

Issues/Observations

1) Must find the MLE for all $(2n-5)!!$ trees, tree space is huge!

NP-hard

2) Maybe there is not a unique MLE...

3) Optimization techniques for a fixed T are well-developed. (Felsenstein pruning algorithm)

4) ML is an attractive method since

- it's well grounded in Statistics

- model assumptions are explicit

Eg: Is a stable base dist. reasonable?

Common Q?

5) Searching tree space ... How?

6) Is iid reasonable for data analysis?

Since for even moderately-sized n = number of taxa, a full search of

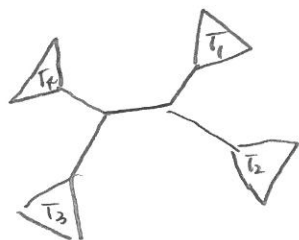
tree space is impossible, for ML and parsimony searches certain

"moves" in tree space are typically used. We discuss three of

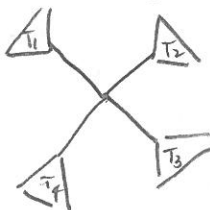
these from the most local to the most general.

(I) Nearest Neighbor Interchanges or NNI moves

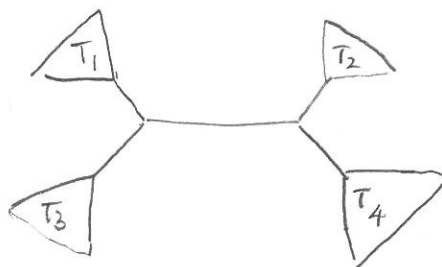
Start: Pick an internal edge



• Collapse edge



• Resolve polytomy by random
(Pick 2 out of 4 subtree
to be sister trees.)



(II) Subtree Prune and Regraft or SPR moves

Handout

(III) Tree Bisection and Reconnection or TBR moves.

An approximate count of the number of such moves from a fixed tree is

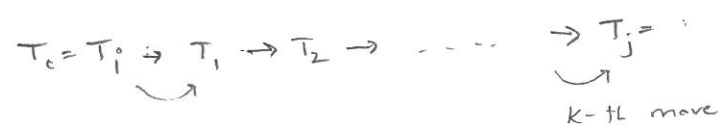
NNI	$2(n-3)$	2 for each internal edge
SPR	$2(n-3)(2n-7) = O(n^2)$	
TBR	$\approx n^3$	$O(n^3)$

Almost all phylogeny software uses NNI as default search strategy, but allow the user to change to a different strategy.

Each of these types of moves defines a metric on tree space

and for a fixed tree there are roughly $O(n), O(n^2), O(n^3)$ tree is a neighborhood 1-move away.

Defn: Given 2 n-taxon trees, then the distance d_{NNI} (or (d_{SPR}, d_{TBR}) is $d_{NNI}(T_i, T_j)$ is the smallest number k such that you can get from T_i to T_j in k moves



In general, particularly for TBR moves, these distances are very hard to compute.

There is therefore another distance - the Robinson-Goulds distance - on tree space that is often used to quantify similarity/distance between trees

Defn: For two trees T_1, T_2 , the Robinson-Goulds distance between them is

$$d_{RF}(T_1, T_2) = \# \text{ of splits that occur on exactly 1 of trees} \\ (\text{but not both}).$$

This is also called the SPLITS METRIC, SYMMETRIC DISTANCE, etc.

It is easy to compute and is often used in simulation studies to compare reconstruction accuracies of various tree construction methods.

Eg. Compute the RF-distance for all pairs of 4-taxon trees 0, 2, 2