The Depth of the Most Recent Common Ancestor of a Random Sample of Species

(joint with Mike Steel)

Michael Fuchs

Department of Mathematical Sciences
Chengchi University
Taipei, Taiwan

NATIONAL CHENGCHI UNIVERSITY

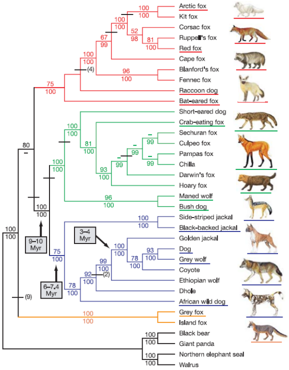
August 4th, 2025

What is a Phylogenetic Tree (PT)?

X ... a finite set.

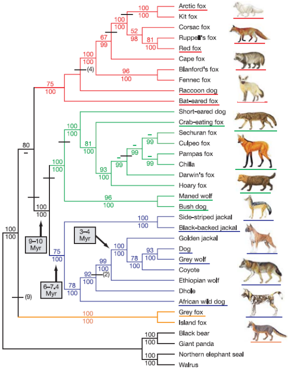
What is a Phylogenetic Tree (PT)?

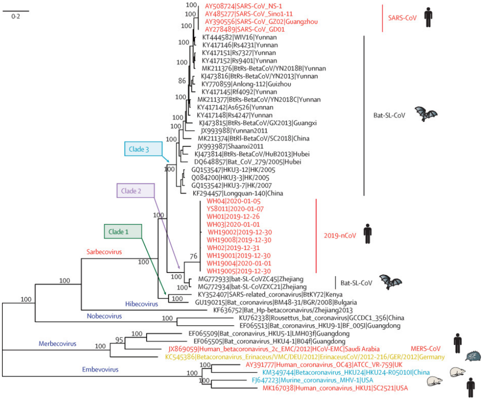
X ... a finite set.



What is a Phylogenetic Tree (PT)?

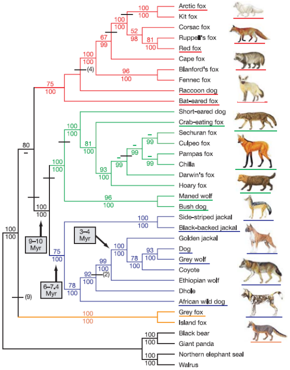
X ... a finite set.

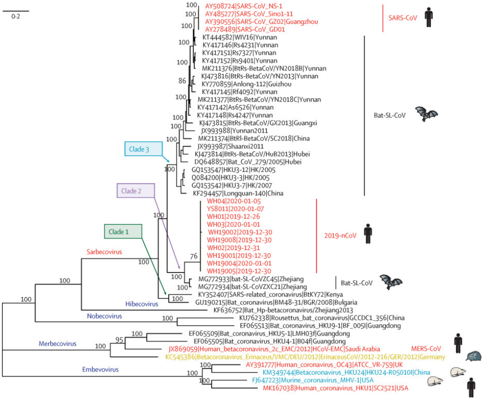




What is a Phylogenetic Tree (PT)?

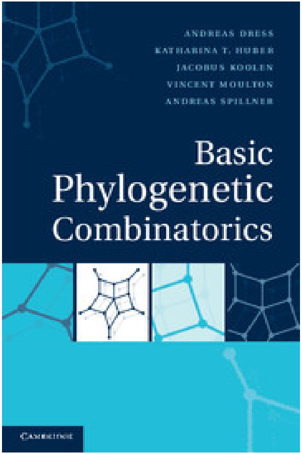
X ... a finite set.

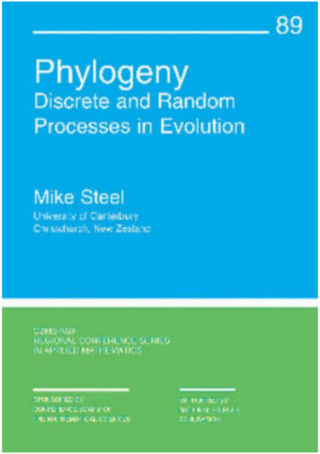




Phylogenetic tree: rooted, binary, non-plane tree with leaves labeled by X.

Phylogenetics





Yule-Harding Model

Yule-Harding Model

- Ranking of phylogenetic tree t:

  Increasing labeling of internal nodes.

Yule-Harding Model

- Ranking of phylogenetic tree t:

  Increasing labeling of internal nodes.

- Yule-Harding Model:

  Phylogenetic tree t of size n is sampled with probability proportional to the number of rankings of t.

Yule-Harding Model

- Ranking of phylogenetic tree t:

  Increasing labeling of internal nodes.

- Yule-Harding Model:

  Phylogenetic tree t of size n is sampled with probability proportional to the number of rankings of t.

  Thus, higher probability is assigned to more “balanced" trees.

Yule-Harding Model

- Ranking of phylogenetic tree t:

  Increasing labeling of internal nodes.

- Yule-Harding Model:

  Phylogenetic tree t of size n is sampled with probability proportional to the number of rankings of t.

  Thus, higher probability is assigned to more "balanced" trees.

- Probability of t under Yule-Harding model:

$$\mathbb{P}(t) = \frac{2^{n-1}}{n!\prod_{r=3}^{n}(r-1)^{d_r(t)}},$$

where $d_r(t)$ is the number of nodes of r with r descendant leaves.

Sanderson's work

Definition

The most recent common ancestor (MRCA) of k leaves of a PT is the root of the leaf-induced subtree of this set of leaves.

Sanderson's work

Definition

The most recent common ancestor (MRCA) of k leaves of a PT is the root of the leaf-induced subtree of this set of leaves.

Question: How many leaves have to be sampled in a large tree such that the MRCA is the root?

Sanderson's work

Definition

The most recent common ancestor (MRCA) of k leaves of a PT is the root of the leaf-induced subtree of this set of leaves.

Question: How many leaves have to be sampled in a large tree such that the MRCA is the root?

Theorem (Sanderson; 1996)

Let k leaves be randomly sampled from a random PT of size n under the Yule-Harding model. Then, as $n \to \infty$,

$$\mathbb{P}(MRCA = root) \sim 1 - \frac{2}{k+1}.$$

Sanderson's work

Definition

The *most recent common ancestor (MRCA)* *of* k *leaves of a PT is the root of the leaf-induced subtree of this set of leaves.*

Question: How many leaves have to be sampled in a large tree such that the MRCA is the root?

Theorem (Sanderson; 1996)

Let k *leaves be randomly sampled from a random PT of size* n *under the Yule-Harding model. Then, as* $n \to \infty$*,*

$$\mathbb{P}(MRCA = root) \sim 1 - \frac{2}{k+1}.$$

E.g. with $k = 40$, the probability equals $\approx 0.9512\cdots$.

Aldous β-splitting Model (i)

Aldous β-splitting Model (i)

- Let f be probability density on $[0,1]$ which is symmetric (i.e. $f(x) = f(1-x)$)

Aldous β-splitting Model (i)

- Let f be probability density on $[0,1]$ which is symmetric (i.e. $f(x) = f(1-x)$)
- Throw n balls uniformly at random into $[0,1]$.

Aldous β-splitting Model (i)

- Let f be probability density on $[0,1]$ which is symmetric (i.e. $f(x) = f(1-x)$)
- Throw n balls uniformly at random into $[0,1]$.
- Split $[0,1]$ into two subintervals according to f; if one subinterval contains no ball repeat.

Aldous β-splitting Model (i)

- Let f be probability density on $[0,1]$ which is symmetric (i.e. $f(x) = f(1-x)$)
- Throw n balls uniformly at random into $[0,1]$.
- Split $[0,1]$ into two subintervals according to f; if one subinterval contains no ball repeat.
- Recursively continue with the subintervals, where a subinterval $[a,b]$ is split at $a + X(b-a)$ with X having distribution f.

Aldous β-splitting Model (i)

- Let f be probability density on $[0,1]$ which is symmetric (i.e. $f(x) = f(1-x)$)
- Throw n balls uniformly at random into $[0,1]$.
- Split $[0,1]$ into two subintervals according to f; if one subinterval contains no ball repeat.
- Recursively continue with the subintervals, where a subinterval $[a,b]$ is split at $a + X(b-a)$ with X having distribution f.
- Stop when a subinterval contains only one ball.

Aldous β-splitting Model (i)

- Let f be probability density on $[0,1]$ which is symmetric (i.e. $f(x) = f(1-x)$)
- Throw n balls uniformly at random into $[0,1]$.
- Split $[0,1]$ into two subintervals according to f; if one subinterval contains no ball repeat.
- Recursively continue with the subintervals, where a subinterval $[a,b]$ is split at $a + X(b-a)$ with X having distribution f.
- Stop when a subinterval contains only one ball.

$\longrightarrow$ This gives a probability distribution on PTs of size n.

Aldous β-splitting Model (ii)

T ... random PT.

Aldous β-splitting Model (ii)

T ... random PT.

Choose a β-distribution $(\beta > -1)$:

$$f(x) = \frac{\Gamma(2\beta+2)}{\Gamma^2(\beta+1)} x^{\beta}(1-x)^{\beta}, \qquad x \in [0,1].$$

Aldous β-splitting Model (ii)

T ... random PT.

Choose a β-distribution $(\beta > -1)$:

$$f(x) = \frac{\Gamma(2\beta+2)}{\Gamma^2(\beta+1)} x^{\beta}(1-x)^{\beta}, \qquad x \in [0,1].$$

Let $\pi_{n,i}$ be the probability that left subtree has size i.

Aldous β-splitting Model (ii)

T … random PT.

Choose a β-distribution $(\beta > -1)$:

$$f(x) = \frac{\Gamma(2\beta+2)}{\Gamma^2(\beta+1)} x^{\beta}(1-x)^{\beta}, \qquad x \in [0,1].$$

Let $\pi_{n,i}$ be the probability that left subtree has size i.

Then,

$$\pi_{n,i} = \frac{1}{\pi_n(\beta)} \frac{\Gamma(\beta+i+1)\Gamma(\beta+n-i+1)}{i!(n-i)!}, \qquad (1 \le i \le n-1),$$

where $\pi_n(\beta)$ is a suitable constant.

Aldous β-splitting Model (ii)

T ... random PT.

Choose a β-distribution ($\beta > -1$):

$$f(x) = \frac{\Gamma(2\beta+2)}{\Gamma^2(\beta+1)} x^{\beta}(1-x)^{\beta}, \qquad x \in [0,1].$$

Let $\pi_{n,i}$ be the probability that left subtree has size i.

Then,

$$\pi_{n,i} = \frac{1}{\pi_n(\beta)} \frac{\Gamma(\beta+i+1)\Gamma(\beta+n-i+1)}{i!(n-i)!}, \qquad (1 \le i \le n-1),$$

where $\pi_n(\beta)$ is a suitable constant.

Note that the above expression makes also sense for $-2 < \beta \le -1$.

Special Cases

- $\beta = 0$: Yule-Harding model:

$$\pi_{n,i} = \frac{1}{n-1}, \qquad 1 \le i \le n-1.$$

Special Cases

- $\beta = 0$: Yule-Harding model:

$$\pi_{n,i} = \frac{1}{n-1}, \qquad 1 \le i \le n-1.$$

- $\beta = -3/2$: Uniform or PDA model:

$$\pi_{n,i} = \frac{C_{i-1}C_{n-i-1}}{C_{n-1}}, \qquad (1 \le i \le n-1),$$

where $C_n = \binom{2n}{n}/(n+1)$ are the Catalan numbers.

Special Cases

- $\beta = 0$: Yule-Harding model:

$$\pi_{n,i} = \frac{1}{n-1}, \qquad 1 \le i \le n-1.$$

- $\beta = -3/2$: Uniform or PDA model:

$$\pi_{n,i} = \frac{C_{i-1}C_{n-i-1}}{C_{n-1}}, \qquad (1 \le i \le n-1),$$

where $C_n = \binom{2n}{n}/(n+1)$ are the Catalan numbers.

- $\beta = -1$: with H_n the harmonic numbers:

$$\pi_{n,i} = \frac{n}{2H_{n-1}} \cdot \frac{1}{i(n-i)}, \qquad (1 \le i \le n-1).$$

This model seems to have the best match with "real" trees.

Extensions of Sanderson's Result (i)

$D_{n,k}$ ... depth of MRCA of a random sample of k leaves of a random PT.

Extensions of Sanderson's Result (i)

$D_{n,k}$ ... depth of MRCA of a random sample of k leaves of a random PT.

Theorem (F. & Steel; 2025+)

(i) *For* $\beta = -1$*:*

$$\mathbb{P}(D_{n,k} = 0) = \frac{H_{k-1}}{H_{n-1}},$$

where H_m *denotes the* m*-th harmonic number.*

(ii) *For* $\beta \neq -1$*:*

$$\mathbb{P}(D_{n,k} = 0) = 1 - \frac{2(\beta+1)\cdots(\beta+k)}{k!\binom{n}{k}} \times \frac{\binom{n+2\beta+1}{n-k} - \binom{n+\beta}{n-k}}{\binom{n+2\beta+1}{n} - 2\binom{n+\beta}{n}}.$$

Extensions of Sanderson's Result (i)

$D_{n,k}$ ... depth of MRCA of a random sample of k leaves of a random PT.

Theorem (F. & Steel; 2025+)

(i) *For* $\beta = -1$*:*

$$\mathbb{P}(D_{n,k} = 0) = \frac{H_{k-1}}{H_{n-1}},$$

where H_m *denotes the* m*-th harmonic number.*

(ii) *For* $\beta \neq -1$*:*

$$\mathbb{P}(D_{n,k} = 0) = 1 - \frac{2(\beta+1)\cdots(\beta+k)}{k!\binom{n}{k}} \times \frac{\binom{n+2\beta+1}{n-k} - \binom{n+\beta}{n-k}}{\binom{n+2\beta+1}{n} - 2\binom{n+\beta}{n}}.$$

Thus, $\lim_{n\to\infty} \mathbb{P}(D_{n,k} = 0) = 0$ iff $-2 < \beta \leq -1$.

Extensions of Sanderson's Result (ii)

Corollary (F. & Steel; 2025+)

(i) *For* $\beta > -1$*:* $\lim_{n\to\infty} \mathbb{P}(D_{n,k} = 0) = 1 - \frac{(\beta+2)\cdots(\beta+k)}{(2\beta+3)\cdots(2\beta+k+1)}$.

(ii) *For* $\beta = -1$*:* $\lim_{n\to\infty} \mathbb{P}(D_{n,k} = 0) = \alpha$ *if* $k \sim n^{\alpha}$.

(iii) *For* $-2 < \beta < -1$*:*

$$\lim_{n\to\infty} \mathbb{P}(D_{n,k} = 0) = \begin{cases} c^{-\beta-1}, & \text{if } k \sim cn; \\ 0, & \text{if } k = o(n). \end{cases}$$

Extensions of Sanderson's Result (ii)

Corollary (F. & Steel; 2025+)

(i) *For* $\beta > -1$*:* $\lim_{n\to\infty} \mathbb{P}(D_{n,k} = 0) = 1 - \frac{(\beta+2)\cdots(\beta+k)}{(2\beta+3)\cdots(2\beta+k+1)}$.

(ii) *For* $\beta = -1$*:* $\lim_{n\to\infty} \mathbb{P}(D_{n,k} = 0) = \alpha$ *if* $k \sim n^{\alpha}$.

(iii) *For* $-2 < \beta < -1$*:*

$$\lim_{n\to\infty} \mathbb{P}(D_{n,k} = 0) = \begin{cases} c^{-\beta-1}, & \text{if } k \sim cn; \\ 0, & \text{if } k = o(n). \end{cases}$$

n	10	10^2	10^3	10^4	10^5	10^6
$\beta = 0$	8	29	38	39	39	39
$\beta = -1$	9	78	688	6131	54635	486930
$\beta = -3/2$	10	91	903	9026	90251	902501

Figure: Values of k such that $\mathbb{P}(D_{n,k} = 0) \geq 0.95$.

Limit Laws for $D_{n,k}$ (i)

Question: How about limit laws for $D_{n,k}$?

Limit Laws for $D_{n,k}$ (i)

Question: How about limit laws for $D_{n,k}$?

Theorem (F. & Steel; 2025+)

As $n \to \infty$,

$$D_{n,k} \xrightarrow{d} G_k,$$

where G_k *is geometric with success probability* $1 - q(\beta, k)$, *where*

$$q(\beta, k) = \frac{(\beta+2)\cdots(\beta+k)}{(2\beta+3)\cdots(2\beta+k+1)}.$$

Limit Laws for $D_{n,k}$ (i)

Question: How about limit laws for $D_{n,k}$?

Theorem (F. & Steel; 2025+)

As $n \to \infty$,

$$D_{n,k} \xrightarrow{d} G_k,$$

where G_k is geometric with success probability $1 - q(\beta, k)$, where

$$q(\beta, k) = \frac{(\beta+2)\cdots(\beta+k)}{(2\beta+3)\cdots(2\beta+k+1)}.$$

Proof. By induction on r,

$$\mathbb{P}(D_{n,k} \geq r) = q(\beta, k)^r,$$

where one uses that with high probability no subtrees of the root is small.

Limit Laws for $D_{n,k}$ (ii)

Theorem (F. & Steel; 2025+)

(i) *For* $\beta = -1$,

$$\frac{H_{k-1} D_{n,k}}{\log n} \xrightarrow{d} \mathrm{Exp}(1),$$

where $\mathrm{Exp}(1)$ *is the standard exponential distribution.*

(ii) *For* $\beta = -3/2$,

$$\frac{D_{n,k}}{\sqrt{n}} \xrightarrow{d} D_k,$$

where D_k *has the three-parameter Mittag-Leffler distribution* $\mathrm{ML}(1/2, 1/2, k-1)$.

Limit Laws for $D_{n,k}$ (ii)

Theorem (F. & Steel; 2025+)

(i) *For* $\beta = -1$,

$$\frac{H_{k-1} D_{n,k}}{\log n} \xrightarrow{d} \mathrm{Exp}(1),$$

where $\mathrm{Exp}(1)$ *is the standard exponential distribution.*

(ii) *For* $\beta = -3/2$,

$$\frac{D_{n,k}}{\sqrt{n}} \xrightarrow{d} D_k,$$

where D_k *has the three-parameter Mittag-Leffler distribution* $\mathrm{ML}(1/2, 1/2, k-1)$.

Both results are proved with the method of moments.

Sketch of Proof (i)

We have,

$$(D_{n,k}|I_n = j) \stackrel{d}{=} \begin{cases} D_{j,k}+1, & \text{with probability } \binom{j}{k}/\binom{n}{k}; \\ D_{n-j,k}+1, & \text{with probability } \binom{n-j}{k}/\binom{n}{k}; \\ 0, & \text{otherwise,} \end{cases}$$

where I_n is the size of the left subtree of the root.

Sketch of Proof (i)

We have,

$$(D_{n,k}|I_n = j) \stackrel{d}{=} \begin{cases} D_{j,k}+1, & \text{with probability } \binom{j}{k}/\binom{n}{k}; \\ D_{n-j,k}+1, & \text{with probability } \binom{n-j}{k}/\binom{n}{k}; \\ 0, & \text{otherwise,} \end{cases}$$

where I_n is the size of the left subtree of the root.

For $\beta = -3/2$, set

$$D^{[m]}(z) := \sum_{n\geq 1} C_{n-1}\binom{n}{k}\mathbb{E}(D_{n,k}^m)z^n.$$

Sketch of Proof (i)

We have,

$$(D_{n,k}|I_n = j) \stackrel{d}{=} \begin{cases} D_{j,k}+1, & \text{with probability } \binom{j}{k}/\binom{n}{k}; \\ D_{n-j,k}+1, & \text{with probability } \binom{n-j}{k}/\binom{n}{k}; \\ 0, & \text{otherwise,} \end{cases}$$

where I_n is the size of the left subtree of the root.

For $\beta = -3/2$, set

$$D^{[m]}(z) := \sum_{n\geq 1} C_{n-1}\binom{n}{k}\mathbb{E}(D_{n,k}^m)z^n.$$

Then,

$$D^{[m]}(z) = \sum_{\ell=1}^{m}\binom{m}{\ell}D^{[m-\ell]}(z)((1-4z)^{-1/2}-1).$$

Sketch of Proof (ii)

Proposition

As $z \to 1/4$,

$$D^{[m]}(z) \sim \frac{m! C_{k-1} 4^{-k}}{(1-4z)^{k+(m-1)/2}}.$$

Sketch of Proof (ii)

Proposition

As $z \to 1/4$,

$$D^{[m]}(z) \sim \frac{m! C_{k-1} 4^{-k}}{(1-4z)^{k+(m-1)/2}}.$$

Corollary

As $n \to \infty$,

$$\mathbb{E}(D_{n,k}^m) \sim \frac{m! C_{k-1} 4^{1-k} k! \sqrt{\pi}}{\Gamma(k+(m-1)/2)} n^{m/2}.$$

Sketch of Proof (ii)

Proposition

As $z \to 1/4$,

$$D^{[m]}(z) \sim \frac{m!C_{k-1}4^{-k}}{(1-4z)^{k+(m-1)/2}}.$$

Corollary

As $n \to \infty$,

$$\mathbb{E}(D_{n,k}^m) \sim \frac{m!C_{k-1}4^{1-k}k!\sqrt{\pi}}{\Gamma(k+(m-1)/2)} n^{m/2}.$$

Remark: The sequence

$$\frac{m!C_{k-1}4^{1-k}k!\sqrt{\pi}}{\Gamma(k+(m-1)/2)}$$

is the (unique) moment sequence of $\mathrm{ML}(1/2, 1/2, k-1)$.

Sketch of Proof (iii)

For $\beta = -1$, all moments of $D_{n,k}$ satisfy the recurrence

$$a_n = \frac{1}{H_{n-1}} \sum_{j=1}^{n-1} \frac{a_j}{n-j} + b_n.$$

Sketch of Proof (iii)

For $\beta = -1$, all moments of $D_{n,k}$ satisfy the recurrence

$$a_n = \frac{1}{H_{n-1}} \sum_{j=1}^{n-1} \frac{a_j}{n-j} + b_n.$$

This recurrence was recently studied in:

D. Aldous and B. Pittel (2025). The critical beta-splitting random tree I: Heights and related results, Ann. Appl. Probab. 35(1): 158–195.

Sketch of Proof (iii)

For $\beta = -1$, all moments of $D_{n,k}$ satisfy the recurrence

$$a_n = \frac{1}{H_{n-1}} \sum_{j=1}^{n-1} \frac{a_j}{n-j} + b_n.$$

This recurrence was recently studied in:

D. Aldous and B. Pittel (2025). The critical beta-splitting random tree I: Heights and related results, Ann. Appl. Probab. 35(1): 158–195.

Proposition

Let $t \in \mathbb{N}$ and $s \in \mathbb{Z}$.

(i) *If $b_n = O(n^t \log^s n)$, then $a_n = O(n^t \log^{s+1} n)$.*

(ii) *If $b_n = cn^t \log^s n$, then $a_n \sim cn^t \log^{s+1} n / H_t$.*

Sketch of the Proof (iv)

Let

$$A_{n,k}^{[m]} := (n-1)_{(k-1)}\mathbb{E}(D_{n,k}^{m}).$$

Sketch of the Proof (iv)

Let

$$A_{n,k}^{[m]} := (n-1)_{(k-1)}\mathbb{E}(D_{n,k}^{m}).$$

Then,

$$A_{n,k}^{[m]} = \frac{1}{H_{n-1}}\sum_{j=1}^{n-1}\frac{A_{j,k}^{[m]}}{n-j} + \frac{1}{H_{n-1}}\sum_{\ell=0}^{m-1}\binom{m}{\ell}\sum_{j=1}^{n-1}\frac{A_{j,k}^{[\ell]}}{n-j}.$$

Sketch of the Proof (iv)

Let

$$A_{n,k}^{[m]} := (n-1)_{(k-1)}\mathbb{E}(D_{n,k}^m).$$

Then,

$$A_{n,k}^{[m]} = \frac{1}{H_{n-1}}\sum_{j=1}^{n-1}\frac{A_{j,k}^{[m]}}{n-j} + \frac{1}{H_{n-1}}\sum_{\ell=0}^{m-1}\binom{m}{\ell}\sum_{j=1}^{n-1}\frac{A_{j,k}^{[\ell]}}{n-j}.$$

Thus,

$$b_n := \frac{1}{H_{n-1}}\sum_{\ell=0}^{m-1}\binom{m}{\ell}\sum_{j=1}^{n-1}\frac{A_{j,k}^{[\ell]}}{n-j}.$$

Sketch of the Proof (iv)

Let

$$A_{n,k}^{[m]} := (n-1)_{(k-1)}\mathbb{E}(D_{n,k}^{m}).$$

Then,

$$A_{n,k}^{[m]} = \frac{1}{H_{n-1}}\sum_{j=1}^{n-1}\frac{A_{j,k}^{[m]}}{n-j} + \frac{1}{H_{n-1}}\sum_{\ell=0}^{m-1}\binom{m}{\ell}\sum_{j=1}^{n-1}\frac{A_{j,k}^{[\ell]}}{n-j}.$$

Thus,

$$b_n := \frac{1}{H_{n-1}}\sum_{\ell=0}^{m-1}\binom{m}{\ell}\sum_{j=1}^{n-1}\frac{A_{j,k}^{[\ell]}}{n-j}.$$

E.g., for $m=1$:

$$b_n = \frac{1}{H_{n-1}}\sum_{j=1}^{n-1}\frac{(j-1)_{(k-1)}}{n-j} = (n-1)_{(k-1)}\left(1-\frac{H_{k-1}}{H_{n-1}}\right)$$

and thus,

$$\mathbb{E}(D_{n,k}) \sim \log n / H_{k-1}.$$

Sketch of Proof (v)

Using "moment pumping", one obtains the following.

Proposition

Set

$$A_{n,k}^{[m]} := (n-1)_{(k-1)}\mathbb{E}(D_{n,k}^{m}).$$

Then, as $n \to \infty$,

$$A_{n,k}^{[m]} \sim \frac{m!}{H_{k-1}^{m}} n^{k-1} \log^{m} n.$$

Sketch of Proof (v)

Using "moment pumping", one obtains the following.

Proposition

Set

$$A_{n,k}^{[m]} := (n-1)_{(k-1)}\mathbb{E}(D_{n,k}^{m}).$$

Then, as $n \to \infty$,

$$A_{n,k}^{[m]} \sim \frac{m!}{H_{k-1}^{m}} n^{k-1} \log^{m} n.$$

Thus, as $n \to \infty$,

$$\mathbb{E}(D_{n,k}^{m}) \sim \frac{m!}{H_{k-1}^{m}} \log^{m} n$$

which implies the claimed limit law.

A Conjecture

Conjecture

For $-2 < \beta < -1$,

$$\frac{D_{n,k}}{n^{-\beta-1}} \xrightarrow{d} D_k,$$

where D_k *is uniquely characterized by the moment sequence*

$$c_m = m! \prod_{j=1}^{m} e(\beta, k, j)$$

and

$$e(\beta, k, m) := \frac{\Gamma((-\beta-1)m + \beta + k + 1)}{\Gamma((-\beta-1)m + 2\beta + k + 2)} - \frac{\Gamma(\beta+2)}{\Gamma(2\beta+3)}.$$

A Conjecture

Conjecture

For $-2 < \beta < -1$,

$$\frac{D_{n,k}}{n^{-\beta-1}} \xrightarrow{d} D_k,$$

where D_k *is uniquely characterized by the moment sequence*

$$c_m = m! \prod_{j=1}^{m} e(\beta, k, j)$$

and

$$e(\beta, k, m) := \frac{\Gamma((-\beta-1)m+\beta+k+1)}{\Gamma((-\beta-1)m+2\beta+k+2)} - \frac{\Gamma(\beta+2)}{\Gamma(2\beta+3)}.$$

Remark: Relatively little is known so far for the range $-2 < \beta < -1$.

Random Sample of Species (i)

Before, we randomly picked a set of k species.

Random Sample of Species (i)

Before, we randomly picked a set of k species.

Now, assume that each species is included in the set of species with probability p.

Random Sample of Species (i)

Before, we randomly picked a set of k species.

Now, assume that each species is included in the set of species with probability p.

D_n ... depth of MRCA of the random set of species.

Random Sample of Species (i)

Before, we randomly picked a set of k species.

Now, assume that each species is included in the set of species with probability p.

D_n ... depth of MRCA of the random set of species.

Lemma

We have,

$$\mathbb{P}(D_n = 0) = \frac{1 - 2\mathbb{E}(q^{I_n}) + q^n}{1 - q^n},$$

where I_n is the size of the left subtree.

Random Sample of Species (i)

Before, we randomly picked a set of k species.

Now, assume that each species is included in the set of species with probability p.

D_n ... depth of MRCA of the random set of species.

Lemma

We have,

$$\mathbb{P}(D_n = 0) = \frac{1 - 2\mathbb{E}(q^{I_n}) + q^n}{1 - q^n},$$

where I_n is the size of the left subtree.

Let $p = \lambda / n$.

Random Sample of Species (ii)

Theorem (F. & Steel; 2025+)

(i) *For* $\beta > -1$,

$$\mathbb{P}(D_n = 0) \sim \frac{1 - 2\Gamma(2\beta+2)\int_0^1 e^{-\lambda x} x^{\beta}(1-x)^{\beta}\mathrm{d}x/\Gamma(\beta+1)^2 + e^{-\lambda}}{1 - e^{-\lambda}}.$$

(ii) *For* $\beta = -1$,

$$\mathbb{P}(D_n = 0) \sim \frac{E_1(\lambda) + \log\lambda + \gamma - e^{-\lambda}(\mathrm{Ei}(\lambda) - \lambda - \gamma)}{(1 - e^{-\lambda})H_{n-1}}.$$

(iii) *For* $-2 < \beta < -1$,

$$\mathbb{P}(D_n = 0) \sim \frac{cn^{\beta+1}}{\Gamma(\beta+1)(1 - e^{-\lambda})}.$$

Random Sample of Species (iii)

Corollary (F. & Steel; 2025+)

(i) *For* $\beta > -1$,

$$\frac{1 - 2\Gamma(2\beta+2)\int_0^1 e^{-\lambda x} x^\beta (1-x)^\beta \mathrm{d}x / \Gamma(\beta+1)^2 + e^{-\lambda}}{1 - e^{-\lambda}} \sim 1 - \frac{2\Gamma(2\beta+2)}{\lambda^{\beta+1}\Gamma(\beta+1)}.$$

(ii) *For* $\beta = -1$,

$$\frac{E_1(\lambda) + \log\lambda + \gamma - e^{-\lambda}(\mathrm{Ei}(\lambda) - \lambda - \gamma)}{(1 - e^{-\lambda})H_{n-1}} \sim \frac{\log\lambda}{H_{n-1}}.$$

(iii) *For* $-2 < \beta < -1$,

$$\frac{cn^{\beta+1}}{\Gamma(\beta+1)(1-e^{-\lambda})} \sim \left(\frac{\lambda}{n}\right)^{-\beta-1}.$$

Reference

M. Fuchs and M. Steel. Predicting the depth of the most recent common ancestor of a random sample of k species: the impact of phylogenetic tree shape, J. Math. Biol., in press.

Reference

M. Fuchs and M. Steel. Predicting the depth of the most recent common ancestor of a random sample of k species: the impact of phylogenetic tree shape, J. Math. Biol., in press.

Thanks for the attention!