On the number of subtrees on the fringe of random trees

(partly joined with Huilan Chang)

Michael Fuchs

Institute of Applied Mathematics
National Chiao Tung University

Hsinchu, Taiwan

April 16th, 2008

The number of subtrees

$X_{n,k}$ =number of subtrees of size k on the fringe of random binary search trees of size n.

The number of subtrees

$X_{n,k}$ =number of subtrees of size k on the fringe of random binary search trees of size n.

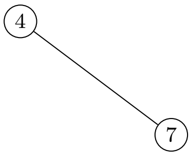
Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$

4

The number of subtrees

$X_{n,k}$ =number of subtrees of size k on the fringe of random binary search trees of size n.

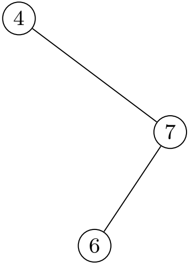
Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$



The number of subtrees

$X_{n,k}$ =number of subtrees of size k on the fringe of random binary search trees of size n.

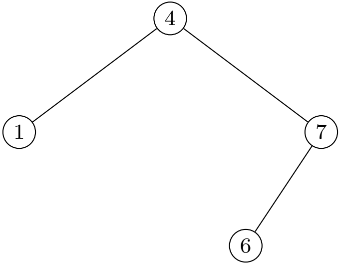
Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$



The number of subtrees

$X_{n,k}$ =number of subtrees of size k on the fringe of random binary search trees of size n.

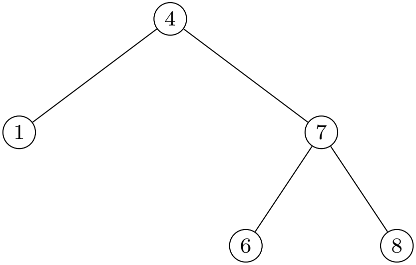
Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$



The number of subtrees

$X_{n,k}$ =number of subtrees of size k on the fringe of random binary search trees of size n.

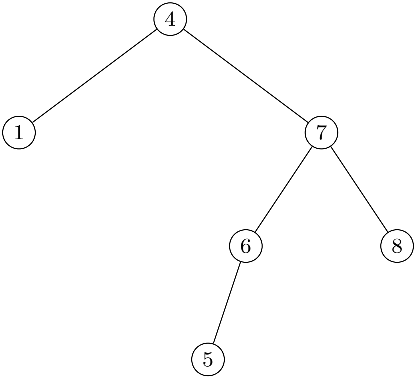
Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$



The number of subtrees

$X_{n,k}$ =number of subtrees of size k on the fringe of random binary search trees of size n.

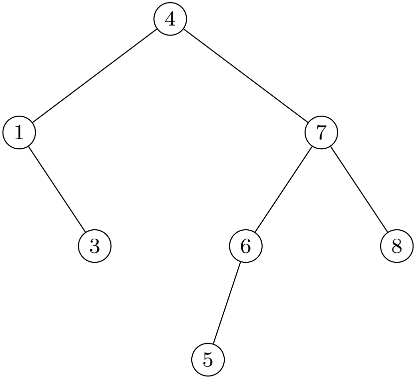
Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$



The number of subtrees

$X_{n,k}$ =number of subtrees of size k on the fringe of random binary search trees of size n.

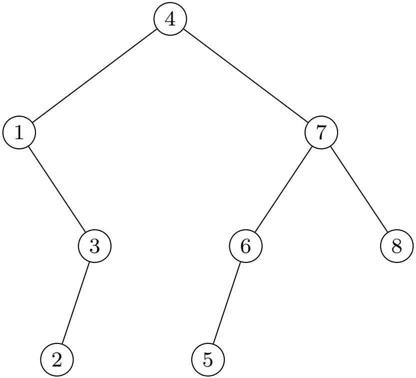
Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$



The number of subtrees

$X_{n,k}$ =number of subtrees of size k on the fringe of random binary search trees of size n.

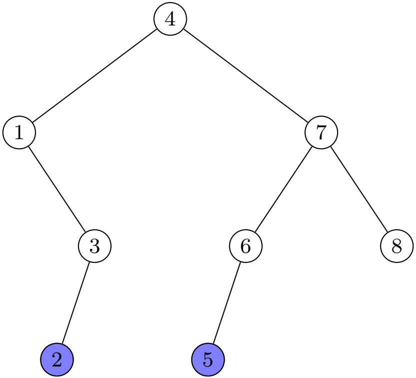
Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$



The number of subtrees

$X_{n,k}$ =number of subtrees of size k on the fringe of random binary search trees of size n.

Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$

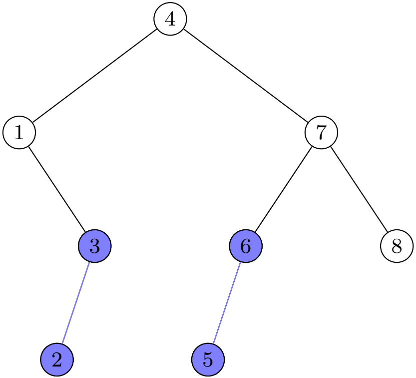


$X_{8,1} = 2$

The number of subtrees

$X_{n,k}$ =number of subtrees of size k on the fringe of random binary search trees of size n.

Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$



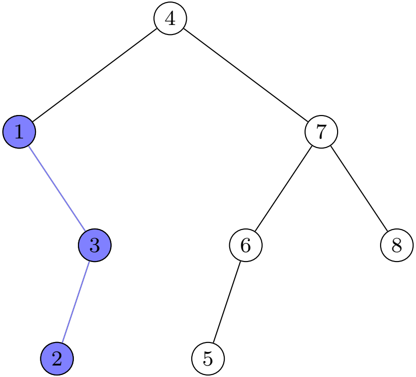
$X_{8,1} = 2$

$X_{8,2} = 2$

The number of subtrees

$X_{n,k}$ =number of subtrees of size k on the fringe of random binary search trees of size n.

Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$

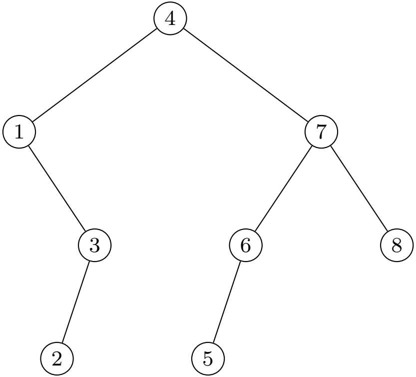


$$X_{8,1} = 2$$
$$X_{8,2} = 2$$
$$X_{8,3} = 1$$

The number of subtrees

$X_{n,k}$ =number of subtrees of size k on the fringe of random binary search trees of size n.

Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$

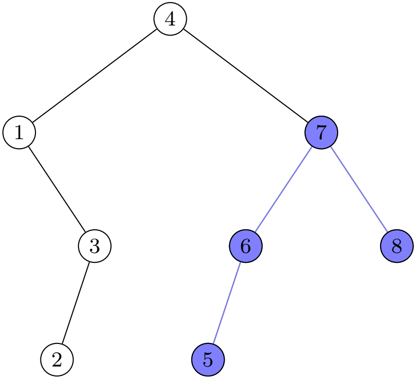


$X_{8,1} = 2$
$X_{8,2} = 2$
$X_{8,3} = 1$
$X_{8,4} = 0$

The number of subtrees

$X_{n,k}$ =number of subtrees of size k on the fringe of random binary search trees of size n.

Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$

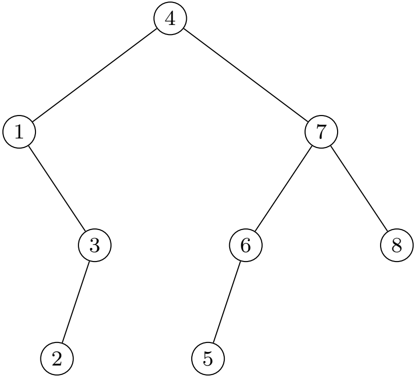


$$X_{8,1} = 2$$
$$X_{8,2} = 2$$
$$X_{8,3} = 1$$
$$X_{8,4} = 0$$
$$X_{8,5} = 1$$

The number of subtrees

$X_{n,k}$ =number of subtrees of size k on the fringe of random binary search trees of size n.

Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$

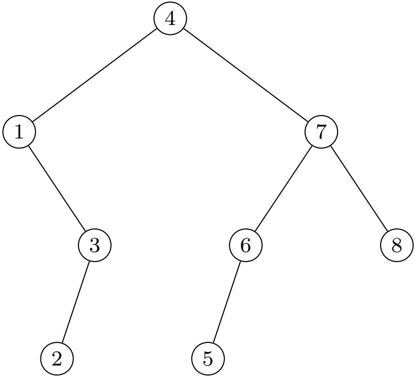


$$X_{8,1} = 2$$
$$X_{8,2} = 2$$
$$X_{8,3} = 1$$
$$X_{8,4} = 0$$
$$X_{8,5} = 1$$
$$X_{8,6} = 0$$

The number of subtrees

$X_{n,k}$ =number of subtrees of size k on the fringe of random binary search trees of size n.

Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$

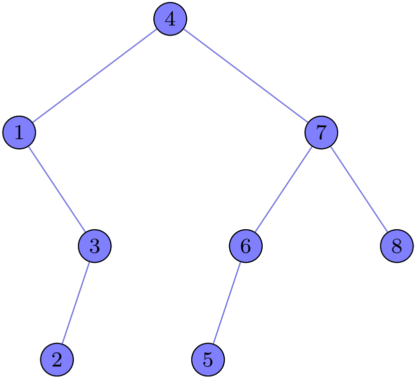


$$X_{8,1} = 2$$
$$X_{8,2} = 2$$
$$X_{8,3} = 1$$
$$X_{8,4} = 0$$
$$X_{8,5} = 1$$
$$X_{8,6} = 0$$
$$X_{8,7} = 0$$

The number of subtrees

$X_{n,k}$ =number of subtrees of size k on the fringe of random binary search trees of size n.

Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$



$$X_{8,1} = 2$$
$$X_{8,2} = 2$$
$$X_{8,3} = 1$$
$$X_{8,4} = 0$$
$$X_{8,5} = 1$$
$$X_{8,6} = 0$$
$$X_{8,7} = 0$$
$$X_{8,8} = 1$$

Mean value and variance

$X_{n,k}$ satisfies

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

where $X_{k,k} = 1$, $X_{I_n,k}$ and $X^*_{n-1-I_n,k}$ are conditionally independent given I_n, and $I_n = \text{Unif}\{0, \ldots, n-1\}$.

Mean value and variance

$X_{n,k}$ satisfies

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

where $X_{k,k} = 1$, $X_{I_n,k}$ and $X^*_{n-1-I_n,k}$ are conditionally independent given I_n, and $I_n = \text{Unif}\{0, \ldots, n-1\}$.

This yields

$$\mu_{n,k} := \mathbf{E}(X_{n,k}) = \frac{2(n+1)}{(k+1)(k+2)}, \qquad (n > k),$$

and

$$\sigma^2_{n,k} := \mathrm{Var}(X_{n,k}) = \frac{2k(4k^2+5k-3)(n+1)}{(k+1)(k+2)^2(2k+1)(2k+3)}$$

for $n > 2k+1$.

Some previous results

- Aldous (1991): Weak law of large numbers

$$\frac{X_{n,k}}{\mu_{n,k}} \longrightarrow 1 \qquad \text{in probability.}$$

Some previous results

- Aldous (1991): Weak law of large numbers

$$\frac{X_{n,k}}{\mu_{n,k}} \longrightarrow 1 \qquad \text{in probability.}$$

- Devroye (1991): Central limit theorem

$$\frac{X_{n,k} - \mu_{n,k}}{\sigma_{n,k}} \xrightarrow{d} \mathcal{N}(0,1).$$

Some previous results

- Aldous (1991): Weak law of large numbers

$$\frac{X_{n,k}}{\mu_{n,k}} \longrightarrow 1 \qquad \text{in probability.}$$

- Devroye (1991): Central limit theorem

$$\frac{X_{n,k} - \mu_{n,k}}{\sigma_{n,k}} \xrightarrow{d} \mathcal{N}(0,1).$$

- Flajolet, Gourdon, Martinez (1997):

  Central limit theorem with optimal Berry-Esseen bound and LLT

Some previous results

- Aldous (1991): Weak law of large numbers

$$\frac{X_{n,k}}{\mu_{n,k}} \longrightarrow 1 \qquad \text{in probability.}$$

- Devroye (1991): Central limit theorem

$$\frac{X_{n,k} - \mu_{n,k}}{\sigma_{n,k}} \xrightarrow{d} \mathcal{N}(0,1).$$

- Flajolet, Gourdon, Martinez (1997):

  Central limit theorem with optimal Berry-Esseen bound and LLT

$\longrightarrow$ All the above results are for fixed k.

Results for $k = k_n$

Theorem (Feng, Mahmoud, Panholzer (2008))

(i) *(Normal range) Let* $k = o\left(\sqrt{n}\right)$ *and* $k \to \infty$ *as* $n \to \infty$. *Then,*

$$\frac{X_{n,k} - \mu_{n,k}}{\sqrt{2n/k^2}} \xrightarrow{d} \mathcal{N}(0,1).$$

(ii) *(Poisson range) Let* $k \sim c\sqrt{n}$ *as* $n \to \infty$. *Then,*

$$X_{n,k} \xrightarrow{d} \mathrm{Poisson}(2c^{-2}).$$

(iii) *(Degenerate range) Let* $k < n$ *and* $\sqrt{n} = o(k)$ *as* $n \to \infty$. *Then,*

$$X_{n,k} \xrightarrow{L_1} 0.$$

Why are we interested in $X_{n,k}$?

- $X_{n,k}$ is a new kind of profile of a tree.

Why are we interested in $X_{n,k}$?

- $X_{n,k}$ is a new kind of profile of a tree.
- The phase change from normal to Poisson is a universal phenomenon expected to hold for many classes of random trees.

Why are we interested in $X_{n,k}$?

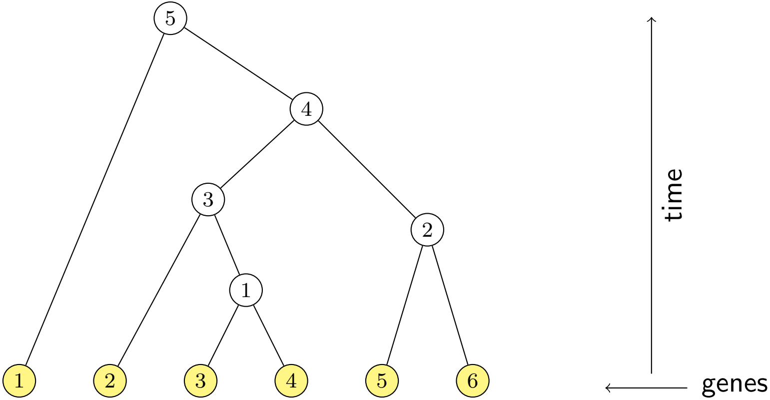
- $X_{n,k}$ is a new kind of profile of a tree.
- The phase change from normal to Poisson is a universal phenomenon expected to hold for many classes of random trees.
- The methods for proving phase change results might be applicable to other parameters which are expected to exhibit the same phase change behavior as well.

Why are we interested in $X_{n,k}$?

- $X_{n,k}$ is a new kind of profile of a tree.
- The phase change from normal to Poisson is a universal phenomenon expected to hold for many classes of random trees.
- The methods for proving phase change results might be applicable to other parameters which are expected to exhibit the same phase change behavior as well.
- $X_{n,k}$ is related to parameters arising in genetics.

Yule generated random genealogical trees

Example:



Yule generated random genealogical trees

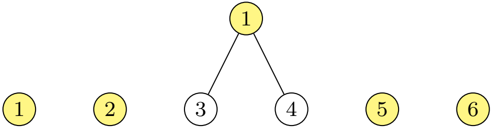
Example:

Random model:

At every time point, two yellow nodes uniformly coalescent.

Yule generated random genealogical trees

Example:

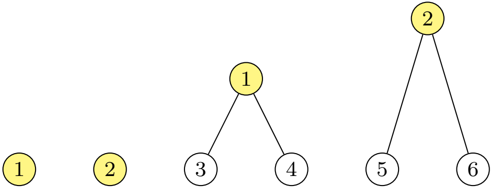


Random model:

At every time point, two yellow nodes uniformly coalescent.

Yule generated random genealogical trees

Example:

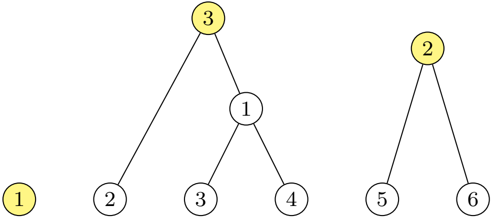


Random model:

At every time point, two yellow nodes uniformly coalescent.

Yule generated random genealogical trees

Example:

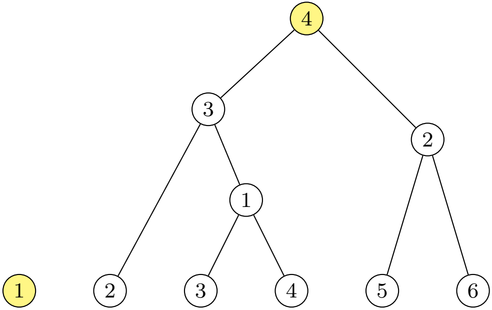


Random model:

At every time point, two yellow nodes uniformly coalescent.

Yule generated random genealogical trees

Example:

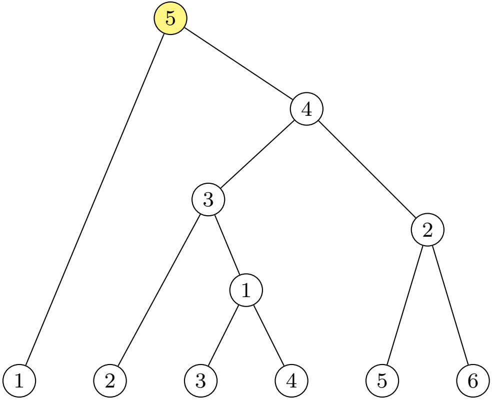


Random model:

At every time point, two yellow nodes uniformly coalescent.

Yule generated random genealogical trees

Example:

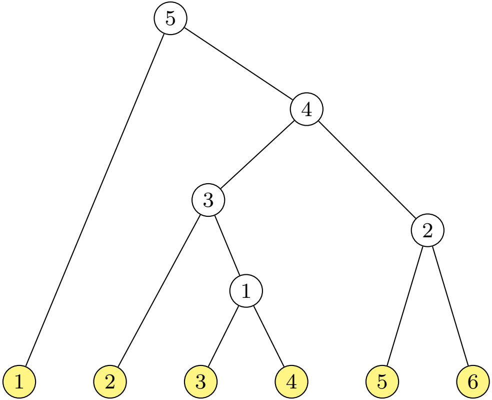


Random model:

At every time point, two yellow nodes uniformly coalescent.

Yule generated random genealogical trees

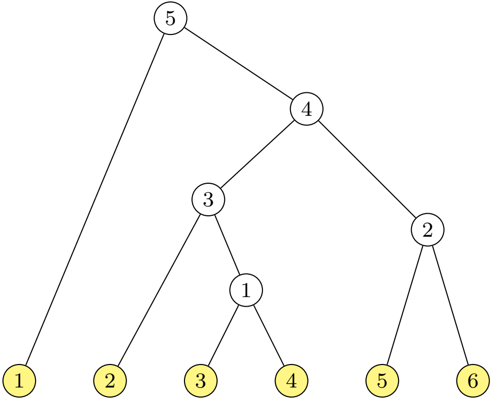
Example:



Random model:

At every time point, two yellow nodes uniformly coalescent.

Example:



Random model:

At every time point, two yellow nodes uniformly coalescent.

Same model as random binary search tree model!

Shape parameters of genealogical trees

- k-pronged nodes (Rosenberg 2006):

  Nodes with an induced subtree with $k-1$ internal nodes.

Shape parameters of genealogical trees

- k-pronged nodes (Rosenberg 2006):

  Nodes with an induced subtree with $k-1$ internal nodes.

- k-caterpillars (Rosenberg 2006):

  Correspond to nodes whose induced subtree has $k-1$ internal nodes all of them with out-degree either 0 or 1.

Shape parameters of genealogical trees

- k-pronged nodes (Rosenberg 2006):

  Nodes with an induced subtree with $k-1$ internal nodes.

- k-caterpillars (Rosenberg 2006):

  Correspond to nodes whose induced subtree has $k-1$ internal nodes all of them with out-degree either 0 or 1.

- Nodes with minimal clade size k (Blum and François (2005)):

  If $k \geq 3$, then they are internal nodes with induced subtree of size $k-1$ and either an empty right subtree or empty left subtree.

Counting pattern in random binary search trees

Consider $X_{n,k}$ with

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

where $X_{k,k} = \text{Bernoulli}(p_k)$, $X_{I_n,k}$ and $X^*_{n-1-I_n,k}$ are conditionally independent given I_n, and $I_n = \text{Unif}\{0, \ldots, n-1\}$.

Counting pattern in random binary search trees

Consider $X_{n,k}$ with

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

where $X_{k,k} = \text{Bernoulli}(p_k)$, $X_{I_n,k}$ and $X^*_{n-1-I_n,k}$ are conditionally independent given I_n, and $I_n = \text{Unif}\{0, \dots, n-1\}$.

Then,

p_k	shape parameter
1	# of $k+1$-pronged nodes
$2/k$	# of nodes with minimal clade size $k+1$
$2^{k-1}/k!$	# of $k+1$ caterpillars

Underlying recurrence and solution

All (centered or non-centered) moments satisfy

$$a_{n,k} = \frac{2}{n} \sum_{j=0}^{n-1} a_{j,k} + b_{n,k},$$

where $a_{k,k}$ is given and $a_{n,k} = 0$ for $n < k$.

Underlying recurrence and solution

All (centered or non-centered) moments satisfy

$$a_{n,k} = \frac{2}{n}\sum_{j=0}^{n-1} a_{j,k} + b_{n,k},$$

where $a_{k,k}$ is given and $a_{n,k} = 0$ for $n < k$.

We have

$$a_{n,k} = \frac{2(n+1)}{(k+1)(k+2)} a_{k,k} + 2(n+1)\sum_{k<j<n} \frac{b_{j,k}}{(j+1)(j+2)} + b_{n,k},$$

where $n > k$.

Mean value and variance

We have

$$\mathbf{E}(X_{n,k}) = \frac{2(n+1)}{(k+1)(k+2)}p_k, \qquad (n > k),$$

and

$$\mathrm{Var}(X_{n,k}) = \frac{2p_k(4k^3+16k^2+19k+6-(11k^2+22k+6)p_k)(n+1)}{(k+1)(k+2)^2(2k+1)(2k+3)}$$

for $n > 2k+1$.

Mean value and variance

We have

$$\mathbf{E}(X_{n,k}) = \frac{2(n+1)}{(k+1)(k+2)} p_k, \qquad (n > k),$$

and

$$\mathrm{Var}(X_{n,k}) = \frac{2p_k(4k^3+16k^2+19k+6-(11k^2+22k+6)p_k)(n+1)}{(k+1)(k+2)^2(2k+1)(2k+3)}$$

for $n > 2k+1$.

Note that

$$\mathbf{E}(X_{n,k}) \sim \mathrm{Var}(X_{n,k}) \sim \frac{2p_k}{k^2} n$$

for $n > 2k+1$ and $k \to \infty$ as $n \to \infty$.

Higher moments

Denote by

$$A_{n,k}^{(m)} := \mathbf{E}(X_{n,k} - \mathbf{E}(X_{n,k}))^m.$$

Higher moments

Denote by

$$A_{n,k}^{(m)} := \mathbf{E}(X_{n,k} - \mathbf{E}(X_{n,k}))^m.$$

Then,

$$A_{n,k}^{(m)} = \frac{2}{n}\sum_{j=0}^{n-1} A_{j,k}^{(m)} + B_{n,k}^{(m)},$$

where

$$B_{n,k}^{(m)} := \sum_{\substack{i_1+i_2+i_3=m\\ 0\le i_1,i_2<m}} \binom{m}{i_1, i_2, i_3} \frac{1}{n}\sum_{j=0}^{n-1} A_{j,k}^{(i_1)} A_{n-1-j,k}^{(i_2)} \Delta_{n,j,k}^{i_3}$$

and

$$\Delta_{n,j,k} = \mathbf{E}(X_{j,k}) + \mathbf{E}(X_{n-1-j,k}) - \mathbf{E}(X_{n,k}).$$

Methods of moments

Theorem

Let Z_n, Z be random variables. Assume that, as $n \to \infty$,

$$\mathbf{E}(Z_n^m) \longrightarrow \mathbf{E}(Z^m)$$

for all $m \geq 1$ and Z is uniquely determined by its moment sequence. Then,

$$Z_n \xrightarrow{d} Z.$$

Methods of moments

Theorem

Let Z_n, Z be random variables. Assume that, as $n \to \infty$,

$$\mathbf{E}(Z_n^m) \longrightarrow \mathbf{E}(Z^m)$$

for all $m \geq 1$ and Z is uniquely determined by its moment sequence. Then,

$$Z_n \xrightarrow{d} Z.$$

Inductive approach based on asymptotic transfers was extensively used in the AofA to prove numerous result ("moment pumping").

Methods of moments

Theorem

Let Z_n, Z be random variables. Assume that, as $n \to \infty$,

$$\mathbf{E}(Z_n^m) \longrightarrow \mathbf{E}(Z^m)$$

for all $m \geq 1$ and Z is uniquely determined by its moment sequence. Then,

$$Z_n \xrightarrow{d} Z.$$

Inductive approach based on asymptotic transfers was extensively used in the AofA to prove numerous result ("moment pumping").

Fuchs, Hwang, Neininger (2007): variation of the above scheme to study the profile of random binary search trees and random recursive trees.

Normal range

Proposition

Uniformly for $n, k, m \geq 1$ *and* $n > k$

$$A_{n,k}^{(m)} = \max\left\{\frac{2p_k n}{k^2}, \left(\frac{2p_k n}{k^2}\right)^{m/2}\right\}.$$

Normal range

Proposition

Uniformly for $n, k, m \geq 1$ *and* $n > k$

$$A_{n,k}^{(m)} = \max\left\{\frac{2p_k n}{k^2}, \left(\frac{2p_k n}{k^2}\right)^{m/2}\right\}.$$

Proposition

For $\mathbf{E}(X_{n,k}) \to \infty$ *as* $n \to \infty$,

$$A_{n,k}^{(2m-1)} = o\left(\left(\frac{2p_k n}{k^2}\right)^{m-1/2}\right), \qquad A_{n,k}^{(2m)} \sim g_m \left(\frac{2p_k n}{k^2}\right)^{m},$$

where

$$g_m = (2m)!/(2^m m!).$$

Poisson range

Consider

$$\bar{A}_{n,k}^{(m)} = \mathbf{E}(X_{n,k}(X_{n,k}-1)\cdots(X_{n,k}-m+1)).$$

Poisson range

Consider

$$\bar{A}_{n,k}^{(m)} = \mathbf{E}(X_{n,k}(X_{n,k}-1)\cdots(X_{n,k}-m+1)).$$

Then, similarly as before:

Proposition

(i) *Uniformly for* $n, k, m \geq 1$ *and* $n > k$

$$\bar{A}_{n,k}^{(m)} = \max\left\{\frac{2p_k n}{k^2}, \left(\frac{2p_k n}{k^2}\right)^m\right\}.$$

(ii) *For* $\mathbf{E}(X_{n,k}) \to c$ *and* $k < n$ *as* $n \to \infty$,

$$\bar{A}_{n,k}^{(m)} \longrightarrow c^m.$$

The phase change

Theorem

(i) *(Normal range) Let* $\mathbf{E}(X_{n,k}) \to \infty$ *and* $k \to \infty$ *as* $n \to \infty$*. Then,*

$$\frac{X_{n,k} - \mathbf{E}(X_{n,k})}{\sqrt{2p_k n/k^2}} \xrightarrow{d} \mathcal{N}(0,1).$$

(ii) *(Poisson range) Let* $\mathbf{E}(X_{n,k}) \to c > 0$ *and* $k < n$ *as* $n \to \infty$*. Then,*

$$X_{n,k} \xrightarrow{d} \mathrm{Poisson}(c).$$

(iii) *(Degenerate range) Let* $\mathbf{E}(X_{n,k}) \to 0$ *as* $n \to \infty$*. Then,*

$$X_{n,k} \xrightarrow{L_1} 0.$$

A comparison of the phase change

For k-caterpillars, we have

$$\mathbf{E}(X_{n,k}) = \frac{2^{k-1}n}{(k+2)!}.$$

Note that either

$$\mathbf{E}(X_{n,k}) \to \infty \qquad \text{or} \qquad \mathbf{E}(X_{n,k}) \to 0.$$

So, there is no Poisson range.

A comparison of the phase change

For k-caterpillars, we have

$$\mathbf{E}(X_{n,k}) = \frac{2^{k-1}n}{(k+2)!}.$$

Note that either

$$\mathbf{E}(X_{n,k}) \to \infty \qquad \text{or} \qquad \mathbf{E}(X_{n,k}) \to 0.$$

So, there is no Poisson range.

shape parameter	location	phase change
k-pronged nodes	$\sqrt{n}$	normal - poisson - degenerate
minimal clade size k	$\sqrt[3]{n}$	normal - poisson - degenerate
k-caterpillars	$\ln n/(\ln\ln n)$	normal - degenerate

Refined results (for # of subtrees)

Define

$$\phi_{n,k}(y) = e^{-\sigma_{n,k}^2 y^2/2}\mathbf{E}\left(e^{(X_{n,k}-\mu_{n,k})y}\right).$$

and

$$\phi_{n,k}^{(m)} = \left.\frac{d^m \phi_{n,k}(y)}{dy^m}\right|_{y=0}.$$

Refined results (for # of subtrees)

Define

$$\phi_{n,k}(y) = e^{-\sigma_{n,k}^2 y^2/2}\mathbf{E}\left(e^{(X_{n,k}-\mu_{n,k})y}\right).$$

and

$$\phi_{n,k}^{(m)} = \left.\frac{d^m \phi_{n,k}(y)}{dy^m}\right|_{y=0}.$$

Proposition

Uniformly for $n, k \geq 1$ *and* $m \geq 0$

$$|\phi_{n,k}^{(m)}| \leq m! A^m \max\left\{\frac{n}{k^2}, \left(\frac{n}{k^2}\right)^{m/3}\right\}$$

for a suitable constant A.

Characteristic function

Let $\varphi_{n,k}(y) = \mathbf{E}(\exp\{(X_{n,k} - \mu_{n,k})iy/\sigma_{n,k}\})$.

Proposition

Let $1 \le k = o(\sqrt{n})$.

(i) *For* n *large*

$$\varphi_{n,k}(y) = e^{-y^2/2}\left(1 + \mathcal{O}\left(|y|^3 \frac{k}{\sqrt{n}}\right)\right),$$

uniformly for y *with* $|y| \le \epsilon n^{1/6}/k^{1/3}$.

(ii) *For* n *large*

$$|\varphi_{n,k}(y)| \le e^{-\epsilon y^2/2},$$

where $|y| \le \pi\sigma_{n,k}$.

Berry-Esseen bound and LLT for the normal range

Theorem (Rate of convergency)

For $1 \leq k = o(\sqrt{n})$ *as* $n \to \infty$,

$$\sup_{x\in\mathbb{R}} \left| P\left(\frac{X_{n,k} - \mu_{n,k}}{\sigma_{n,k}} < x\right) - \Phi(x)\right| = \mathcal{O}\left(\frac{k}{\sqrt{n}}\right).$$

Berry-Esseen bound and LLT for the normal range

Theorem (Rate of convergency)

For $1 \leq k = o(\sqrt{n})$ *as* $n \to \infty$,

$$\sup_{x \in \mathbb{R}} \left| P\left(\frac{X_{n,k} - \mu_{n,k}}{\sigma_{n,k}} < x \right) - \Phi(x) \right| = \mathcal{O}\left(\frac{k}{\sqrt{n}} \right).$$

Theorem (LLT)

For $1 \leq k = o(\sqrt{n})$ *as* $n \to \infty$,

$$P(X_{n,k} = \lfloor \mu_{n,k} + x\sigma_{n,k} \rfloor) = \frac{e^{-x^2/2}}{\sqrt{2\pi}\sigma_{n,k}} \left(1 + \mathcal{O}\left((1 + |x|^3) \frac{k}{\sqrt{n}} \right) \right),$$

uniformly in $x = o(n^{1/6}/k^{1/3})$.

LLT for the Poisson range

Define

$$\bar{\phi}_{n,k}(y) = e^{-\mu_{n,k}(y-1)}\mathbf{E}\left(y^{X_{n,k}}\right).$$

and

$$\phi_{n,k}^{(m)} = \left.\frac{d^m\bar{\phi}_{n,k}(y)}{dy^m}\right|_{y=1}.$$

LLT for the Poisson range

Define

$$\bar{\phi}_{n,k}(y) = e^{-\mu_{n,k}(y-1)}\mathbf{E}\left(y^{X_{n,k}}\right).$$

and

$$\phi_{n,k}^{(m)} = \left.\frac{d^m\bar{\phi}_{n,k}(y)}{dy^m}\right|_{y=1}.$$

Proposition

Uniformly for $n > k$ and $m \geq 0$

$$|\bar{\phi}_{n,k}^{(m)}| \leq m!A^m\left(\frac{n}{k^3}\right)^{m/2}$$

for a suitable constant A.

Poisson approximation

Theorem (LLT)

For $k < n$ and $n \to \infty$,

$$P(X_{n,k} = l) = e^{-\mu_{n,k}} \frac{(\mu_{n,k})^l}{l!} + \mathcal{O}\left(\frac{n}{k^3}\right)$$

uniformly in l.

Poisson approximation

Theorem (LLT)

For $k < n$ and $n \to \infty$,

$$P(X_{n,k} = l) = e^{-\mu_{n,k}} \frac{(\mu_{n,k})^l}{l!} + \mathcal{O}\left(\frac{n}{k^3}\right)$$

uniformly in l.

Theorem (Poisson approximation)

Let $k < n$ and $k \to \infty$ as $n \to \infty$. Then,

$$d_{TV}(X_{n,k}, \mathrm{Poisson}(\mu_{n,k})) \longrightarrow 0.$$

Remark: A rate can be given as well.

Other types of random trees

- Random recursive trees

  Non-plane, labelled trees with every label sequence from the root to a leave increasing; random model is the uniform model.

  Methods works as well (with minor modifications) and similar results can be proved.

Other types of random trees

- Random recursive trees

  Non-plane, labelled trees with every label sequence from the root to a leave increasing; random model is the uniform model.

  Methods works as well (with minor modifications) and similar results can be proved.

- Plane-oriented recursive trees (PORTs)

  Plane, labelled trees with every label sequence from the root to a leave increasing; random model is the uniform model.

  Method works as well, but details more involved.

of subtrees for PORTs

$X_{n,k}$ satisfies

$$X_{n,k} \stackrel{d}{=} \sum_{i=1}^{N} X_{I_i,k}^{(i)},$$

where $X_{k,k} = 1$ and $X_{I_i,k}^{(i)}$ are conditionally independent given $(N, I_1, I_2, \ldots)$.

of subtrees for PORTs

$X_{n,k}$ satisfies

$$X_{n,k} \stackrel{d}{=} \sum_{i=1}^{N} X_{I_i,k}^{(i)},$$

where $X_{k,k} = 1$ and $X_{I_i,k}^{(i)}$ are conditionally independent given $(N, I_1, I_2, \ldots)$.

This can be simplified to

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

where $X_{k,k} = 1$, $X_{I_n,k}$ and $X_{n-I_n,k}^{*}$ are conditionally independent given I_n and

$$P(I_n = j) = \frac{2(n-j)C_j C_{n-j}}{nC_n}.$$

Underlying recurrence and solution

All (centered or non-centered) moments satisfy

$$a_{n,k} = 2\sum_{j=1}^{n-1} \frac{C_j C_{n-j}}{C_n} a_{j,k} + b_{n,k},$$

where $a_{k,k}$ is given and $a_{n,k} = 0$ for $n < k$.

Underlying recurrence and solution

All (centered or non-centered) moments satisfy

$$a_{n,k} = 2\sum_{j=1}^{n-1} \frac{C_j C_{n-j}}{C_n} a_{j,k} + b_{n,k},$$

where $a_{k,k}$ is given and $a_{n,k} = 0$ for $n < k$.

We have

$$a_{n,k} = \frac{C_k(n+1-k)C_{n+1-k}}{C_n} a_{k,k} + \sum_{k<j\leq n} \frac{C_j(n+1-j)C_{n+1-j}}{C_n} b_{n,k},$$

where $n > k$.

Mean value and variance of PORTs

We have,

$$\mu_{n,k} := \mathbf{E}(X_{n,k}) = \frac{2n-1}{4k^2-1}, \qquad (n > k).$$

Mean value and variance of PORTs

We have,

$$\mu_{n,k} := \mathbf{E}(X_{n,k}) = \frac{2n-1}{4k^2-1}, \qquad (n > k).$$

Moreover, for fixed k as $n \to \infty$,

$$\mathrm{Var}(X_{n,k}) \sim c_k n,$$

where

$$c_k = \frac{8k^2-4k-8}{(4k^2-1)^2} - \frac{((2k-3)!!)^2}{((k-1)!)^2 4^{k-1} k(2k+1)},$$

and, for $k < n$ and $k \to \infty$ as $n \to \infty$,

$$\mathbf{E}(X_{n,k}) \sim \mathrm{Var}(X_{n,k}) \sim \frac{n}{2k^2}.$$

The phase change

Theorem

(i) *(Normal range) Let* $k = o(\sqrt{n})$ *and* $k \to \infty$ *as* $n \to \infty$*. Then,*

$$\frac{X_{n,k} - \mu_{n,k}}{\sqrt{n/(2k^2)}} \xrightarrow{d} \mathcal{N}(0,1).$$

(ii) *(Poisson range) Let* $k \sim c\sqrt{n}$ *as* $n \to \infty$*. Then,*

$$X_{n,k} \xrightarrow{d} \mathrm{Poisson}((2c^2)^{-1}).$$

(iii) *(Degenerate range) Let* $k < n$ *and* $\sqrt{n} = o(k)$ *as* $n \to \infty$*. Then,*

$$X_{n,k} \xrightarrow{L_1} 0.$$

More results and future research

- Parameters of genealogical trees under different random models

More results and future research

- Parameters of genealogical trees under different random models

- Universality of the phase change for the number of subtrees

  Very simple classes of increasing trees and more general classes of increasing trees (polynomial varieties, mobile trees, etc.)

Polynomial varieties

Bergeron, Flajolet, Salvy (1992): classes of increasing trees with degree function $\phi(\omega)$ under the uniform model.

Polynomial varieties

Bergeron, Flajolet, Salvy (1992): classes of increasing trees with degree function $\phi(\omega)$ under the uniform model.

Polynomial varieties: class of increasing tree with

$$\phi(\omega) = \phi_d \omega^d + \cdots + \phi_0,$$

where $d \geq 2$ and $\phi_d, \phi_0 \neq 0$.

Polynomial varieties

Bergeron, Flajolet, Salvy (1992): classes of increasing trees with degree function $\phi(\omega)$ under the uniform model.

Polynomial varieties: class of increasing tree with

$$\phi(\omega) = \phi_d \omega^d + \cdots + \phi_0,$$

where $d \geq 2$ and $\phi_d, \phi_0 \neq 0$.

For mean value and variance of the number of subtrees,

$$\mathbf{E}(X_{n,k}) \sim \mathrm{Var}(X_{n,k}) \sim \frac{d}{d-1} \cdot \frac{n}{k^2},$$

where $k \to \infty$ as $n \to \infty$.

Polynomial varieties - phase change

Theorem

(i) *(Normal range) Let* $k = o\left(\sqrt{n}\right)$ *and* $k \to \infty$ *as* $n \to \infty$. *Then,*

$$\frac{X_{n,k} - \mu_{n,k}}{\sqrt{nd/((d-1)k^2)}} \xrightarrow{d} \mathcal{N}(0,1).$$

(ii) *(Poisson range) Let* $k \sim c\sqrt{n}$ *as* $n \to \infty$. *Then,*

$$X_{n,k} \xrightarrow{d} \mathrm{Poisson}(d/((d-1)c^2)).$$

(iii) *(Degenerate range) Let* $k < n$ *and* $\sqrt{n} = o(k)$ *as* $n \to \infty$. *Then,*

$$X_{n,k} \xrightarrow{L_1} 0.$$

Mobile trees - phase change

Theorem

(i) *(Normal range) Let* $k = o\left(\sqrt{n/\ln n}\right)$ *and* $k \to \infty$ *as* $n \to \infty$. *Then,*

$$\frac{X_{n,k} - \mu_{n,k}}{\sqrt{n/(k^2 \ln k)}} \xrightarrow{d} \mathcal{N}(0,1).$$

(ii) *(Poisson range) Let* $k \sim c\sqrt{n/\ln n}$ *as* $n \to \infty$. *Then,*

$$X_{n,k} \xrightarrow{d} \mathrm{Poisson}(2c^{-2}).$$

(iii) *(Degenerate range) Let* $k < n$ *and* $\sqrt{n/\ln n} = o(k)$ *as* $n \to \infty$. *Then,*

$$X_{n,k} \xrightarrow{L_1} 0.$$

More results and future research

- Parameters of genealogical trees under different random models

- Universality of the phase change for the number of subtrees

  Very simple classes of increasing trees and more general classes of increasing trees (polynomial varieties, mobile trees, etc.)

More results and future research

- Parameters of genealogical trees under different random models
- Universality of the phase change for the number of subtrees

  Very simple classes of increasing trees and more general classes of increasing trees (polynomial varieties, mobile trees, etc.)
- Phase change results for the number of nodes with out-degree k

  Important in computer science.

  A phase change from normal to degenerate is expected (no Poisson range).