Topology of Phylogenetic Mixtures

MATH 833 - Fall 2012 *Presenter: Karthik Rao Aroor*

A phylogenetic mixture model is a weighted average of the site pattern frequencies of phylogenetic trees having same or different topologies. We are interested in finding out the topology of the mixture of two phylogenetic trees - specifically, quartet trees. We will discuss two cases: one in which the mixing of two trees gives rise to a tree of same topology and the other in which mixing gives rise to a tree of different topology. For example if we mix two $12|34$ trees, sometimes we can get a $12|34$ tree and sometimes we can get a $13|24$ tree. We will derive the conditions on the branch lengths of the original trees that gives rise to these two cases.

The fidelity, θ_e of an edge e with length $\gamma(e)$ is defined as,

$$\theta_e = \exp[\,-2\gamma(e)\,] \tag{1}$$

The fidelity of an edge varies between 0 and 1 for positive edge lengths.

Hadamard matrix H is the matrix of -1s and 1s and each column is orthogonal to every other column. It is used to compute the Fourier transform of the split probabilities to get $q_A = (H\,p)_A$ where $A \subset \{1, 2 \ldots n\}$.
The Fourier transform of the split probabilities can be written as the product of fidelities (Theorem 8.6.3, Semple and Steel),

$$q_A = \prod_{e \in P(T,A)} \theta_e \tag{2}$$

where $P(T, A)$ is the set of edges that lie in the set of edge disjoint paths connecting the taxa in A to each other.

Proposition 1: The fidelity of a pendant edge a on a quartet $ab|cd$ is given by,

$$\theta_a = \sqrt{\frac{q_{ab}\, q_{ac}}{q_{bc}}}$$

For a given phylogenetic tree topology, the Fourier transform of the split probabilities must satisfy a set of equations called *phylogenetic invariants* and for a quartet tree $ab|cd$, the invariants are given by,

$$q_{abcd} - q_{ab}\, q_{cd} = 0 \tag{3}$$

$$q_{ac}\, q_{bd} - q_{ad}\, q_{bc} = 0 \tag{4}$$

q is the Fourier transform of the site pattern frequencies when the edge fidelities lie between 0 and 1 and they satisfy phylogenetic invariants.

We will now derive the conditions such that the mixture of two $12|34$ trees results in a tree of the same $12|34$ topology. We will use parameters $\{q, t, \theta\}$to represent parameters of one tree and $\{r, s, \psi\}$ to represent parameters of the other tree.
Since the mixture has a $12|34$ topology, it must satisfy the phylogenetic invariants, (3), (4) of the $12|34$ tree. We first solve for invariant (3),

$$(\alpha+1-\alpha)(\alpha\, q_{1234}+(1-\alpha)\, r_{1234})-(\alpha\, q_{12}+(1-\alpha)\, r_{12})(\alpha\, q_{34}+(1-\alpha)\, r_{34}) = 0. \tag{5}$$

This can be simplified to get,

$$q_{1234} + r_{1234} - (q_{12}\, r_{34} + r_{12}\, q_{34}) = 0. \tag{6}$$

$$(q_{12} - r_{12})(q_{34} - r_{34}) = 0. \tag{7}$$

Substituting (2), we get,

$$(\theta_1\, \theta_2 - \psi_1\, \psi_2)(\theta_3\, \theta_4 - \psi_3\, \psi_4) = 0.$$

This means that the branch lengths should satisfy,

$$t_1 + t_2 = s_1 + s_2 \quad \text{or} \quad t_3 + t_4 = s_3 + s_4. \tag{8}$$

We now solve for invariant (4),

$$q_{13}\, r_{24} + r_{13}\, q_{24} - (q_{13}\, r_{24} + q_{13}\, r_{24}) = 0. \tag{9}$$

Substituting (2), we get,

$$(\frac{\theta_1}{\theta_2} - \frac{\psi_1}{\psi_2})(\frac{\theta_3}{\theta_4} - \frac{\psi_3}{\psi_4}) = 0.$$

This means that the branch lengths should satisfy,

$$t_1 - t_2 = s_1 - s_2 \text{ or } t_3 - t_4 = s_3 - s_4$$

So to summarize if we mix two $12|34$ quartet trees, then we will get a $12|34$ tree if the following conditions on the edge lengths are satisfied,

$$t_1 = s_1 \quad \text{and} \quad t_2 = s_2 \tag{10}$$

or

$$t_1 + t_2 = s_1 + s_2 \quad \text{and} \quad t_3 - t_4 = s_3 - s_4 \tag{11}$$

We note that α gets factored out and doesn't appear in any of the later equations. This means that if we have branch lengths such that the mixture satisfies the phylogenetic invariants for a single α then it is satisfied for all α.

Proposition 2: A mixture of two $12|34$ quartet trees with branch lengths $t_1 = s_1$ and $t_2 = s_2$ will have resulting branch lengths $t_1(= s_1)$ and $t_2(= s_2)$ respectively.

Fidelity of the mixture, μ_1 is given by

$$\mu_1 = \sqrt{\frac{(\alpha\theta_1\theta_2 + (1-\alpha)\psi_1\psi_2)(\alpha\theta_1\theta_5\theta_3 + (1-\alpha)\psi_1\psi_5\psi_3)}{\alpha\theta_2\theta_5\theta_3 + (1-\alpha)\psi_2\psi_5\psi_3}}$$

Substitute $\psi_1 = \theta_1$ and $\psi_2 = \theta_2$ to get $\mu_1 = \theta_1$

Proposition 3: If we mix two $12|34$ trees to get a tree having the same $12|34$ topology then the branch length of the mixture can be arbitrarily small even if the corresponding branch lengths in both the original trees are large.

Substitute $\theta_1 = \psi_1$, $\theta_3 = \psi_3$, $\theta_4 = \psi_4$, $\theta_5 = \psi_2$ and $\alpha = 0.5$. Mixture clearly satisfies the phylogenetic invariants. Fidelity of the mixture, μ_1 is given by

$$\mu_1 = \frac{\theta_1|\theta_2 + \theta_5|}{\sqrt{\theta_2\theta_5}}$$

So even if θ_1 is close to 0, if we choose the factor $\frac{\theta_2}{\theta_5}$ to be close to 0, μ_1 will be approximately 1 which means that the edge length of the mixture in pendant 1 is close to 0.

We will specify the conditions (only a subset of all the conditions) for which the mixture of two $12|34$ quartet trees gives a $13|24$ quartet tree.

Say that we can find numbers k_i that satisfy

$$k_1 > k_3 > k_4 > 1 > k_2 > 0$$

$$\frac{1-k_1^2}{k_1}\frac{1-k_4^2}{k_4} + \frac{1-k_2^2}{k_2}\frac{1-k_3^2}{k_3} > 0$$

$$\frac{k_1+k_4}{1+k_1k_4}\frac{k_2+k_3}{1+k_2k_3} > 1.$$

Then for branch lengths t_i and s_i such that $k_i = \exp(-2(t_i - s_i))$ there exists mixing weights α, such that the mixture of the two $12|34$ quartets gives a $13|24$ quartet .

Reference

F. A. Matsen and M. Steel (2007), "*Phylogenetic mixtures on a single tree can mimic a tree of another topology*", Systematic Biology, Vol. 56, 767–775.