Depth and height of Scaled Attachement Random Recursive Trees (SARRT)

Luc Devroye[1]   Omar Fawzi[1]   Nicolas Fraiman[2]

[1]School of Computer Science
McGill University

[2]Department of Mathematics and Statistics
McGill University

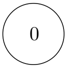
July 1st, 2010

Outline

Outline

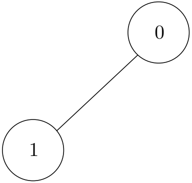
Uniform Random Recursive Trees

- Node labels: $\{0, \ldots, n\}$
- Node i picks one parent uniformly $\{0, \ldots, i-1\}$



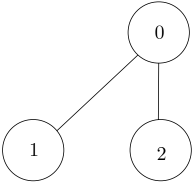
Uniform Random Recursive Trees

- Node labels: $\{0, \ldots, n\}$
- Node i picks one parent uniformly $\{0, \ldots, i-1\}$



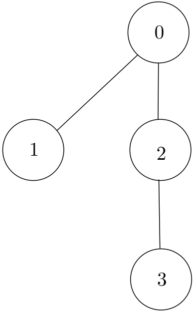
Uniform Random Recursive Trees

- Node labels: $\{0, \ldots, n\}$
- Node i picks one parent uniformly $\{0, \ldots, i-1\}$



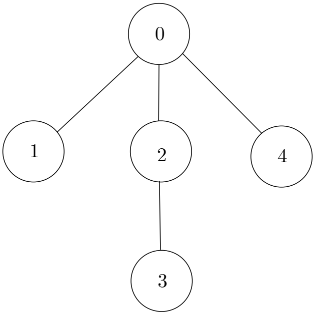
Uniform Random Recursive Trees

- Node labels: $\{0, \ldots, n\}$
- Node i picks one parent uniformly $\{0, \ldots, i-1\}$



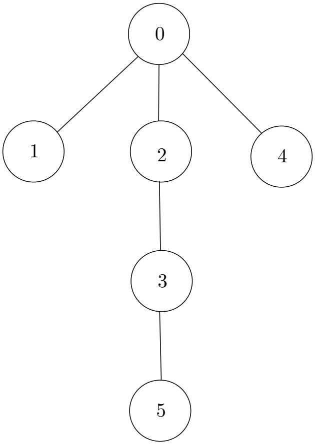
Uniform Random Recursive Trees

- Node labels: $\{0, \ldots, n\}$
- Node i picks one parent uniformly $\{0, \ldots, i-1\}$



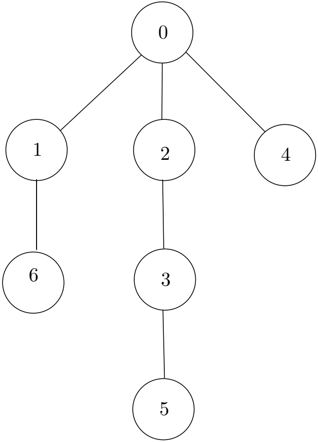
Uniform Random Recursive Trees

- Node labels: $\{0, \ldots, n\}$
- Node i picks one parent uniformly $\{0, \ldots, i-1\}$



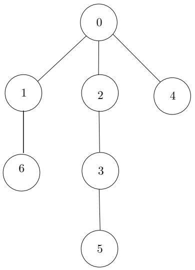
Uniform Random Recursive Trees

- Node labels: $\{0, \ldots, n\}$
- Node i picks one parent uniformly $\{0, \ldots, i-1\}$



Uniform Random Recursive Trees

Questions:



- D_n = depth of node n? $D_6 = 2$
- $H_n = \max_{0 \le i \le n} D_i$? $H_6 = 3$

Answers:

$$\boxed{\frac{D_n}{\log n} \xrightarrow{\mathcal{P}} 1}$$

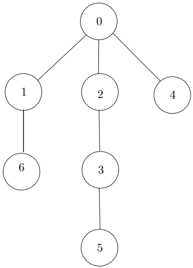
$$\boxed{\frac{H_n}{\log n} \xrightarrow{\mathcal{P}} e}$$

(Devroye 1987, Pittel 1994)

Uniform Random Recursive Trees

Questions:

- D_n = depth of node n? $D_6 = 2$
- $H_n = \max_{0 \leq i \leq n} D_i$? $H_6 = 3$



Answers:

$$\frac{D_n}{\log n} \xrightarrow{\mathcal{P}} 1$$

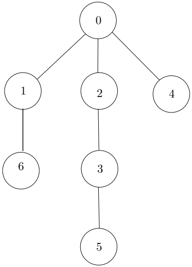
$$\frac{H_n}{\log n} \xrightarrow{\mathcal{P}} e$$

(Devroye 1987, Pittel 1994)

Uniform Random Recursive Trees

Questions:

- D_n = depth of node n? $D_6 = 2$
- $H_n = \max_{0 \leq i \leq n} D_i$? $H_6 = 3$



Answers:

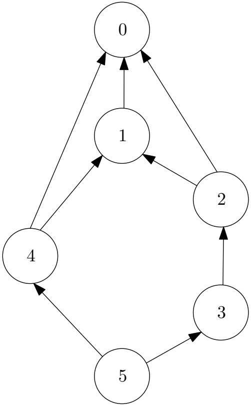
$$\boxed{\frac{D_n}{\log n} \xrightarrow{\mathcal{P}} 1}$$

$$\boxed{\frac{H_n}{\log n} \xrightarrow{\mathcal{P}} e}$$

(Devroye 1987, Pittel 1994)

Uniform Random Recursive DAGs

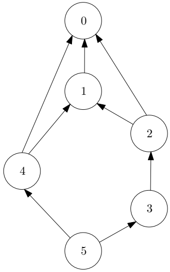
- Directed Acyclic Graph (DAG)
- Node labels: $\{0, \ldots, n\}$
- Node i picks $\underline{k}$ parents indep. unif. $\{0, \ldots, i-1\}$



Uniform Random Recursive DAG

Many paths from node i to the root

- Shortest/longest path
  (Tsukiji and Xhafa, 1996; Arya et al. 1999)
  (D'Souza et al. 2007)
  (Devroye and Janson, 2009)
- "Greedy" shortest path: follow parent with smallest label
  (Devroye and Janson, 2009)
  (Mahmoud, 2009)



Uniform Random Recursive DAG

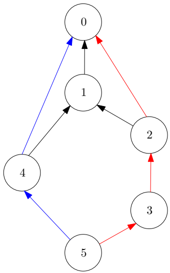
Many paths from node i to the root

- Shortest/longest path

  (Tsukiji and Xhafa, 1996; Arya et al. 1999)
  (D'Souza et al. 2007)
  (Devroye and Janson, 2009)

- "Greedy" shortest path: follow parent with smallest label

  (Devroye and Janson, 2009)
  (Mahmoud, 2009)



SARRTs

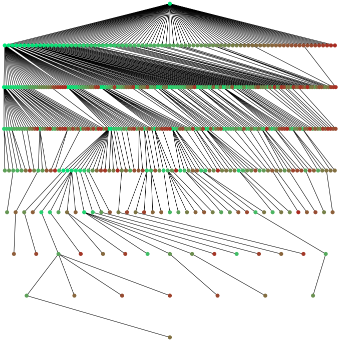
- Fix $X \in (0,1)$
- Node labels: $\{0, \ldots, n\}$
- Node i picks one parent $\lfloor iX_i \rfloor$ where $X_1, \ldots, X_n$ i.i.d.

Examples:

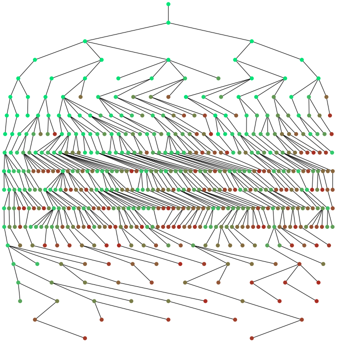
- $X = U$: Uniform RRT
- $X = \min(U_1, U_2)$: D_n = length of greedy shortest path from n to 0 in a 2-URRD

SARRTs

SARRT with $X = U^2$



SARRT with $X = U^{1/2}$



SARRTs: our results

Theorem (First obtained by Mahmoud (2009) for min/max of uniforms)

$$\frac{D_n}{\log n} \xrightarrow{\mathcal{P}} \frac{1}{\mathbf{E}\left\{-\log X\right\}}$$

Theorem

If X has a density, then there exists an explicit $\alpha_{\max}(X)$

$$\frac{H_n}{\log n} \xrightarrow{\mathcal{P}} \alpha_{\max}$$

Remark: no branching processes used

Outline

Introduction

Random recursive trees and DAGs

Scaled Attachment Random Recursive Trees (SARRT)

Main results

Proof sketches

Typical depth in a SARRT

Height of a SARRT

Summary and open problem

The ancestors as arrival times

Want to show:

$$\frac{D_n}{\log n} \xrightarrow{\mathcal{P}} \frac{1}{\mathbf{E}\{-\log X\}}$$

Ancestors of n:

$$n \to \lfloor nX_n \rfloor \to \lfloor \lfloor nX_n \rfloor X_{\lfloor nX_n \rfloor} \rfloor \to \ldots \to 0$$

$$D_n = \text{length}\left[n \to \lfloor nX_n \rfloor \to \lfloor \lfloor nX_n \rfloor X_{\lfloor nX_n \rfloor} \rfloor \to \cdots \to 0\right]$$
$$\approx \text{length}\left[n \to nX_n \to (nX_n)X_{\lfloor nX_n \rfloor} \to \cdots \to \leq 1\right]$$

Apply log:

$$D_n \approx \text{length}\left[\log n \to \log n + \log X_n \to \log n + \log X_n + \log X_{\lfloor nX_n \rfloor} \to \cdots \to \leq 0\right]$$

The ancestors as arrival times

Want to show:

$$\frac{D_n}{\log n} \xrightarrow{\mathcal{P}} \frac{1}{\mathbf{E}\{-\log X\}}$$

Ancestors of n:

$$n \to \lfloor nX_n \rfloor \to \lfloor \lfloor nX_n \rfloor X_{\lfloor nX_n \rfloor} \rfloor \to \ldots \to 0$$

$$\begin{aligned} D_n &= \text{length}\left[n \to \lfloor nX_n \rfloor \to \lfloor \lfloor nX_n \rfloor X_{\lfloor nX_n \rfloor} \rfloor \to \cdots \to 0 \right] \\ &\approx \text{length}\left[n \to nX_n \to (nX_n)X_{\lfloor nX_n \rfloor} \to \cdots \to \leq 1 \right] \end{aligned}$$

Apply log:

$$D_n \approx \text{length}\left[\log n \to \log n + \log X_n \to \log n + \log X_n + \log X_{\lfloor nX_n \rfloor} \to \cdots \to \leq 0 \right]$$

The ancestors as arrival times

Want to show:

$$\frac{D_n}{\log n} \xrightarrow{\mathcal{P}} \frac{1}{\mathbf{E}\{-\log X\}}$$

Ancestors of n:

$$n \to \lfloor nX_n \rfloor \to \lfloor \lfloor nX_n \rfloor X_{\lfloor nX_n \rfloor} \rfloor \to \ldots \to 0$$

$$\begin{aligned} D_n &= \text{length}\left[n \to \lfloor nX_n \rfloor \to \lfloor \lfloor nX_n \rfloor X_{\lfloor nX_n \rfloor} \rfloor \to \cdots \to 0\right] \\ &\approx \text{length}\left[n \to nX_n \to (nX_n)X_{\lfloor nX_n \rfloor} \to \cdots \to \leq 1\right] \end{aligned}$$

Apply log:

$$D_n \approx \text{length}\left[\log n \to \log n + \log X_n \to \log n + \log X_n + \log X_{\lfloor nX_n \rfloor} \to \cdots \to \leq 0\right]$$

The ancestors as arrival times

Want to show:

$$\frac{D_n}{\log n} \xrightarrow{\mathcal{P}} \frac{1}{\mathbf{E}\{-\log X\}}$$

Ancestors of n:

$$n \to \lfloor nX_n \rfloor \to \lfloor \lfloor nX_n \rfloor X_{\lfloor nX_n \rfloor} \rfloor \to \ldots \to 0$$

$$D_n = \text{length}\left[n \to \lfloor nX_n \rfloor \to \lfloor \lfloor nX_n \rfloor X_{\lfloor nX_n \rfloor} \rfloor \to \cdots \to 0 \right]$$
$$\approx \text{length}\left[n \to nX_n \to (nX_n)X_{\lfloor nX_n \rfloor} \to \cdots \to \leq 1 \right]$$

Apply log:

$$D_n \approx \text{length}\left[\log n \to \log n + \log X_n \to \log n + \log X_n + \log X_{\lfloor nX_n \rfloor} \to \cdots \to \leq 0 \right]$$

The ancestors as arrival times

Want to show:

$$\frac{D_n}{\log n} \xrightarrow{\mathcal{P}} \frac{1}{\mathbf{E}\{-\log X\}}$$

Ancestors of n:

$$n \to \lfloor nX_n \rfloor \to \lfloor \lfloor nX_n \rfloor X_{\lfloor nX_n \rfloor} \rfloor \to \dots \to 0$$

$$D_n = \text{length}\left[n \to \lfloor nX_n \rfloor \to \lfloor \lfloor nX_n \rfloor X_{\lfloor nX_n \rfloor} \rfloor \to \cdots \to 0\right]$$
$$\approx \text{length}\left[n \to nX_n \to (nX_n)X_{\lfloor nX_n \rfloor} \to \cdots \to \leq 1\right]$$

Apply log:

$$D_n \approx \text{length}\left[\log n \to \log n + \log X_n \to \log n + \log X_n + \log X_{\lfloor nX_n \rfloor} \to \cdots \to \leq 0\right]$$

The ancestors as arrival times

Want to show:

$$\frac{D_n}{\log n} \xrightarrow{\mathcal{P}} \frac{1}{\mathbf{E}\left\{-\log X\right\}}$$

Ancestors of n:

$$n \to \lfloor nX_n \rfloor \to \left\lfloor \lfloor nX_n \rfloor X_{\lfloor nX_n \rfloor} \right\rfloor \to \ldots \to 0$$

$$D_n = \text{length}\left[n \to \lfloor nX_n \rfloor \to \left\lfloor \lfloor nX_n \rfloor X_{\lfloor nX_n \rfloor} \right\rfloor \to \cdots \to 0\right]$$
$$\approx \text{length}\left[n \to nX_n \to (nX_n)X_{\lfloor nX_n \rfloor} \to \cdots \to \leq 1\right]$$

Apply log:

$$D_n \approx \text{length}\left[\log n \to \log n + \log X_n \to \log n + \log X_n + \log X_{\lfloor nX_n \rfloor} \to \cdots \to \leq 0\right]$$

The ancestors as arrival times

Want to show:

$$\frac{D_n}{\log n} \xrightarrow{\mathcal{P}} \frac{1}{\mathbf{E}\{-\log X\}}$$

Ancestors of n:

$$n \to \lfloor nX_n \rfloor \to \left\lfloor \lfloor nX_n \rfloor X_{\lfloor nX_n \rfloor} \right\rfloor \to \ldots \to 0$$

$$\begin{aligned} D_n &= \text{length}\left[n \to \lfloor nX_n \rfloor \to \left\lfloor \lfloor nX_n \rfloor X_{\lfloor nX_n \rfloor} \right\rfloor \to \cdots \to 0\right] \\ &\approx \text{length}\left[n \to nX_n \to (nX_n)X_{\lfloor nX_n \rfloor} \to \cdots \to \leq 1\right] \end{aligned}$$

Apply log:

$$D_n \approx \text{length}\left[\log n \to \log n + \log X_n \to \log n + \log X_n + \log X_{\lfloor nX_n \rfloor} \to \cdots \to \leq 0\right]$$

The ancestors as arrival times

$$\text{length}\left[\log n \to \log n + \log X_n \to \log n + \log X_n + \log X_{\lfloor nX_n \rfloor} \to \cdots \to \leq 0\right]$$

$$= \text{length}\left[0 \to -\log X_1 \to -\log X_1 - \log X_2 \to \cdots \to \geq \log n\right]$$

View as a sequence of arrivals:

$0 \to$	$-\log X_1 \to$	$-\log X_1 - \log X_2 \to$	$\cdots \to$	$\geq \log n$
	1st arrival	2nd arrival	$\ldots$	$N(\log n)$-th arrival

where $N(t) = \max\left\{m : \sum_{i=1}^{m} -\log X_i \leq t\right\}$ is a *renewal process*

$$\boxed{D_n \approx N(\log n)}$$

The ancestors as arrival times

$$\text{length}\left[\log n \to \log n + \log X_n \to \log n + \log X_n + \log X_{\lfloor nX_n \rfloor} \to \cdots \to \leq 0\right]$$
$$= \text{length}\left[0 \to -\log X_1 \to -\log X_1 - \log X_2 \to \cdots \to \geq \log n\right]$$

View as a sequence of arrivals:

$0 \to$	$-\log X_1 \to$	$-\log X_1 - \log X_2 \to$	$\cdots \to$	$\geq \log n$
	1st arrival	2nd arrival	...	$N(\log n)$-th arrival

where $N(t) = \max\left\{m : \sum_{i=1}^{m} -\log X_i \leq t\right\}$ is a *renewal process*

$$\boxed{D_n \approx N(\log n)}$$

Renewal theory gives D_n

Theorem (Renewal theorem)

$$\frac{N(t)}{t} \xrightarrow{\mathcal{P}} \frac{1}{\mathbf{E}\left\{-\log X\right\}} \qquad \text{as } t \to \infty$$

Recalling $D_n \approx N(\log n)$:

Theorem (Depth of a SARRT)

$$\frac{D_n}{\log n} \xrightarrow{\mathcal{P}} \frac{1}{\mathbf{E}\left\{-\log X\right\}}$$

Remark: can get a central limit theorem for D_n (also in Mahmoud, 2009)

The height of a SARRT

Theorem

If X has a density, then there exists an explicit $\alpha_{\max}(X)$

$$\frac{H_n}{\log n} \xrightarrow{\mathcal{P}} \alpha_{\max}$$

Recall $H_n = \max_{i=1,\dots,n} D_i$

This part of the talk: $\boxed{X = U} \Rightarrow \alpha_{\max} = e$ (URRT)

Plan:

- Upper bound (union bound)

$$\mathbf{P}\left\{H_n > (e + \varepsilon) \log n\right\} \to 0$$

- Lower bound (second moment)

$$\mathbf{P}\left\{H_n \geq (e - \varepsilon) \log n\right\} \to 1$$

The height of a SARRT

Theorem

If X has a density, then there exists an explicit $\alpha_{\max}(X)$

$$\frac{H_n}{\log n} \xrightarrow{\mathcal{P}} \alpha_{\max}$$

Recall $H_n = \max_{i=1,\dots,n} D_i$

This part of the talk: $\boxed{X = U} \Rightarrow \alpha_{\max} = e$ (URRT)

Plan:

- Upper bound (union bound)

$$\mathbf{P}\{H_n > (e+\varepsilon)\log n\} \to 0$$

- Lower bound (second moment)

$$\mathbf{P}\{H_n \geq (e-\varepsilon)\log n\} \to 1$$

The height of a SARRT

Theorem

If X has a density, then there exists an explicit $\alpha_{\max}(X)$

$$\frac{H_n}{\log n} \xrightarrow{\mathcal{P}} \alpha_{\max}$$

Recall $H_n = \max_{i=1,\ldots,n} D_i$

This part of the talk: $\boxed{X = U} \Rightarrow \alpha_{\max} = e$ (URRT)

Plan:

- Upper bound (union bound)

$$\mathbf{P}\{H_n > (e+\varepsilon)\log n\} \to 0$$

- Lower bound (second moment)

$$\mathbf{P}\{H_n \geq (e-\varepsilon)\log n\} \to 1$$

Height upper bound

Union bound:

$$\mathbf{P}\{H_n > t\} \le n \cdot \mathbf{P}\{D_n > t\}$$

Chernoff-type bound gives

$$\mathbf{P}\{D_n > c\log n\} \le n^{-1+c-c\log c}$$

Thus,

$$\begin{aligned}\mathbf{P}\{H_n > c\log n\} &\le n^{c-c\log c}\\ &\to 0 \qquad \text{for } c > e \quad \square\end{aligned}$$

Height lower bound: looking for deep nodes

i is *deep node*: $D_i > (e - \varepsilon) \log n$

Objective:

$$\mathbf{P}\{\exists \text{ deep node}\} \xrightarrow[n \to \infty]{} 1$$

Height lower bound: expected number of deep nodes

Recall Chernoff-type bound

$$\mathbf{P}\{D_n > c\log n\} \leq n^{-1+c-c\log c}$$

This is tight (using Cramér's Theorem):

$$\mathbf{P}\{D_n > c\log n\} \approx n^{-1+c-c\log c}$$

For $c = e - \varepsilon$,

$$\mathbf{P}\{D_n > (e-\varepsilon)\log n\} \geq n^{-1+\delta}$$

$$\boxed{\mathbf{E}\{|\{i : D_i > (e-\varepsilon)\log n\}|\} \geq n^{\delta}}$$

Height lower bound: second moment

We have

$$\mathbf{E}\{\#\text{ deep nodes}\} \to \infty$$

Does this imply $\mathbf{P}\{\#\text{ deep nodes} \geq 1\} \to 1$?

Yes if for $i \neq j$, [i deep node] *independent* of [j deep node]

1. Use Chung-Erdős inequality:

$$\mathbf{P}\{\exists\text{ deep node}\} \geq \frac{\mathbf{E}\{\#\text{ deep nodes}\}^2}{\sum_{i\neq j}\mathbf{P}\{i\text{ and }j\text{ deep}\} + \mathbf{E}\{\#\text{ deep nodes}\}}$$

2. Independence: $\mathbf{P}\{i\text{ and }j\text{ deep}\} = \mathbf{P}\{i\text{ deep}\}\,\mathbf{P}\{j\text{ deep}\}$

But: Enough to show

$$\mathbf{P}\{i\text{ and }j\text{ deep}\} \leq \mathbf{P}\{i\text{ deep}\}\,\mathbf{P}\{j\text{ deep}\} + o(\mathbf{P}\{i\text{ deep}\}\,\mathbf{P}\{j\text{ deep}\})$$

Height lower bound: second moment

We have

$$\mathbf{E}\{\#\text{ deep nodes}\} \to \infty$$

Does this imply $\mathbf{P}\{\#\text{ deep nodes} \geq 1\} \to 1$?

Yes if for $i \neq j$, [i deep node] *independent* of [j deep node]

1. Use Chung-Erdős inequality:

$$\mathbf{P}\{\exists\text{ deep node}\} \geq \frac{\mathbf{E}\{\#\text{ deep nodes}\}^2}{\sum_{i \neq j} \mathbf{P}\{i \text{ and } j \text{ deep}\} + \mathbf{E}\{\#\text{ deep nodes}\}}$$

2. Independence: $\mathbf{P}\{i \text{ and } j \text{ deep}\} = \mathbf{P}\{i \text{ deep}\}\mathbf{P}\{j \text{ deep}\}$

   But: Enough to show

$$\mathbf{P}\{i \text{ and } j \text{ deep}\} \leq \mathbf{P}\{i \text{ deep}\}\mathbf{P}\{j \text{ deep}\} + o(\mathbf{P}\{i \text{ deep}\}\mathbf{P}\{j \text{ deep}\})$$

Height lower bound: second moment

We have

$$\mathbf{E}\{\# \text{ deep nodes}\} \to \infty$$

Does this imply $\mathbf{P}\{\# \text{ deep nodes} \geq 1\} \to 1$?

Yes if for $i \neq j$, [i deep node] *independent* of [j deep node]

1. Use Chung-Erdős inequality:

$$\mathbf{P}\{\exists \text{ deep node}\} \geq \frac{\mathbf{E}\{\# \text{ deep nodes}\}^2}{\sum_{i \neq j} \mathbf{P}\{i \text{ and } j \text{ deep}\} + \mathbf{E}\{\# \text{ deep nodes}\}}$$

2. Independence: $\mathbf{P}\{i \text{ and } j \text{ deep}\} = \mathbf{P}\{i \text{ deep}\}\mathbf{P}\{j \text{ deep}\}$

But: Enough to show

$$\mathbf{P}\{i \text{ and } j \text{ deep}\} \leq \mathbf{P}\{i \text{ deep}\}\mathbf{P}\{j \text{ deep}\} + o(\mathbf{P}\{i \text{ deep}\}\mathbf{P}\{j \text{ deep}\})$$

Height lower bound: second moment

We have

$$\mathbf{E}\{\#\text{ deep nodes}\} \to \infty$$

Does this imply $\mathbf{P}\{\#\text{ deep nodes} \geq 1\} \to 1$?

Yes if for $i \neq j$, [i deep node] *independent* of [j deep node]

1. Use Chung-Erdős inequality:

$$\mathbf{P}\{\exists\text{ deep node}\} \geq \frac{\mathbf{E}\{\#\text{ deep nodes}\}^2}{\sum_{i \neq j} \mathbf{P}\{i\text{ and }j\text{ deep}\} + \mathbf{E}\{\#\text{ deep nodes}\}}$$

2. Independence: $\mathbf{P}\{i\text{ and }j\text{ deep}\} = \mathbf{P}\{i\text{ deep}\}\,\mathbf{P}\{j\text{ deep}\}$

But: Enough to show

$$\mathbf{P}\{i\text{ and }j\text{ deep}\} \leq \mathbf{P}\{i\text{ deep}\}\,\mathbf{P}\{j\text{ deep}\} + o(\mathbf{P}\{i\text{ deep}\}\,\mathbf{P}\{j\text{ deep}\})$$

Height lower bound: second moment

We have

$$\mathbf{E}\{\#\text{ deep nodes}\} \to \infty$$

Does this imply $\mathbf{P}\{\#\text{ deep nodes} \geq 1\} \to 1$?

Yes if for $i \neq j$, [i deep node] *independent* of [j deep node]

1. Use Chung-Erdős inequality:

$$\mathbf{P}\{\exists\text{ deep node}\} \geq \frac{\mathbf{E}\{\#\text{ deep nodes}\}^2}{\sum_{i\neq j}\mathbf{P}\{i\text{ and }j\text{ deep}\} + \mathbf{E}\{\#\text{ deep nodes}\}}$$

2. ~~Independence: $\mathbf{P}\{i\text{ and }j\text{ deep}\} = \mathbf{P}\{i\text{ deep}\}\mathbf{P}\{j\text{ deep}\}$~~

But: Enough to show

$$\mathbf{P}\{i\text{ and }j\text{ deep}\} \leq \mathbf{P}\{i\text{ deep}\}\mathbf{P}\{j\text{ deep}\} + o(\mathbf{P}\{i\text{ deep}\}\mathbf{P}\{j\text{ deep}\})$$

Height lower bound: second moment

We have

$$\mathbf{E}\{\#\text{ deep nodes}\} \to \infty$$

Does this imply $\mathbf{P}\{\#\text{ deep nodes} \geq 1\} \to 1$?

Yes if for $i \neq j$, [i deep node] *independent* of [j deep node]

1. Use Chung-Erdős inequality:

$$\mathbf{P}\{\exists\text{ deep node}\} \geq \frac{\mathbf{E}\{\#\text{ deep nodes}\}^2}{\sum_{i\neq j}\mathbf{P}\{i\text{ and }j\text{ deep}\} + \mathbf{E}\{\#\text{ deep nodes}\}}$$

2. ~~Independence: $\mathbf{P}\{i\text{ and }j\text{ deep}\} = \mathbf{P}\{i\text{ deep}\}\mathbf{P}\{j\text{ deep}\}$~~

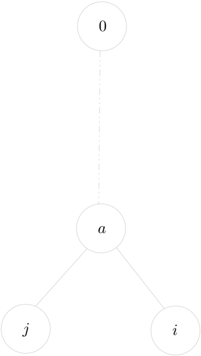
   But: Enough to show

$$\mathbf{P}\{i\text{ and }j\text{ deep}\} \leq \mathbf{P}\{i\text{ deep}\}\mathbf{P}\{j\text{ deep}\} + o(\mathbf{P}\{i\text{ deep}\}\mathbf{P}\{j\text{ deep}\})$$

Height lower bound: joint probability

Want to bound $\mathbf{P}\{i \text{ and } j \text{ deep}\}$

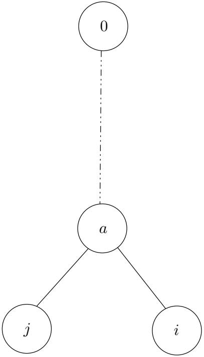
Problem: collisions!



Height lower bound: joint probability

Want to bound $\mathbf{P}\{i \text{ and } j \text{ deep}\}$

Problem: collisions!



Height lower bound: introducing good deep nodes

Notation: $L(i,t)$ = t-th ancestor of i

i deep node: $L(i,t) \geq 1$ where $t = (e - \varepsilon) \log n$

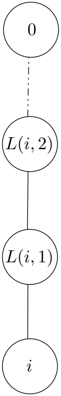
To be able to bound collision probabilities:
New definition: *i good deep node*

$$\forall p = 1, \ldots, t : L(i,p) \geq i \exp\left(-\frac{p}{e-\varepsilon}\right)$$

We have

$$\mathbf{P}\{i \text{ good deep node}\} \approx \mathbf{P}\{i \text{ deep node}\}$$

Similar idea in Devroye and Reed (1995)



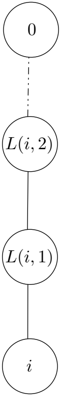
Height lower bound: introducing good deep nodes

Notation: $L(i, t)$ = t-th ancestor of i

i deep node: $L(i, t) \geq 1$ where $t = (e - \varepsilon) \log n$

To be able to bound collision probabilities:
New definition: i *good deep node*

$$\forall p = 1, \ldots, t : L(i, p) \geq i \exp\left(-\frac{p}{e - \varepsilon}\right)$$



We have

$$\mathbf{P}\{i \text{ good deep node}\} \approx \mathbf{P}\{i \text{ deep node}\}$$

Similar idea in Devroye and Reed (1995)

Height lower bound: introducing good deep nodes

Notation: $L(i, t)$ = t-th ancestor of i

i deep node: $L(i, t) \geq 1$ where $t = (e - \varepsilon) \log n$

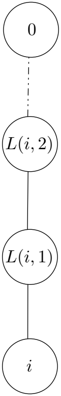
To be able to bound collision probabilities:
New definition: i *good deep node*

$$\forall p = 1, \ldots, t : L(i, p) \geq i \exp\left(-\frac{p}{e - \varepsilon}\right)$$

We have

$$\mathbf{P}\{i \text{ good deep node}\} \approx \mathbf{P}\{i \text{ deep node}\}$$

Similar idea in Devroye and Reed (1995)

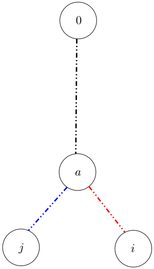


Height lower bound: joint probability

By bounding the collision probability:

$$
\begin{aligned}
&\mathbf{P}\{i \text{ and } j \text{ good deep}\} \\
&= \mathbf{P}\{\text{no collision at} \leq t \text{ and } i, j \text{ good deep}\} \\
&+ \sum_{p=1}^{t} \mathbf{P}\{\text{collision at } p \text{ and } i, j \text{ good deep}\} \\
&\vdots \\
&\sim \mathbf{P}\{i \text{ good deep}\} \mathbf{P}\{j \text{ good deep}\}
\end{aligned}
$$

for well chosen parameters □



Outline

Introduction

Random recursive trees and DAGs

Scaled Attachment Random Recursive Trees (SARRT)

Main results

Proof sketches

Typical depth in a SARRT

Height of a SARRT

Summary and open problem

Summary

- SARRT with attachment X: node i picks one parent $\lfloor iX_i \rfloor$
- URRT = SARRT with $X = U$

Theorem

$$\frac{D_n}{\log n} \xrightarrow{\mathcal{P}} \frac{1}{\mathbf{E}\{-\log X\}}$$

Theorem

If X has a density, then there exists an explicit $\alpha_{\max}(X)$

$$\frac{H_n}{\log n} \xrightarrow{\mathcal{P}} \alpha_{\max}$$

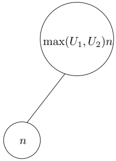
Application: distances for 2-URRD

- Five natural distances
- For each one: D_n, $\max_{1 \le i \le n} D_i$ and $\min_{n/2 \le i \le n} D_i$
- The table of constants $\frac{\square_n}{\log n} \xrightarrow{\mathcal{P}} \alpha$

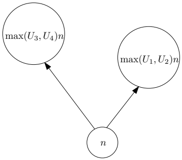
	$\min_{n/2 \le i \le n} D_i$	D_n	$\max_{1 \le i \le n} D_i$
Shortest (DJ09)	0	0.373...	0.373...
Greedy shortest	0	2/3	1.673...
Random	0	1	e
Greedy longest	0.373...	2	4.311...
Longest	2.155...(BF)	4.311...	2e (TX96)

Open problem

- Random binary search tree (RBST)



- Greedy best-of-two choices RBST



- Height same as SARRT with
  $X = \min(\max(U_1, U_2), \max(U_3, U_4))$?

Extra

Definition of $\alpha_{\max}$

$$\alpha_{\max} = \inf \left\{ c > \frac{1}{\mathbf{E}\left\{-\log X\right\}} : \ \Psi(c) > 1 \right\}$$

where

$$\Psi(c) = c \cdot \sup_{\lambda \in \mathbb{R}} \left\{ -\frac{\lambda}{c} - \log \mathbf{E}\left\{ X^{\lambda} \right\} \right\}$$