Clades and clans probability in Yule trees

Sha (Joe) Zhu, Dr James Degnan, and Prof Mike Steel
University of Canterbury

November 9, 2011

Content

Clade

Definition

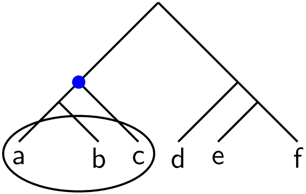
A clade of a rooted tree is a subset of X that corresponds to the set of leaves that are descended from any internal vertex.

Clade

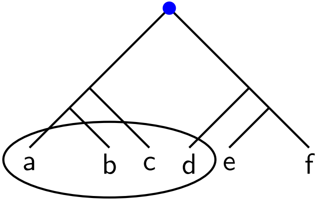
Definition

A clade of a rooted tree is a subset of X that corresponds to the set of leaves that are descended from any internal vertex.

Example:



Set $\{a, b, c\}$ is a clade.



Set $\{a, b, c, d\}$ is not a clade.

Content

Yule Tree

Definition

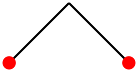
Yule tree construction starts from a X_2 tree, choose a pendant edge randomly, and add a node onto this pendant edge. Repeat this process on a X_k tree, choose a pendent edge at random, and add the $(k+1)$th node onto it to form a X_{k+1} tree.

Yule Tree

Definition

Yule tree construction starts from a X_2 tree, choose a pendant edge randomly, and add a node onto this pendant edge. Repeat this process on a X_k tree, choose a pendent edge at random, and add the $(k+1)$th node onto it to form a X_{k+1} tree.

Example:

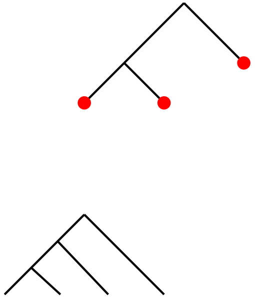


Yule Tree

Definition

Yule tree construction starts from a X_2 tree, choose a pendant edge randomly, and add a node onto this pendant edge. Repeat this process on a X_k tree, choose a pendent edge at random, and add the $(k+1)$th node onto it to form a X_{k+1} tree.

Example:

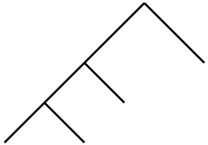


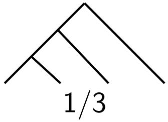
Yule Tree

Definition

Yule tree construction starts from a X_2 tree, choose a pendant edge randomly, and add a node onto this pendant edge. Repeat this process on a X_k tree, choose a pendent edge at random, and add the $(k+1)$th node onto it to form a X_{k+1} tree.

Example:



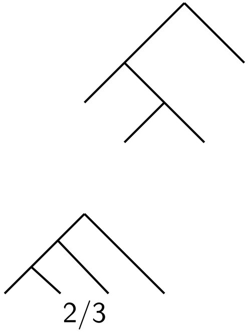


Yule Tree

Definition

Yule tree construction starts from a X_2 tree, choose a pendant edge randomly, and add a node onto this pendant edge. Repeat this process on a X_k tree, choose a pendent edge at random, and add the $(k+1)$th node onto it to form a X_{k+1} tree.

Example:

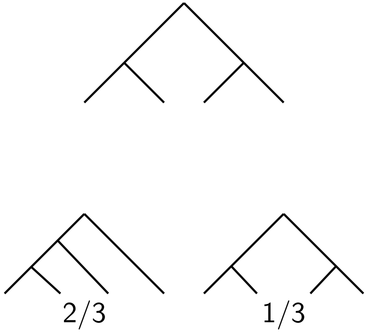


Yule Tree

Definition

Yule tree construction starts from a X_2 tree, choose a pendant edge randomly, and add a node onto this pendant edge. Repeat this process on a X_k tree, choose a pendent edge at random, and add the $(k+1)$th node onto it to form a X_{k+1} tree.

Example:



Coalescent Tree

Definition

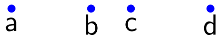
The coalescent process starts from the bottom of a tree.
Randomly choose two lineage to coalesce into one, and repeat this process on the remaing lineages until all lineages coalesce into one.

Coalescent Tree

Definition

The coalescent process starts from the bottom of a tree.
Randomly choose two lineage to coalesce into one, and repeat this process on the remaing lineages until all lineages coalesce into one.

Example:

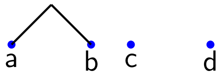


Coalescent Tree

Definition

The coalescent process starts from the bottom of a tree.
Randomly choose two lineage to coalesce into one, and repeat this process on the remaing lineages until all lineages coalesce into one.

Example:

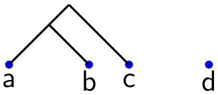


Coalescent Tree

Definition

The coalescent process starts from the bottom of a tree.
Randomly choose two lineage to coalesce into one, and repeat this process on the remaing lineages until all lineages coalesce into one.

Example:

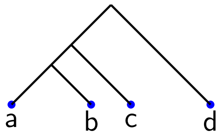


Coalescent Tree

Definition

The coalescent process starts from the bottom of a tree.
Randomly choose two lineage to coalesce into one, and repeat this process on the remaing lineages until all lineages coalesce into one.

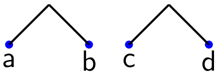
Example:



Definition

The coalescent process starts from the bottom of a tree.
Randomly choose two lineage to coalesce into one, and repeat this process on the remaing lineages until all lineages coalesce into one.

Example:

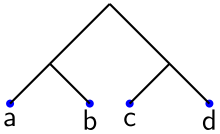


Coalescent Tree

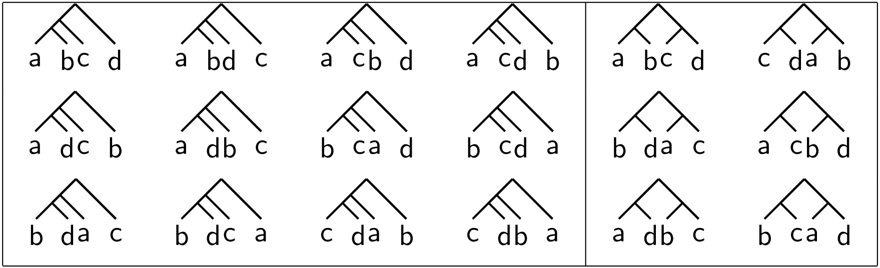
Definition

The coalescent process starts from the bottom of a tree.
Randomly choose two lineage to coalesce into one, and repeat this process on the remaing lineages until all lineages coalesce into one.

Example:



Coalescent Tree



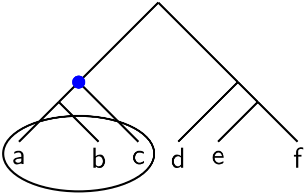
Content

Clade

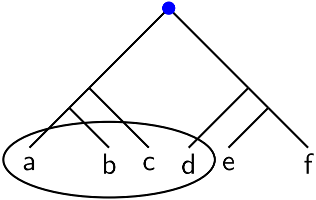
Definition

A clade of a rooted tree is a subset of X that corresponds to the set of leaves that are descended from any internal vertex.

Example:



Set $\{a, b, c\}$ is a clade.



Set $\{a, b, c, d\}$ is not a clade.

Yule process properties (I) (Aldous, 1995)

(EP) If T' is obtained from T by permuting its leaves, then

$$\mathbb{P}(\mathcal{T} = T') = \mathbb{P}(\mathcal{T} = T).$$

Yule process properties (I) (Aldous, 1995)

(EP) If T' is obtained from T by permuting its leaves, then

$$\mathbb{P}(\mathcal{T} = T') = \mathbb{P}(\mathcal{T} = T).$$

Example:
Let $p_n(A)$ be the probability that A is a clade of T, and $a = |A|$, then

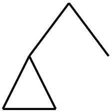
$$p_n(A) = p_n(a).$$

From Rosenberg (2003) we have:

$$p_n(a) = \begin{cases} \dfrac{2n}{a(a+1)} \dbinom{n}{a}^{-1}, & \text{if } 1 \leq a \leq n-1; \\ 0, & \text{otherwise.} \end{cases}$$

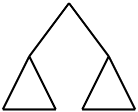
Clades probabilities

- Probability of A is a clade of T (Rosenberg, 2003; Brown, 1994)



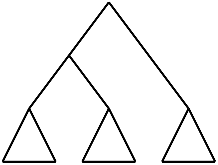
Clades probabilities

- Probability of A is a clade of T (Rosenberg, 2003; Brown, 1994)
- Probability of A and $X - A$ are clades of T (Rosenberg, 2003; Brown, 1994)



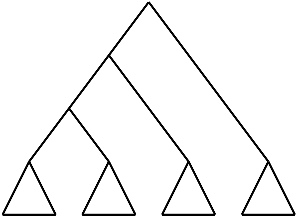
Clades probabilities

- Probability of A is a clade of T (Rosenberg, 2003; Brown, 1994)
- Probability of A and $X - A$ are clades of T (Rosenberg, 2003; Brown, 1994)
- Probability of three clades, and four clades (Brown, 1994)



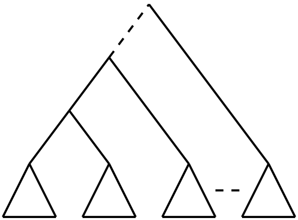
Clades probabilities

- Probability of A is a clade of T (Rosenberg, 2003; Brown, 1994)
- Probability of A and $X - A$ are clades of T (Rosenberg, 2003; Brown, 1994)
- Probability of three clades, and four clades (Brown, 1994)



Clades probabilities

- Probability of A is a clade of T (Rosenberg, 2003; Brown, 1994)
- Probability of A and $X - A$ are clades of T (Rosenberg, 2003; Brown, 1994)
- Probability of three clades, and four clades (Brown, 1994)
- k clades



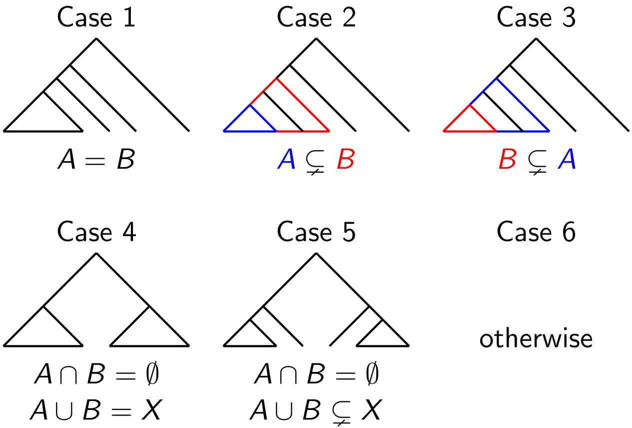
Yule process properties (II) (Aldous, 1995)

(GE) For any proper (and non-empty) subset A of X, and any rooted binary phylogenetic tree T with leaf set $X - A$:

$$\mathbb{P}(\mathcal{T}_{X|(X-A)} = T | A \in c(\mathcal{T})) = \mathbb{P}(\mathcal{T}_{(X-A)} = T).$$

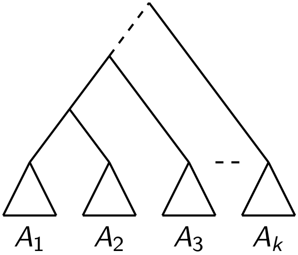
(SC) $\mathbb{P}(\mathcal{T}_{X|A} = T) = \mathbb{P}(\mathcal{T}_A = T).$

Clades probability of A and B, $p_n(a, b)$



More than k clades,

Consider set $A_1, A_2, \ldots, A_k$ are clades, and $\bigcup_{i=1}^{k} A_i = X$

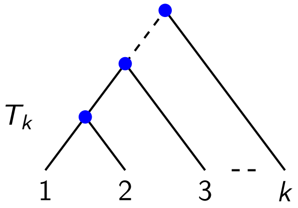


$$p(a_1, \ldots, a_k; T_k) = \frac{2^{k-1} \prod_{i=1}^{k} a_i!}{n!} \prod_{v \in \mathcal{I}(T_k)} \left(\frac{1}{\sum_{i=1}^{k} a_i l_v(A_i) - 1} \right),$$

$$p(a_1, \ldots, a_k) = \sum_{T_k} p(a_1, \ldots, a_k; T_k).$$

More than k clades,

Consider set $A_1, A_2, \ldots, A_k$ are clades, and $\bigcup_{i=1}^{k} A_i = X$

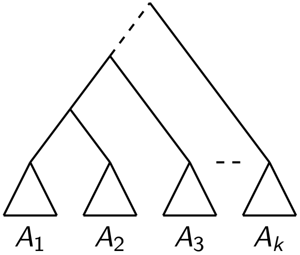


$$p(a_1, \ldots, a_k; T_k) = \frac{2^{k-1} \prod_{i=1}^{k} a_i!}{n!} \prod_{v \in \mathcal{I}(T_k)} \left(\frac{1}{\sum_{i=1}^{k} a_i l_v(A_i) - 1} \right),$$

$$p(a_1, \ldots, a_k) = \sum_{T_k} p(a_1, \ldots, a_k; T_k).$$

More than k clades,

Consider set $A_1, A_2, \ldots, A_k$ are clades, and $\bigcup_{i=1}^{k} A_i = X$



$$p(a_1, \ldots, a_k; T_k) = \frac{2^{k-1}\prod_{i=1}^{k} a_i!}{n!} \prod_{v \in \mathcal{I}(T_k)} \left(\frac{1}{\sum_{i=1}^{k} a_i I_v(A_i) - 1} \right),$$

$$p(a_1, \ldots, a_k) = \sum_{T_k} p(a_1, \ldots, a_k; T_k).$$

Correlation

Corollary

For any two strict subsets A, B *of* X*, the correlation* $\rho_n(A, B)$ *is:*

- *strictly negative, if* A, B *are not compatible, and undefined if* $|A| = 1$ *or* $|B| = 1$*.*
- *strictly positive, otherwise.*

Correlation

Corollary

For any two strict subsets A, B of X, the correlation $\rho_n(A, B)$ is:

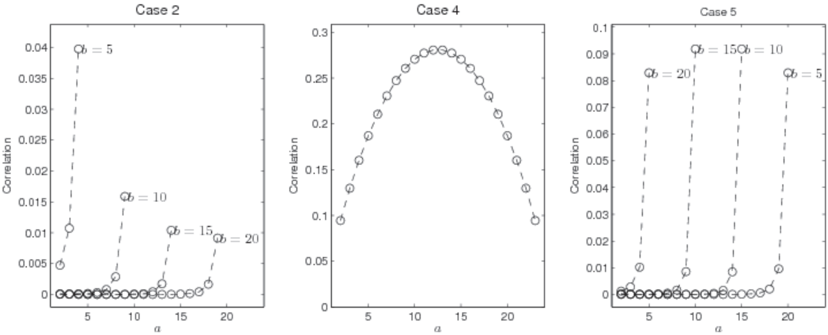
- *strictly negative, if A, B are not compatible, and undefined if $|A| = 1$ or $|B| = 1$.*
- *strictly positive, otherwise.*

$$\rho_n(A, B) = \frac{p_n(A, B) - p_n(A)p_n(B)}{\sqrt{p_n(A)(1 - p_n(A))p_n(B)(1 - p_n(B))}}.$$

Corollary

For any two strict subsets A, B of X, the correlation $\rho_n(A, B)$ is:

- *strictly negative, if A, B are not compatible, and undefined if $|A| = 1$ or $|B| = 1$.*
- *strictly positive, otherwise.*



Content

Clan

Definition

We say that a subset A of X is a clan of an unrooted phylogenetic $X-$tree $T^{-\rho}$ if $A \mid (X - A)$ is a split of $T^{-\rho}$.

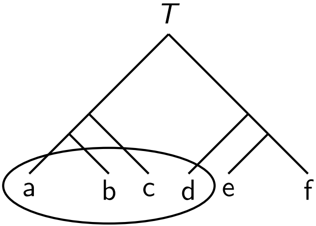
Clan

Definition

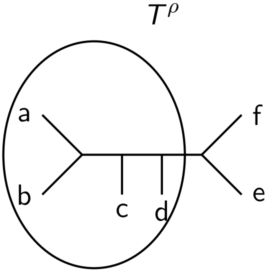
We say that a subset A of X is a clan of an unrooted phylogenetic $X-$tree $T^{-\rho}$ if $A \mid (X - A)$ is a split of $T^{-\rho}$.

Example:

Let $X = \{a, b, c, d, e, f\}$, $A = \{a, b, c, d\}$, and $X - A = \{e, f\}$.



Set A is not a clade.



Set A is a clan.

Clans probability

Lemma

A set A is a clan of a unrooted tree $T^{-\rho}$ if and only if either A is a clade of T or $X - A$ is a clade of T.

$$q_n(A) = p_n(A) + p_n(X - A) - p_n(A, X - A)$$

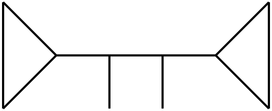
If $A \cap B = \emptyset$ and $A \cup B = X$,

$$q_n(A, B) = q_n(A)$$

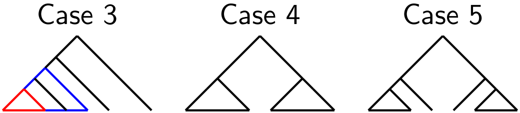
Clans probability

If $A \cap B = \emptyset$ and $A \cup B \subsetneq X$,

$$q_n(A, B)$$



Recall:



Content

In this project,

- we derived exact probabilities for two compatible clades with any sizes;
- we have shown there is positive correlation between any two compatible clades;
- we have extended clades probabilities to clans probabilities.

Reference

Aldous, D. (1995). *Random Discrete Structures*. Springer.

Brown, J. K. M. (1994). Probabilities of evolutionary trees. *Systematic Biology 43*, 8–91.

Rosenberg, N. A. (2003). The shapes of neutral gene genealogies in two species: probabilities of monophyly, paraphyly and polyphyly in a coalescent model. *Evolution 57*(7), 1465–1477.

Zhu, S., J. H. Degnan, and M. Steel (2011). Clades, clans and reciprocal monophyly under neutral evolutionary models. *Theoretical Population Biology 79*, 220–227.

Acknowledgments

- My supervisors: Mike and James.
- Marsden fund.
- Organizers: Barbara and Jeremy.