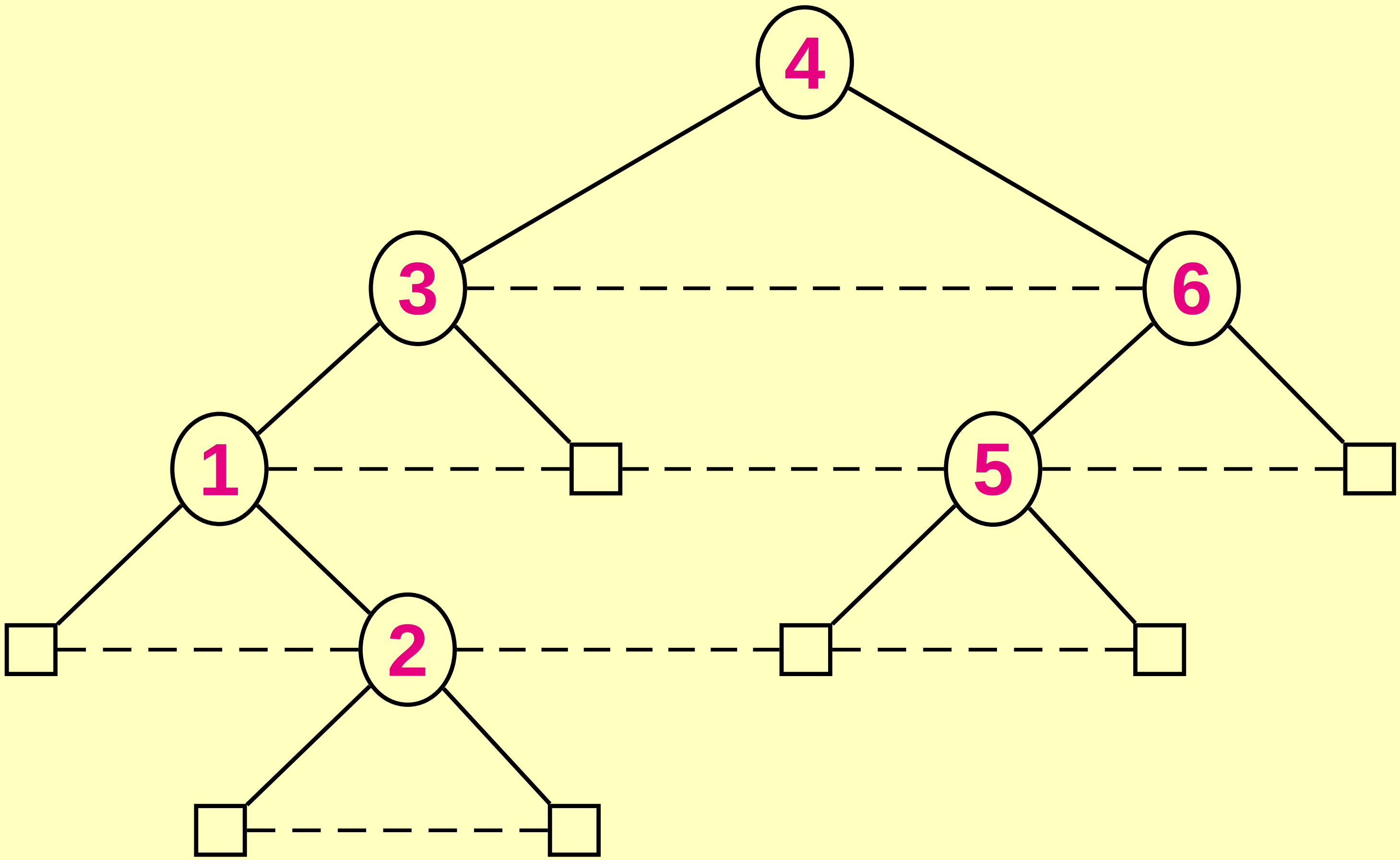
THE BINARY SEARCH TREE CONSTRUCTED FROM $\{$6,2,4,8,7,1,5,3,10,9$\}$



THE BINARY SEARCH TREE CONSTRUCTED FROM $\{$6,2,4,8,7,1,5,3,10,9$\}$



INTERNAL NODES AND EXTERNAL NODES



PROFILE OF RANDOM BINARY SEARCH TREES

The model: Assume that all permutations of n elements are equally likely. Construct the BST from a random permutation.

$$\begin{cases} Y_{n,k} & := \text{the number of external nodes at level } k \\ Z_{n,k} & := \text{the number of internal nodes at level } k \end{cases}$$

PROFILE OF RANDOM BINARY SEARCH TREES

The model: **Assume that all permutations of n elements are equally likely. Construct the BST from a random permutation.**

$$\begin{cases} Y_{n,k} & := \text{the number of external nodes at level } k \\ Z_{n,k} & := \text{the number of internal nodes at level } k \end{cases}$$

Main question: **probabilistic properties of $Y_{n,k}$ and $Z_{n,k}$?**

A CONNECTION

$X_{n,k} \stackrel{d}{=} \#$ **of nodes at** k **left branches away from the root in random BSTs of** $n-1$ **keys**

A CONNECTION

$X_{n,k} \stackrel{d}{=} \#$ of nodes at k left branches away from the root in random BSTs of $n-1$ keys by (i) the usual transformation from a multiway tree to a binary tree (first branch $\longrightarrow$ left, sibling $\longrightarrow$ right)

A CONNECTION

$X_{n,k} \stackrel{d}{=} \#$ **of nodes at** k **left branches away from the root in random BSTs of** $n-1$ **keys by** (i) **the usual transformation from a multiway tree to a binary tree (first branch** $\longrightarrow$ **left, sibling** $\longrightarrow$ **right) and** (ii) **the bijection between binary increasing trees and binary search trees.**

A CONNECTION

$X_{n,k} \overset{d}{=} \#$ **of nodes at** k **left branches away from the root in random BSTs of** $n-1$ **keys by** (i) **the usual transformation from a multiway tree to a binary tree (first branch** $\longrightarrow$ **left, sibling** $\longrightarrow$ **right) and** (ii) **the bijection between binary increasing trees and binary search trees.**

It suffices to look at BSTs.

A CONNECTION

$X_{n,k} \stackrel{d}{=} \#$ of nodes at k left branches away from the root in random BSTs of $n-1$ keys by (i) the usual transformation from a multiway tree to a binary tree (first branch $\longrightarrow$ left, sibling $\longrightarrow$ right) and (ii) the bijection between binary increasing trees and binary search trees.

It suffices to look at BSTs.

But (i) it's much simpler to start from the better structured recursive trees; (ii) behaviors not identical in all ranges.

SUMMARY OF MAIN PHENOMENA

Write throughout $\alpha_{n,k} = \dfrac{k}{\log n}$ **and** $\lim_n \alpha_{n,k} = \alpha$**.**

SUMMARY OF MAIN PHENOMENA

Write throughout $\alpha_{n,k} = \dfrac{k}{\log n}$ **and** $\lim_n \alpha_{n,k} = \alpha$**.**

➠ $\mathbb{E}(X_{n,k})$ **unimodal, but** $\mathbb{V}(X_{n,k})$ **bimodal**

SUMMARY OF MAIN PHENOMENA

Write throughout $\alpha_{n,k} = \dfrac{k}{\log n}$ **and** $\lim_n \alpha_{n,k} = \alpha$.

- $\mathbb{E}(X_{n,k})$ **unimodal, but** $\mathbb{V}(X_{n,k})$ **bimodal**
- **If** $0 \le \alpha < e$, **then** $\dfrac{X_{n,k}}{\mathbb{E}(X_{n,k})} \xrightarrow{d} X_\alpha$.

SUMMARY OF MAIN PHENOMENA

Write throughout $\alpha_{n,k} = \dfrac{k}{\log n}$ **and** $\lim_n \alpha_{n,k} = \alpha$**.**

- $\mathbb{E}(X_{n,k})$ **unimodal, but** $\mathbb{V}(X_{n,k})$ **bimodal**
- **If** $0 \le \alpha < e$**, then** $\dfrac{X_{n,k}}{\mathbb{E}(X_{n,k})} \xrightarrow{d} X_\alpha$**.**
- **If** $0 \le \alpha \le 1$**, then** $\dfrac{X_{n,k}}{\mathbb{E}(X_{n,k})} \xrightarrow{m} X_\alpha$**.**

SUMMARY OF MAIN PHENOMENA

- **For** $k = o(\log n)$, $\dfrac{X_{n,k} - \mathbb{E}(X_{n,k})}{\sqrt{\mathbb{V}(X_{n,k})}} \xrightarrow{m} N(0,1)$.

SUMMARY OF MAIN PHENOMENA

- **For** $k = o(\log n)$, $\dfrac{X_{n,k} - \mathbb{E}(X_{n,k})}{\sqrt{\mathbb{V}(X_{n,k})}} \xrightarrow{m} N(0,1)$.
- **If** $k = \log n + o(\log n)$ **and** $|k - \log n| \to \infty$, **then** $\dfrac{X_{n,k} - \mathbb{E}(X_{n,k})}{\sqrt{\mathbb{V}(X_{n,k})}} \xrightarrow{m} X_1'$.

SUMMARY OF MAIN PHENOMENA

- **For** $k = o(\log n)$, $\dfrac{X_{n,k} - \mathbb{E}(X_{n,k})}{\sqrt{\mathbb{V}(X_{n,k})}} \xrightarrow{m} N(0,1)$.
- **If** $k = \log n + o(\log n)$ **and** $|k - \log n| \to \infty$, **then** $\dfrac{X_{n,k} - \mathbb{E}(X_{n,k})}{\sqrt{\mathbb{V}(X_{n,k})}} \xrightarrow{m} X_1'$.
- **For** $k = \log n + O(1)$, **the limit law of** $\dfrac{X_{n,k} - \mathbb{E}(X_{n,k})}{\sqrt{\mathbb{V}(X_{n,k})}}$ **does not exist.**

RECURRENCE OF $X_{n,k}$ ☯

$$X_{n,k} \stackrel{d}{=} X_{\mathsf{uniform}[1,n-1],k-1} + X^*_{n-\mathsf{uniform}[1,n-1],k}$$

RECURRENCE OF $X_{n,k}$

$$X_{n,k} \stackrel{d}{=} X_{\text{uniform}[1,n-1],k-1} + X^*_{n-\text{uniform}[1,n-1],k}$$

Three proofs: (i) **bijection conditioned on the size of the subtree rooted at 2;** (ii) **above-mentioned transformation;** (iii) **algebraic**

$$X_{n,k} \stackrel{d}{=} \sum_{s\geq 1} \frac{1}{s!} \sum_{j_1+\cdots+j_s=n-1} \underbrace{\binom{n-1}{j_1,\ldots,j_s}\frac{(j-1)!\cdots(j_s-1)!}{(n-1)!}}_{\mathbb{P}(\text{s subtrees have sizes } j_1,\ldots,j_s)}$$
$$\times\left(X_{j_1,k-1}+\cdots+X_{j_s,k-1}\right).$$

EXPECTED VALUE OF $X_{n,k}$

Meir, Moon (1978) (implicit): for $0 \le k < n$

$$\mu_{n,k} := \mathbb{E}(X_{n,k}) = \frac{\text{Stirling1}(n, k+1)}{(n-1)!};$$

also in Moon (1974) and Dondajewski, Szymański (1982).

EXPECTED VALUE OF $X_{n,k}$

Meir, Moon (1978) (implicit): for $0 \le k < n$

$$\mu_{n,k} := \mathbb{E}(X_{n,k}) = \frac{\text{Stirling1}(n, k+1)}{(n-1)!};$$

also in Moon (1974) and Dondajewski, Szymański (1982).

By known estimates for Stirling first numbers (H. 1995)

$$\mu_{n,k} = \frac{(\log n)^k}{\Gamma(1 + \frac{k}{\log n})k!}\left(1 + O\left(\frac{1}{\log n}\right)\right),$$

uniformly for $0 \le k \le K \log n$.

A ROUGH DESCRIPTION OF SHAPE

The root has about $\log n$ **subtrees,**

A ROUGH DESCRIPTION OF SHAPE

The root has about $\log n$ subtrees, each of them "attracting" about the same number of new keys.

A ROUGH DESCRIPTION OF SHAPE

The root has about $\log n$ subtrees, each of them "attracting" about the same number of new keys.

Also $\dfrac{\log \mu_{n,k}}{\log n} \to \alpha(1 - \log \alpha)$ **for** $0 \leq \alpha \leq K$**,**

A ROUGH DESCRIPTION OF SHAPE

The root has about $\log n$ subtrees, each of them "attracting" about the same number of new keys.

Also $\dfrac{\log \mu_{n,k}}{\log n} \to \alpha(1-\log\alpha)$ **for** $0 \le \alpha \le K$**, and**

$\mu_{n,k} \to \infty$ **when** $1 \le k \le e\log n - \dfrac{1}{2}\log\log n + O(1)$**.**

TWO COROLLARIES

The estimate for $\mu_{n,k}$ also implies

☞ **an LLT for the depth; see Devroye (1988), Szymański (1990), Mahmoud (1991) for CLT, and Dobrow, Smythe (1996) for Poisson approximation;**

TWO COROLLARIES

The estimate for $\mu_{n,k}$ also implies

- ☞ **an LLT for the depth; see Devroye (1988), Szymański (1990), Mahmoud (1991) for CLT, and Dobrow, Smythe (1996) for Poisson approximation;**
- ☞ **that the expected height is bounded above by**

$$\mathbb{E}(H_n) \le e \log n - \frac{1}{2} \log \log n + O(1).$$

(Roughly, the range when $\mu_{n,k} \to \infty$.)

SECOND MOMENT OF $X_{n,k}$

Meir, Moon (1978) (implicit)

$$\mathbb{E}(X_{n,k}^2) = \sum_{0 \le j \le k} \binom{2j}{j} \frac{\text{Stirling1}(n, k+j+1)}{(n-1)!};$$

see also van der Hofstad et al. (2002);

SECOND MOMENT OF $X_{n,k}$

Meir, Moon (1978) (implicit)

$$\mathbb{E}(X_{n,k}^2) = \sum_{0\le j\le k} \binom{2j}{j} \frac{\text{Stirling1}(n, k+j+1)}{(n-1)!};$$

see also van der Hofstad et al. (2002); and for $k = O(1)$

$$\mathbb{V}(X_{n,k}) \sim \frac{(\log n)^{2k-1}}{(2k-1)(k-1)!^2}.$$

VARIANCE OF $X_{n,k}$: MIDDLE RANGE

Uniformly for $1 \leq k \leq 2\log n - K\sqrt{\log n}$

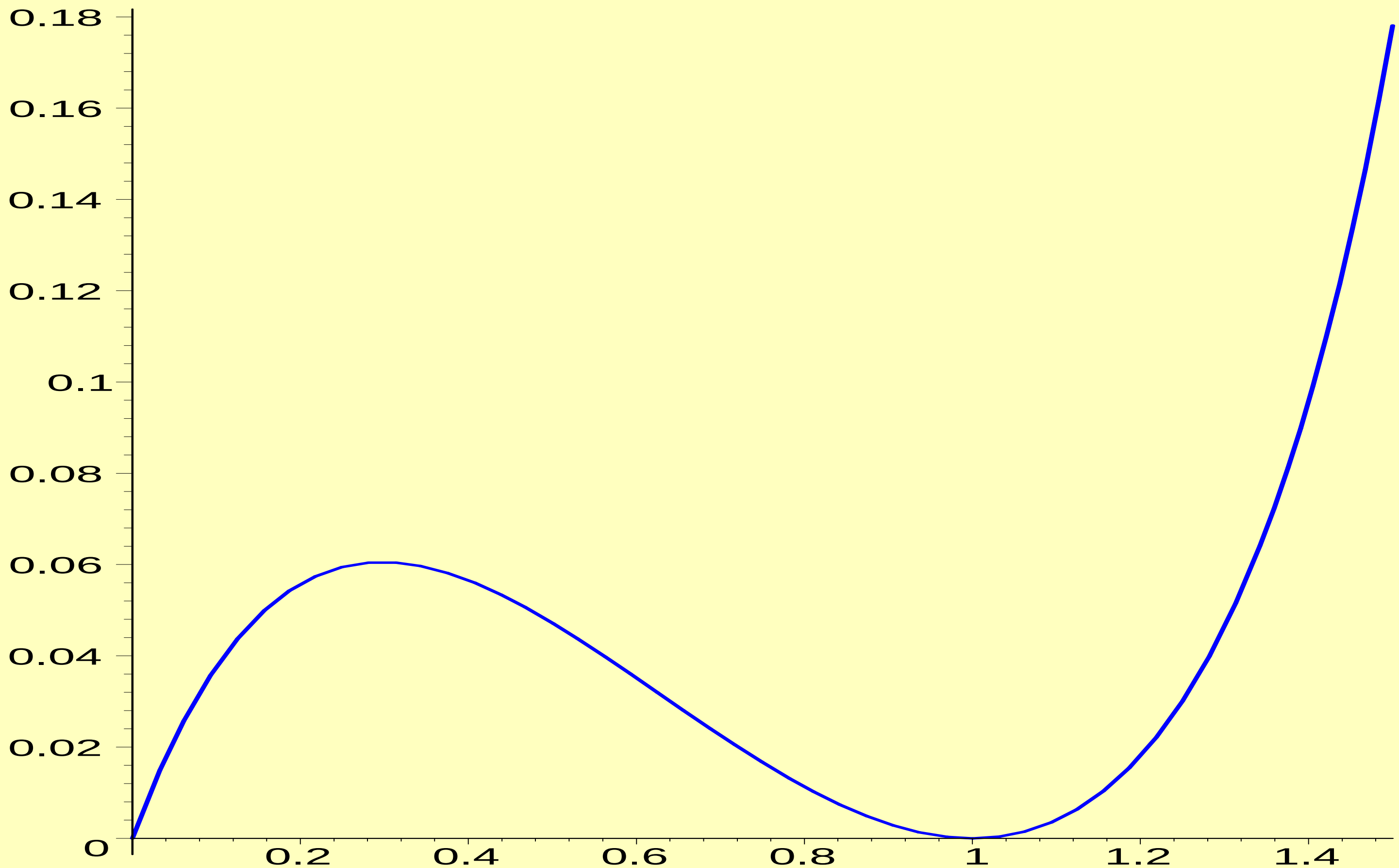
$$\mathbb{V}(X_{n,k}) \sim \phi(\alpha_{n,k})\mu_{n,k}^2,$$

where

$$\phi(x) := \frac{\Gamma(x+1)^2}{(1-\frac{x}{2})\Gamma(2x+1)} - 1.$$

A full asymptotic expansion can be derived.

$$\phi(0) = \phi(1) = \phi'(1) = 0$$



MORE PRECISE ESTIMATES

$$\mathbb{V}(X_{n,k}) \sim \begin{cases} \dfrac{(\log n)^{2k-1}}{(2k-1)(k-1)!^2}, & \textbf{if } k = o(\log n), \\ p(t_n)\left(\dfrac{(\log n)^{k-1}}{k!}\right)^2, & \textbf{if } \begin{aligned} t_n &:= k - \log n \\ &= o(\log n), \end{aligned} \end{cases}$$

where

$$p(t_n) := \left(2 - \frac{\pi^2}{6}\right) t_n^2 - \left(2\zeta(3) + 4\gamma + \frac{\pi^2}{3}(1-\gamma) - 6\right) t_n + 2\gamma^2 - 6\gamma + 8 - 2\zeta(3)(1-\gamma) - \frac{\pi^2}{6}\left(\gamma^2 - 2\gamma + 3\right) - \frac{\pi^4}{360}.$$

MORE PRECISE ESTIMATES

$$\mathbb{V}(X_{n,k}) \sim \begin{cases} \dfrac{(\log n)^{2k-1}}{(2k-1)(k-1)!^2}, & \textbf{if } k = o(\log n), \\ p(t_n)\left(\dfrac{(\log n)^{k-1}}{k!}\right)^2, & \textbf{if } \begin{aligned} t_n &:= k - \log n \\ &= o(\log n), \end{aligned} \end{cases}$$

where

$$p(t_n) := \left(2 - \frac{\pi^2}{6}\right) t_n^2 - \left(2\zeta(3) + 4\gamma + \frac{\pi^2}{3}(1-\gamma) - 6\right) t_n + 2\gamma^2 - 6\gamma + 8 - 2\zeta(3)(1-\gamma) - \frac{\pi^2}{6}\left(\gamma^2 - 2\gamma + 3\right) - \frac{\pi^4}{360}.$$

$$\frac{\phi''(1)}{2} = 2 - \frac{\pi^2}{6}$$

BIMODALITY OF VARIANCE WHEN $\alpha = 1$

$$\min_{k=\log n+O(\sqrt{\log n})} \mathbb{V}(X_{n,k}) \asymp \frac{n^2}{(\log n)^3},$$

BIMODALITY OF VARIANCE WHEN $\alpha = 1$

$$\min_{k=\log n+O(\sqrt{\log n})} \mathbb{V}(X_{n,k}) \asymp \frac{n^2}{(\log n)^3},$$

$$\max_{1\le k<n} \mathbb{V}(X_{n,k}) \sim \frac{12-\pi^2}{12\pi e} \cdot \frac{n^2}{(\log n)^2}.$$

BIMODALITY OF VARIANCE WHEN $\alpha = 1$

$$\min_{k=\log n+O(\sqrt{\log n})} \mathbb{V}(X_{n,k}) \asymp \frac{n^2}{(\log n)^3},$$

$$\max_{1\le k<n} \mathbb{V}(X_{n,k}) \sim \frac{12-\pi^2}{12\pi e} \cdot \frac{n^2}{(\log n)^2}.$$

Note that $\mathbb{V}(X_{n,k}) \sim \frac{p(t_n)}{(\log n)^2} \cdot \mu_{n,k}^2$.

BIMODALITY OF VARIANCE WHEN $\alpha = 1$

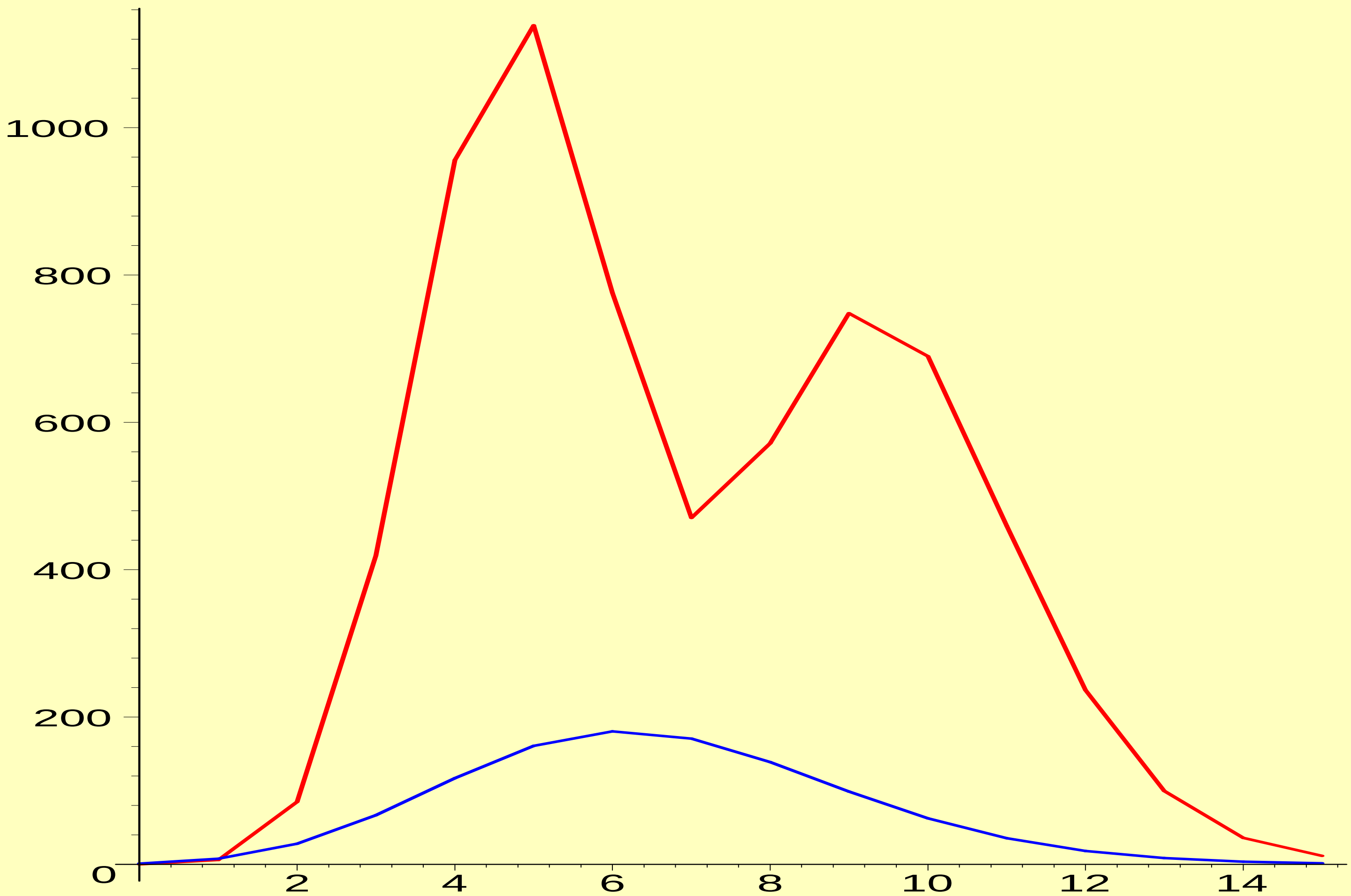
$$\min_{k=\log n+O(\sqrt{\log n})} \mathbb{V}(X_{n,k}) \asymp \frac{n^2}{(\log n)^3},$$

$$\max_{1\le k<n} \mathbb{V}(X_{n,k}) \sim \frac{12-\pi^2}{12\pi e} \cdot \frac{n^2}{(\log n)^2}.$$

Note that $\mathbb{V}(X_{n,k}) \sim \frac{p(t_n)}{(\log n)^2} \cdot \mu_{n,k}^2$**.**

More precise estimates can be derived.

$\mathbb{E}(X_{500,k})$ AND $\mathbb{V}(X_{500,k})$



A GLOBAL DESCRIPTION OF $\mathbb{V}(X_{n,k})$

$$\frac{\log \mathbb{V}(X_{n,k})}{\log n} \to \begin{cases} 2\alpha(1-\log\alpha), & \textbf{if } 0 \le \alpha \le 2; \\ 4-\alpha\log 4, & \textbf{if } 2 \le \alpha \le 4; \\ \alpha(1-\log\alpha), & \textbf{if } 4 \le \alpha \le K. \end{cases}$$

A GLOBAL DESCRIPTION OF $\mathbb{V}(X_{n,k})$

$$\frac{\log \mathbb{V}(X_{n,k})}{\log n} \to \begin{cases} 2\alpha(1-\log\alpha), & \textbf{if } 0 \le \alpha \le 2; \\ 4-\alpha\log 4, & \textbf{if } 2 \le \alpha \le 4; \\ \alpha(1-\log\alpha), & \textbf{if } 4 \le \alpha \le K. \end{cases}$$

Thus $\mathbb{V}(X_{n,k}) \begin{cases} \asymp \mu_{n,k}^2, & \textbf{if } 0 \le \alpha < 2; \\ \gg \mu_{n,k}^2, \mu_{n,k}, & \textbf{if } 2 \le \alpha \le 4; \\ \asymp \mu_{n,k}, & \textbf{if } 4 < \alpha \le K. \end{cases}$

LIMIT DISTRIBUTION: GENESIS

Let $\bar{X}_{n,k} := X_{n,k}/\mu_{n,k}$ **and** $I_n :=$ uniform$[1, n-1]$.

LIMIT DISTRIBUTION: GENESIS

Let $\bar{X}_{n,k} := X_{n,k}/\mu_{n,k}$ **and** $I_n :=$ uniform$[1, n-1]$.

Then from

$$X_{n,k} \stackrel{d}{=} X_{I_n,k-1} + X^*_{n-I_n,k},$$

it follows that

$$\bar{X}_{n,k} \stackrel{d}{=} \frac{\mu_{I_n,k-1}}{\mu_{n,k}} \bar{X}_{I_n,k-1} + \frac{\mu_{n-I_n,k}}{\mu_{n,k}} \bar{X}^*_{n-I_n,k}.$$

LIMIT DISTRIBUTION: GENESIS

Since $\mu_{n,k} \approx (\log n)^k/k!$ **and** $I_n = \lceil (n-1)U \rceil$**, we expect that**

$$\frac{\mu_{I_n,k-1}}{\mu_{n,k}} \approx \frac{k}{\log n}\left(\frac{\log n + \log U}{\log n}\right)^{k-1} \xrightarrow{d} \alpha U^{\alpha},$$

LIMIT DISTRIBUTION: GENESIS

Since $\mu_{n,k} \approx (\log n)^k/k!$ **and** $I_n = \lceil (n-1)U \rceil$**, we expect that**

$$\frac{\mu_{I_n,k-1}}{\mu_{n,k}} \approx \frac{k}{\log n}\left(\frac{\log n + \log U}{\log n}\right)^{k-1} \xrightarrow{d} \alpha U^{\alpha},$$

and similarly $\dfrac{\mu_{n-I_n,k}}{\mu_{n,k}} \xrightarrow{d} (1-U)^{\alpha}$**.**

LIMIT DISTRIBUTION: GENESIS

Since $\mu_{n,k} \approx (\log n)^k/k!$ **and** $I_n = \lceil (n-1)U \rceil$**, we expect that**

$$\frac{\mu_{I_n,k-1}}{\mu_{n,k}} \approx \frac{k}{\log n} \left(\frac{\log n + \log U}{\log n} \right)^{k-1} \xrightarrow{d} \alpha U^\alpha,$$

and similarly $\dfrac{\mu_{n-I_n,k}}{\mu_{n,k}} \xrightarrow{d} (1-U)^\alpha$**. Thus if**

$\bar{X}_{n,k} \xrightarrow{d} X_\alpha$**, then** $X_\alpha \stackrel{d}{=} \alpha U^\alpha X_\alpha + (1-U)^\alpha X_\alpha^*$**.**

LIMIT DISTRIBUTIONS: NEW RESULTS

For $0 \leq \alpha < e$

$$\frac{X_{n,k}}{\mu_{n,k}} \xrightarrow{d} X_\alpha,$$

with convergence of the first m **moments for** $0 \leq \alpha < m^{1/(m-1)}$**.**

LIMIT DISTRIBUTIONS: NEW RESULTS

For $0 \le \alpha < e$

$$\frac{X_{n,k}}{\mu_{n,k}} \xrightarrow{d} X_\alpha,$$

with convergence of the first m **moments for** $0 \le \alpha < m^{1/(m-1)}$**.**

In particular, convergence of all moments holds only for $\alpha \in [0, 1]$**.**

LIMIT DISTRIBUTIONS: NEW RESULTS

For $0 \le \alpha < e$

$$\frac{X_{n,k}}{\mu_{n,k}} \xrightarrow{d} X_\alpha,$$

with convergence of the first m **moments for** $0 \le \alpha < m^{1/(m-1)}$.

In particular, convergence of all moments holds only for $\alpha \in [0, 1]$.

$$X_0 = X_1 \equiv 1$$

RECURRENCE OF MOMENTS

Let $\nu_m := \mathbb{E}(X_\alpha^m)$**. Then** $\nu_0 = \nu_1 = 1$ **and**

$$\nu_m = \frac{1}{m - \alpha^{m-1}} \sum_{1 \le j < m} \binom{m}{j} \nu_j \nu_{m-j} \alpha^{j-1} \times \frac{\Gamma(j\alpha + 1)\Gamma((m-j)\alpha + 1)}{\Gamma(m\alpha + 1)} \qquad (m \ge 2),$$

for $0 \le \alpha < m^{1/(m-1)}$**.**

ASYMPTOTIC NORMALITY WHEN $\alpha = 0$

If $1 \le k = o(\log n)$, then

$$\sup_x \left| \mathbb{P}\left(\frac{X_{n,k} - \frac{(\log n)^k}{k!}}{\sqrt{\frac{(\log n)^{2k-1}}{(k-1)!^2(2k-1)}}} < x \right) - \Phi(x) \right| = O\left(\sqrt{\frac{k}{\log n}} \right),$$

where $\Phi(x)$ denotes the standard normal distribution.

ASYMPTOTIC NORMALITY WHEN $\alpha = 0$

If $1 \le k = o(\log n)$, then

$$\sup_x \left| \mathbb{P}\left(\frac{X_{n,k} - \frac{(\log n)^k}{k!}}{\sqrt{\frac{(\log n)^{2k-1}}{(k-1)!^2(2k-1)}}} < x \right) - \Phi(x) \right| = O\left(\sqrt{\frac{k}{\log n}} \right),$$

where $\Phi(x)$ denotes the standard normal distribution.

In particular, $\begin{cases} X_{n,1} \sim N(\log n, \log n) \\ X_{n,2} \sim N(\frac{1}{2}(\log n)^2, \frac{1}{3}(\log n)^3) \end{cases}$ **...**

A QUICKSORT-TYPE LIMIT LAW WHEN $\alpha = 1$

If $t_n := k - \log n = o(\log n)$ **and** $t_n \to \infty$**, then**

$$\frac{X_{n,k} - \mu_{n,k}}{t_n \frac{(\log n)^{k-1}}{k!}} \xrightarrow{m} X_1',$$

where $X_1' = (\mathbf{d}X_\alpha / \mathbf{d}\alpha)|_{\alpha=1}$ **or**

$$X_1' \stackrel{d}{=} UX_1' + (1-U)X_1'^{*} + U + U\log U + (1-U)\log(1-U).$$

A QUICKSORT-TYPE LIMIT LAW WHEN $\alpha = 1$

If $t_n := k - \log n = o(\log n)$ **and** $t_n \to \infty$**, then**

$$\frac{X_{n,k} - \mu_{n,k}}{t_n \frac{(\log n)^{k-1}}{k!}} \xrightarrow{m} X_1',$$

where $X_1' = (\mathbf{d} X_\alpha / \mathbf{d}\alpha)|_{\alpha=1}$ **or**

$$X_1' \stackrel{d}{=} UX_1' + (1-U)X_1'^* + U + U\log U + (1-U)\log(1-U).$$

Same law as total path length (or left path length in BST; Dobrow, Fill, 1999) and cost of an in-situ permutation algorithm.

NONEXISTENCE OF LIMIT LAW WHEN $k = \log n + O(1)$

If $k = \log n + O(1)$, then the limit distribution of $(X_{n,k} - \mu_{n,k})/\sqrt{\mathbb{V}(X_{n,k})}$ does not exist.

NONEXISTENCE OF LIMIT LAW WHEN $k = \log n + O(1)$

If $k = \log n + O(1)$, then the limit distribution of $(X_{n,k} - \mu_{n,k})/\sqrt{\mathbb{V}(X_{n,k})}$ does not exist.

Main step of the proof:

$$\mathbb{E}(X_{n,k} - \mu_{n,k})^m \sim \underbrace{\text{Polynomial}}_{\text{degree}=m}(t_n) \left(\frac{(\log n)^{k-1}}{k!} \right)^m ;$$

the remaining proof is similar to that used for the space requirement of random m-ary search trees when $m \geq 27$.

APPROACHES USED

Convergence in distribution: contraction method

APPROACHES USED

Convergence in distribution: contraction method

Convergence of moments: method of moments

APPROACHES USED

Convergence in distribution: contraction method

Convergence of moments: method of moments

Common to both approaches is the resolution of the double-indexed recurrence

$$a_{n,k} = b_{n,k} + \frac{1}{n-1} \sum_{1 \le j < n} (a_{j,k-1} + a_{j,k}).$$

APPROACHES USED

Convergence in distribution: contraction method

Convergence of moments: method of moments

Common to both approaches is the resolution of the double-indexed recurrence

$$a_{n,k} = b_{n,k} + \frac{1}{n-1} \sum_{1 \le j < n} (a_{j,k-1} + a_{j,k}).$$

(Martingale arguments also apply to recursive trees.)

SOLUTION OF THE RECURRENCE

If

$$a_{n,k} = \frac{1}{n-1} \sum_{1 \le j < n} \left(a_{j,k} + a_{j,k-1} \right) + b_{n,k}, \quad (n \ge 2; k \ge 1),$$

with $a_{n,k} = b_{n,k}$ **for** $n = 1$ **and** $k = 0$**,**

SOLUTION OF THE RECURRENCE

If

$$a_{n,k} = \frac{1}{n-1} \sum_{1 \le j < n} (a_{j,k} + a_{j,k-1}) + b_{n,k}, \quad (n \ge 2; k \ge 1),$$

with $a_{n,k} = b_{n,k}$ **for** $n = 1$ **and** $k = 0$**, then**

$$a_{n,k} = b_{n,k} + \sum_{2 \le j < n} j^{-1} \sum_{0 \le r \le k} b_{j,k-r} [w^r](w+1) \prod_{j < \ell < n} \left(1 + \frac{w}{\ell}\right).$$

SOLUTION OF THE RECURRENCE

If

$$a_{n,k} = \frac{1}{n-1} \sum_{1 \le j < n} (a_{j,k} + a_{j,k-1}) + b_{n,k}, \quad (n \ge 2; k \ge 1),$$

with $a_{n,k} = b_{n,k}$ **for** $n = 1$ **and** $k = 0$**, then**

$$a_{n,k} = b_{n,k} + \sum_{2 \le j < n} j^{-1} \sum_{0 \le r \le k} b_{j,k-r} [w^r](w+1) \prod_{j < \ell < n} \left(1 + \frac{w}{\ell}\right).$$

Proof by GF

WHICHEVER APPROACH REQUIRES MEAN VALUE

$\mu_{n,k}$ **satisfies the recurrence with** $b_{n,k} = \delta_{0,k}$**.**

WHICHEVER APPROACH REQUIRES MEAN VALUE

$\mu_{n,k}$ **satisfies the recurrence with** $b_{n,k} = \delta_{0,k}$**. Thus**

$$\mu_{n,k} = [w^k](w+1) \sum_{2 \le j < n} \frac{1}{j} \prod_{j < \ell < n} \left(1 + \frac{w}{\ell}\right)$$

$$= \frac{\text{Stirling1}(n, k+1)}{(n-1)!}$$

$$= \frac{(\log n)^k}{\Gamma(1 + \frac{k}{\log n}) k!} \left(1 + O\left(\frac{1}{\log n}\right)\right).$$

WHICHEVER APPROACH REQUIRES MEAN VALUE

$\mu_{n,k}$ **satisfies the recurrence with** $b_{n,k} = \delta_{0,k}$**. Thus**

$$\mu_{n,k} = [w^k](w+1) \sum_{2 \le j < n} \frac{1}{j} \prod_{j < \ell < n} \left(1 + \frac{w}{\ell}\right)$$

$$= \frac{\text{Stirling1}(n, k+1)}{(n-1)!}$$

$$= \frac{(\log n)^k}{\Gamma(1 + \frac{k}{\log n})k!} \left(1 + O\left(\frac{1}{\log n}\right)\right).$$

Sufficient for all ranges except $k = \log n + O(1)$**.**

ASYMPTOTICS OF HIGHER MOMENTS

For method of moments: all moments (centered or not) satisfy the same type of recurrences

$$a_{n,k} = \frac{1}{n-1} \sum_{1 \le j < n} (a_{j,k} + a_{j,k-1}) + b_{n,k},$$

with different $b_{n,k}$, and we need the $\sim$-transfer :

ASYMPTOTICS OF HIGHER MOMENTS

For method of moments: all moments (centered or not) satisfy the same type of recurrences

$$a_{n,k} = \frac{1}{n-1} \sum_{1\le j<n} (a_{j,k} + a_{j,k-1}) + b_{n,k},$$

with different $b_{n,k}$, and we need the $\sim$-transfer:

if $b_{n,k} \sim c\left(\frac{(\log n)^k}{\Gamma(1+\alpha)k!}\right)^m$**, then** $a_{n,k} \sim c\frac{m\alpha+1}{m\alpha-\alpha^m}\left(\frac{(\log n)^k}{\Gamma(1+\alpha)k!}\right)^m$**.**

$$c = c(\alpha)$$

CONTRACTION METHOD: CONVERGENCE IN DISTRIBUTION

For contraction method: (i) **choose** $s \in (1, 2]$ **such that**

$$\mathbb{E}\left(\alpha^s U^{s\alpha} + (1-U)^{\alpha s}\right) = \frac{\alpha^s + 1}{\alpha s + 1} < 1,$$

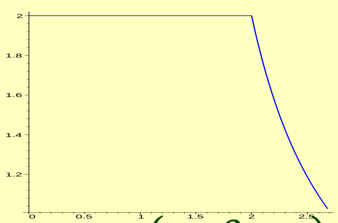
or, equivalently, $s - \alpha^{s-1} > 0$**.**

CONTRACTION METHOD: CONVERGENCE IN DISTRIBUTION

For contraction method: (i) **choose** $s \in (1,2]$ **such that**

$$\mathbb{E}\left(\alpha^s U^{s\alpha} + (1-U)^{\alpha s}\right) = \frac{\alpha^s + 1}{\alpha s + 1} < 1,$$

or, equivalently, $s - \alpha^{s-1} > 0$.



(ii) **need the O-transfer: If** $b_{n,k} = o(\mu_{n,k}^s)$**, where** $s > 1$ **satisfies** $s - \alpha^{s-1} > 0$**, then** $a_{n,k} = o(\mu_{n,k}^s)$**.**

CONTRACTION METHOD: CONVERGENCE RATE

For contraction method, a convergence rate in Zolotarev distance can be derived when $0 \le \alpha < 2$ by proving the stronger O-transfer:

$$\text{if } b_{n,k} = O\left(\frac{|k - \alpha \log n| \vee 1}{\log n} \mu_{n,k}^s\right),$$

$$\text{then } a_{n,k} = O\left(\frac{|k - \alpha \log n| \vee 1}{\log n} \mu_{n,k}^s\right),$$

where $s > 1$ satisfies $s - \alpha^{s-1} > 0$.

THE CENTRAL RANGE $\alpha = 1$

When $k = \log n + o(\log n)$**, same moments approach but starting with** $\bar{P}_{n,k}(y) := \mathbb{E}(e^{(X_{n,k}-\mu_{n,k})y})$**, which satisfies**

$$\bar{P}_{n,k}(y) = \frac{1}{n-1} \sum_{1 \le j < n} \bar{P}_{j,k-1}(y) \bar{P}_{n-j,k}(y) e^{-\Delta_{n,k}(j)y},$$

for $n \ge 2, k \ge 1$**, where** $\Delta_{n,k}(j) := \mu_{j,k-1} + \mu_{n-j,k} - \mu_{n,k}$**.**

THE CENTRAL RANGE $\alpha = 1$

When $k = \log n + o(\log n)$, same moments approach but starting with $\bar{P}_{n,k}(y) := \mathbb{E}(e^{(X_{n,k}-\mu_{n,k})y})$, which satisfies

$$\bar{P}_{n,k}(y) = \frac{1}{n-1} \sum_{1 \le j < n} \bar{P}_{j,k-1}(y) \bar{P}_{n-j,k}(y) e^{-\Delta_{n,k}(j)y},$$

for $n \ge 2, k \ge 1$, where $\Delta_{n,k}(j) := \mu_{j,k-1} + \mu_{n-j,k} - \mu_{n,k}$.

A more uniform estimate for $\Delta_{n,k}(j)$ is needed.

THE NORMAL RANGE: $\alpha = 0$

Let $\sigma_{n,k} := \sqrt{\frac{(\log n)^{2k-1}}{(2k-1)(k-1)!^2}}$ **and** $\Lambda := \frac{\log n}{k}$**.**

Two estimates are needed: (i) **uniformly for** $|\theta| \leq \varepsilon \Lambda^{1/6}$

$$\left| \mathbb{E}\left(e^{\frac{X_{n,k} - (\log n)^k / k!}{\sigma_{n,k}} i\theta} \right) - e^{-\theta^2/2} \right| = O\left(\frac{|\theta| + |\theta|^3}{\sqrt{\Lambda}} e^{-\theta^2/2} + n^{-\varepsilon} \right);$$

THE NORMAL RANGE: $\alpha = 0$

Let $\sigma_{n,k} := \sqrt{\frac{(\log n)^{2k-1}}{(2k-1)(k-1)!^2}}$ **and** $\Lambda := \frac{\log n}{k}$**.**

Two estimates are needed: (i) **uniformly for** $|\theta| \le \varepsilon\Lambda^{1/6}$

$$\left|\mathbb{E}\left(e^{\frac{X_{n,k}-(\log n)^k/k!}{\sigma_{n,k}} i\theta}\right) - e^{-\theta^2/2}\right| = O\left(\frac{|\theta| + |\theta|^3}{\sqrt{\Lambda}} e^{-\theta^2/2} + n^{-\varepsilon}\right);$$

and (ii) **uniformly for** $\varepsilon\Lambda^{1/6} \le |\theta| \le \varepsilon\sqrt{\Lambda}$

$$\mathbb{E}\left(e^{\frac{X_{n,k}-(\log n)^k/k!}{\sigma_{n,k}} i\theta}\right) = O(e^{-\theta^2/4} + n^{-\varepsilon}).$$

IDEA OF PROOF

Let $Q_k(z,y) := \sum_{n\ge 1} \mathbb{E}(e^{X_{n,k}y}) z^{n-1}$**. Then**

$$\begin{cases} Q_0(z,s) = \dfrac{e^s}{1-z}, \\ Q_k(z,s) = \exp\left(\displaystyle\int_0^z Q_{k-1}(t,s)\,\mathbf{d}t\right), \qquad (k \ge 1). \end{cases}$$

IDEA OF PROOF

Decompose Q_k in two ways as follows.

$$Q_k(z,s) := \exp\left(\sum_{m\geq 0} \frac{V_{k,m}(-\log(1-z))}{m!} s^m\right)$$

$$:= \frac{1}{1-z} \sum_{m\geq 0} \frac{W_{k,m}(-\log(1-z))}{m!} s^m.$$

Then manipulate inductively the recurrences of V and W.

PROFILE OF BST

Our approach based on contraction method and method of moments is applicable to BSTs.

$$\begin{cases} Y_{n,k} & := \textbf{the number of external nodes at level } k \\ Z_{n,k} & := \textbf{the number of internal nodes at level } k \end{cases}$$

which satisfies the same type of recurrences

$$C_{n,k} \stackrel{d}{=} C_{\text{uniform}[0,n-1],k-1} + C^*_{n-1-\text{uniform}[0,n-1],k-1},$$

with different initial conditions.

MEAN VALUES

Mean values known since 1960's: **Lynch (1965), Knuth (1998), Brown, Shubert (1984), Mahmoud, Pittel (1984), Pittel (1984), Louchard (1987), Devroye (1988).**

MEAN VALUES

Mean values known since 1960's: Lynch (1965), Knuth (1998), Brown, Shubert (1984), Mahmoud, Pittel (1984), Pittel (1984), Louchard (1987), Devroye (1988).

$$\begin{aligned}\mathbb{E}(Y_{n,k}) &= \frac{2^k}{n!}\,\text{Stirling1}(n,k)\\ &= \frac{(2\log n)^k}{\Gamma(\alpha_{n,k})k!n}\left(1+O\left(\frac{1}{\log n}\right)\right),\end{aligned}$$

uniformly for $0 \le k \le K\log n$**.**

THE BINARY SEARCH TREE CONSTANTS

$$\frac{\log \mathbb{E}(Y_{n,k})}{\log n} \to \lambda(\alpha) := \alpha - 1 - \alpha \log(\alpha/2).$$

Thus $\mathbb{E}(Y_{n,k}) \to \infty$ **when** $\alpha_- < \alpha < \alpha_+$**, where** $0 < \alpha_- < 1 < \alpha_+$ **are the two real zeros of the equation** $z - 1 - z\log(z/2)$ **or** $e^{(z-1)/z} = z/2$**.**

THE BINARY SEARCH TREE CONSTANTS

$$\frac{\log \mathbb{E}(Y_{n,k})}{\log n} \to \lambda(\alpha) := \alpha - 1 - \alpha \log(\alpha/2).$$

Thus $\mathbb{E}(Y_{n,k}) \to \infty$ **when** $\alpha_- < \alpha < \alpha_+$**, where** $0 < \alpha_- < 1 < \alpha_+$ **are the two real zeros of the equation** $z - 1 - z\log(z/2)$ **or** $e^{(z-1)/z} = z/2$**.**

Also the estimate for $\mu_{n,k}$ **implies that the expected height is bounded above by**

$$\mathbb{E}(H_n) \le \alpha_+ \log n - \frac{\alpha_+}{2(\alpha_+ - 1)} \log\log n + O(1).$$

LIMIT DISTRIBUTIONS

Chauvin et al. (2001): almost sure convergence of

$$\frac{Y_{n,k}}{\mathbb{E}(Y_{n,k})}, \frac{Z_{n,k}}{\mathbb{E}(Z_{n,k})} \xrightarrow{d} Y_\alpha,$$

for $1.2 \le \alpha \le 2.8$**, where**

$$Y_\alpha \stackrel{d}{=} \frac{\alpha}{2} U^{\alpha-1} Y_\alpha + \frac{\alpha}{2}(1-U)^{\alpha-1} Y_\alpha^*.$$

LIMIT DISTRIBUTIONS

Chauvin et al. (2001): almost sure convergence of

$$\frac{Y_{n,k}}{\mathbb{E}(Y_{n,k})}, \frac{Z_{n,k}}{\mathbb{E}(Z_{n,k})} \xrightarrow{d} Y_\alpha,$$

for $1.2 \le \alpha \le 2.8$**, where**

$$Y_\alpha \stackrel{d}{=} \frac{\alpha}{2} U^{\alpha-1} Y_\alpha + \frac{\alpha}{2}(1-U)^{\alpha-1} Y_\alpha^*.$$

Chauvin et al. (2003+): convergence in probability for $Y_{n,k}/\mathbb{E}(Y_{n,k})$ **when** $k = \alpha \log n + O(\sqrt{\log n})$**,** $\alpha_- < \alpha < \alpha_+$**.**

NEW RESULTS

Define $\bar{Z}_{n,k} := \begin{cases} 2^k - Z_{n,k}, & \textbf{if } \alpha_- \le \alpha < 1; \\ Z_{n,k}, & \textbf{if } 1 \le \alpha < \alpha_+. \end{cases}$

NEW RESULTS

Define $\bar{Z}_{n,k} := \begin{cases} 2^k - Z_{n,k}, & \textbf{if } \alpha_- \le \alpha < 1; \\ Z_{n,k}, & \textbf{if } 1 \le \alpha < \alpha_+. \end{cases}$

If $k = \alpha \log n + o(\log n)$**, where** $\alpha_- < \alpha < \alpha_+$**, then**

$$\frac{Y_{n,k}}{\mathbb{E}(Y_{n,k})}, \frac{\bar{Z}_{n,k}}{\mathbb{E}(\bar{Z}_{n,k})} \xrightarrow{d} Y_\alpha,$$

with convergence of all moments for $\alpha \in [1, 2]$ **but not for** α **outside** $[1, 2]$**.**

MOMENTS OF THE LIMIT LAW

$\eta_0 = \eta_1 = 1$ **and for** $m \geq 2$

$$\eta_m = \frac{(\alpha/2)^m}{m(\alpha-1)+1-2(\alpha/2)^m} \times \sum_{1\leq j<m} \binom{m}{j} \eta_j \eta_{m-j} \frac{\Gamma(j(\alpha-1)+1)\Gamma((m-j)(\alpha-1)+1)}{\Gamma(m(\alpha-1)+1)}.$$

MOMENTS OF THE LIMIT LAW

$\eta_0 = \eta_1 = 1$ **and for** $m \geq 2$

$$\eta_m = \frac{(\alpha/2)^m}{m(\alpha-1)+1-2(\alpha/2)^m} \times \sum_{1 \leq j < m} \binom{m}{j} \eta_j \eta_{m-j} \frac{\Gamma(j(\alpha-1)+1)\Gamma((m-j)(\alpha-1)+1)}{\Gamma(m(\alpha-1)+1)}.$$

The polynomial $m(z-1)+1-2(z/2)^m$ **has two positive zeros** z_m^- **and** z_m^+**, where** $z_m^- \uparrow 1$**, and** $z_m^+ \downarrow 2$**.**

THE QUICKSORT LIMIT LAW WHEN $\alpha = 2$

Since $Y_1 = Y_2 = 1$, we need to refine the limit result.

THE QUICKSORT LIMIT LAW WHEN $\alpha = 2$

Since $Y_1 = Y_2 = 1$, we need to refine the limit result.

If $k = 2\log n + t_n$, where $t_n = o(\log n)$ and $t_n \to \infty$, then

$$\frac{Y_{n,k} - \mathbb{E}(Y_{n,k})}{t_n n^{\lambda(\alpha_{n,k})}/\sqrt{4\pi(\log n)^3}}, \frac{Z_{n,k} - \mathbb{E}(Z_{n,k})}{t_n n^{\lambda(\alpha_{n,k})}/\sqrt{4\pi(\log n)^3}} \xrightarrow{m} Y_2',$$

where $Y_2' := (dY_\alpha/d\alpha)|_{\alpha=2}$.

THE QUICKSORT LIMIT LAW WHEN $\alpha = 2$

Since $Y_1 = Y_2 = 1$, we need to refine the limit result.

If $k = 2\log n + t_n$, where $t_n = o(\log n)$ and $t_n \to \infty$, then

$$\frac{Y_{n,k} - \mathbb{E}(Y_{n,k})}{t_n n^{\lambda(\alpha_{n,k})}/\sqrt{4\pi(\log n)^3}}, \frac{Z_{n,k} - \mathbb{E}(Z_{n,k})}{t_n n^{\lambda(\alpha_{n,k})}/\sqrt{4\pi(\log n)^3}} \xrightarrow{m} Y_2',$$

where $Y_2' := (dY_\alpha/d\alpha)|_{\alpha=2}$.

The limit law Y_2' is essentially the quicksort limit law

$$Y_2' \stackrel{d}{=} UY_2' + (1-U)Y_2'^{*} + \frac{1}{2} + U\log U + (1-U)\log(1-U).$$

NONEXISTENCE OF LIMIT DISTRIBUTION

If $k = 2\log n + O(1)$**, then neither of the limit distribution of**

$$\left\{ \frac{Y_{n,k} - \mathbb{E}(Y_{n,k})}{\sqrt{\mathbb{V}(Y_{n,k})}}, \frac{Z_{n,k} - \mathbb{E}(Z_{n,k})}{\sqrt{\mathbb{V}(Z_{n,k})}} \right\}$$

exists.

THE RANGE $\alpha = 1$

If $k = \log n + t_n$**, where** $t_n = o(\log n)$ **and** $t_n \to \infty$**, then**

$$\frac{Y_{n,k} - \mathbb{E}(Y_{n,k})}{t_n n^{\lambda(\alpha_{n,k})}/\sqrt{2\pi(\log n)^3}} \xrightarrow{m} Y_1',$$

where $Y_1' := (dY_\alpha/d\alpha)|_{\alpha=1}$**;**

THE RANGE $\alpha = 1$

If $k = \log n + t_n$**, where** $t_n = o(\log n)$ **and** $t_n \to \infty$**, then**

$$\frac{Y_{n,k} - \mathbb{E}(Y_{n,k})}{t_n n^{\lambda(\alpha_{n,k})}/\sqrt{2\pi(\log n)^3}} \xrightarrow{m} Y_1',$$

where $Y_1' := (\mathrm{d}Y_\alpha/\mathrm{d}\alpha)|_{\alpha=1}$**; if** $t_n = O(1)$**, then the limit distribution of** $\frac{Y_{n,k}-\mathbb{E}(Y_{n,k})}{\sqrt{\mathbb{V}(Y_{n,k})}}$ **does not exist.**

THE RANGE $\alpha = 1$

If $k = \log n + t_n$, where $t_n = o(\log n)$ and $t_n \to \infty$, then

$$\frac{Y_{n,k} - \mathbb{E}(Y_{n,k})}{t_n n^{\lambda(\alpha_{n,k})}/\sqrt{2\pi(\log n)^3}} \xrightarrow{m} Y_1',$$

where $Y_1' := (dY_\alpha/d\alpha)|_{\alpha=1}$; if $t_n = O(1)$, then the limit distribution of $\frac{Y_{n,k}-\mathbb{E}(Y_{n,k})}{\sqrt{\mathbb{V}(Y_{n,k})}}$ does not exist.

$$Y_1' \stackrel{d}{=} \frac{1}{2}Y_1' + \frac{1}{2}Y_1'^{*} + 1 + \frac{1}{2}\log U + \frac{1}{2}\log(1-U).$$

Any naturally defined RVs on trees leading to Y_1'?

DIFFERENT BEHAVIOR FOR INTERNAL NODES

For internal nodes, if $k = \log n + t_n$, then, uniformly for $t_n = o(\log n)$,

$$\frac{Z_{n,k} - \mathbb{E}(Z_{n,k})}{n^{\lambda(\alpha_{n,k})}/\sqrt{2\pi \log n}} \xrightarrow{m} Y_1'.$$

DIFFERENT BEHAVIOR FOR INTERNAL NODES

For internal nodes, if $k = \log n + t_n$, then, uniformly for $t_n = o(\log n)$,

$$\frac{Z_{n,k} - \mathbb{E}(Z_{n,k})}{n^{\lambda(\alpha_{n,k})}/\sqrt{2\pi \log n}} \xrightarrow{m} Y_1'.$$

The normalizing standard variation differs by a factor of $t_n / \log n$.

DIFFERENT BEHAVIOR FOR INTERNAL NODES

For internal nodes, if $k = \log n + t_n$, then, uniformly for $t_n = o(\log n)$,

$$\frac{Z_{n,k} - \mathbb{E}(Z_{n,k})}{n^{\lambda(\alpha_{n,k})}/\sqrt{2\pi \log n}} \xrightarrow{m} Y_1'.$$

The normalizing standard variation differs by a factor of $t_n/\log n$.

No normal range for BST-profile

GENERALITY (METHODOLOGY)

Our tools (contraction $+$ moments), with suitable development of "asymptotic transfer", are applicable to most BST relatives: m-ary search trees, median-of-$(2t+1)$ BSTs, Devroye's simplex trees and quadtrees. Technicalities are more involved.

GENERALITY (METHODOLOGY)

Our tools (contraction $+$ moments), with suitable development of "asymptotic transfer", are applicable to most BST relatives: m-ary search trees, median-of-$(2t+1)$ BSTs, Devroye's simplex trees and quadtrees. Technicalities are more involved.

All asymptotic tools needed are based on solving asymptotically (exact solution being less useful) the underlying double-indexed recurrence.

GENERALITY (METHODOLOGY)

Our tools (contraction $+$ moments), with suitable development of "asymptotic transfer", are applicable to most BST relatives: m-ary search trees, median-of-$(2t+1)$ BSTs, Devroye's simplex trees and quadtrees. Technicalities are more involved.

All asymptotic tools needed are based on solving asymptotically (exact solution being less useful) the underlying double-indexed recurrence.

No martingale in general

GENERALITY (METHODOLOGY)

✎ **m-ary search trees, simplex trees and median-of-$(2t+1)$ BSTs: GF satisfies a DE of Cauchy-Euler type; but more uniform estimates are needed.**

GENERALITY (METHODOLOGY)

- m-ary search trees, simplex trees and median-of-$(2t+1)$ BSTs: GF satisfies a DE of Cauchy-Euler type; but more uniform estimates are needed.
- quadtrees: extend Flajolet et al. (1993)'s direct DE and Flajolet et al. (1995)'s Euler-transform approaches for asymptotics of moments.

GENERALITY (PHENOMENA)

For the profiles of these random trees of logarithmic depth

1. **bimodality of variance near the central range**

GENERALITY (PHENOMENA)

For the profiles of these random trees of logarithmic depth

1. **bimodality of variance near the central range**
2. **cv in distribution in the range when mean** $\longrightarrow \infty$

GENERALITY (PHENOMENA)

For the profiles of these random trees of logarithmic depth

1. **bimodality of variance near the central range**
2. **cv in distribution in the range when mean** $\longrightarrow \infty$
3. **cv of all moments in some smaller range, say** $[\alpha_1, \alpha_2]$

GENERALITY (PHENOMENA)

For the profiles of these random trees of logarithmic depth

1. **bimodality of variance near the central range**
2. **cv in distribution in the range when mean $\longrightarrow \infty$**
3. **cv of all moments in some smaller range, say $[\alpha_1, \alpha_2]$**
4. **convergence of all moments in the two extreme cases (to the derivatives of the fixed-point equation)**

GENERALITY (PHENOMENA)

For the profiles of these random trees of logarithmic depth

1. **bimodality of variance near the central range**
2. **cv in distribution in the range when mean** $\to \infty$
3. **cv of all moments in some smaller range, say** $[\alpha_1, \alpha_2]$
4. **convergence of all moments in the two extreme cases (to the derivatives of the fixed-point equation)**
5. **nonexistence of limit law when** $k = \alpha_2 \log n + O(1)$

ARE THERE LOG-TREES WITH DIFFERENT BEHAVIOR FOR THEIR PROFILE?

REPRESENTATIVE CLASSES OF INCREASING TREES

Three representative classes of increasing trees (Bergeron, Flajolet, Salvy, 1992):

- ✔ **recursive trees:** $y' = e^y$

REPRESENTATIVE CLASSES OF INCREASING TREES

Three representative classes of increasing trees (Bergeron, Flajolet, Salvy, 1992):

- ✔ **recursive trees:** $y' = e^y$
- ✔ **binary increasing trees:** $y' = (1 + y)^2$

REPRESENTATIVE CLASSES OF INCREASING TREES

Three representative classes of increasing trees (Bergeron, Flajolet, Salvy, 1992):

- ✔ **recursive trees:** $y' = e^y$
- ✔ **binary increasing trees:** $y' = (1+y)^2$
- ☛ **ordered recursive (or heap-ordered, plane-oriented) trees:** $y' = \dfrac{1}{1-y}$

HEAP-ORDERED (ORDERED RECURSIVE) TREES

The profile $V_{n,k}$ satisfies the recurrence

$$V_{n,k} \stackrel{d}{=} V_{J_n,k-1} + V^*_{n-J_n,k},$$

where ($C_n := \binom{2n-2}{n-1}/n$)

$$\mathbb{P}(J_n = j) = \frac{2(n-j)C_j C_{n-j}}{nC_n} \qquad (1 \le j < n).$$

HEAP-ORDERED (ORDERED RECURSIVE) TREES

The profile $V_{n,k}$ satisfies the recurrence

$$V_{n,k} \stackrel{d}{=} V_{J_n,k-1} + V^*_{n-J_n,k},$$

where ($C_n := \binom{2n-2}{n-1}/n$)

$$\mathbb{P}(J_n = j) = \frac{2(n-j)C_j C_{n-j}}{nC_n} \qquad (1 \le j < n).$$

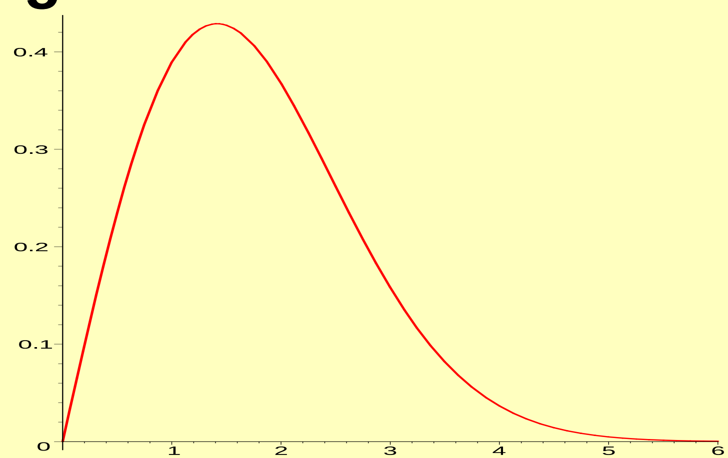
Note that $J_n \xrightarrow{d} J$, where $\mathbb{P}(J = h) = 2C_h/4^h$, but no convergence of any moment of order ≥ 1.

METHOD OF MOMENTS: $k = o(\log n)$

The method of moments applies: for $1 \leq k = o(\log n)$

$$\frac{V_{n,k}}{\sqrt{n}(\frac{1}{2}\log n)^{k-1}/(k-1)!} \longrightarrow \mathcal{R},$$

where $\mathcal{R}$ **is Rayleigh distributed with the density** $te^{-t^2/4}/2$**.**



METHOD OF MOMENTS: $0 \le \alpha \le 1/2$

If $0 \le \alpha \le 1/2$ then $\dfrac{V_{n,k}}{\mathbb{E}(V_{n,k})} \xrightarrow{m} V_\alpha$, for some V_α whose moments can be recursively computed.

Same bimodality for variance, as well as other log-profile phenomena as above for $\alpha = 1/2$.

METHOD OF MOMENTS: $0 \leq \alpha \leq 1/2$

But so far no contraction proof for $1/2 < \alpha < c$, where $c \approx 1.79556$ solves the equation $\frac{1}{2} + z - z\log(2z) = 0$, because it's not obvious how to write a fixed-point equation from the moment sequence of the limit law.

METHOD OF MOMENTS: $0 \le \alpha \le 1/2$

But so far no contraction proof for $1/2 < \alpha < c$, where $c \approx 1.79556$ solves the equation $\frac{1}{2} + z - z\log(2z) = 0$, because it's not obvious how to write a fixed-point equation from the moment sequence of the limit law.

New method is needed for proving the convergence in distribution (anticipated) of $\dfrac{V_{n,k}}{\mathbb{E}(V_{n,k})}$ in the range $(1/2, c)$.

PROFILES OF RANDOM TREES WITH $\sqrt{n}$ HEIGHT

Much has been known for such trees (random binary trees being typical): Stepanov (1969), Takács (1991), Aldous (1993), Drmota, Gittenberger (1997), Pitman (1998), Gittenberger (1998), Kersting (1998), . . .

PROFILES OF RANDOM TREES WITH $\sqrt{n}$ HEIGHT

Much has been known for such trees (random binary trees being typical): Stepanov (1969), Takács (1991), Aldous (1993), Drmota, Gittenberger (1997), Pitman (1998), Gittenberger (1998), Kersting (1998), . . .

For example, for random binary trees,

$$\begin{cases} \dfrac{X_{n,k}}{k/4} \xrightarrow{d} \text{Gamma}(1), & \textbf{if } k \to \infty, k = o(\sqrt{n}), \\ 2\sqrt{2}\dfrac{X_{n,k}}{\sqrt{n}} \xrightarrow{d} \text{Stepanov}_\alpha, & \textbf{if } \dfrac{k}{\sqrt{n}} \to 2\sqrt{2}\alpha. \end{cases}$$

OPEN QUESTIONS

- **What happens at the boundary $\alpha = \alpha_-, \alpha_+$?**

OPEN QUESTIONS

- **What happens at the boundary $\alpha = \alpha_-, \alpha_+$?**
- **More "humps" in the central range for higher moments?**

OPEN QUESTIONS

- **What happens at the boundary $\alpha = \alpha_-, \alpha_+$?**
- **More "humps" in the central range for higher moments?**
- **Are there good process approximations?**

OPEN QUESTIONS

- **What happens at the boundary $\alpha = \alpha_-, \alpha_+$?**
- **More “humps” in the central range for higher moments?**
- **Are there good process approximations?**
- **How to prove almost-sure convergence in general?**

OPEN QUESTIONS

- **What happens at the boundary $\alpha = \alpha_-, \alpha_+$?**
- **More "humps" in the central range for higher moments?**
- **Are there good process approximations?**
- **How to prove almost-sure convergence in general?**
- **More asymptotic transfer for the double-indexed recurrence?**

OPEN QUESTIONS

- **What happens at the boundary $\alpha = \alpha_-, \alpha_+$?**
- **More "humps" in the central range for higher moments?**
- **Are there good process approximations?**
- **How to prove almost-sure convergence in general?**
- **More asymptotic transfer for the double-indexed recurrence?**
- **How to plot or simulate the limit law?**

PROFILE OF RANDOM TREES: A RICH SOURCE OF INTRIGUING PHENOMENA