

Master thesis proposal

Fitch's algorithm for ordinal variables

Krzysztof Bartoszek

October 6, 2022

Background — Fitch's algorithm

Fitch's algorithm [1] is one of the most well known methods for finding the most parsimonious assignment of ancestral states on a phylogeny for categorical traits (e.g. compound or simple eyes). The most parsimonious assignment means that the fewest amount of changes had to take place over the course of evolution. However many categorical variables have a natural hierarchy connected to them, e.g. small, medium, large habitat size. Hence, this implies that some levels are more similar to each other.

Thesis project

The aim of the project would be to first develop appropriate simulation methods for ordinal variables evolving on a phylogeny. The framework would be a branching Markov chain (for which there are multiple implementations, e.g. in the **phytools** R package), but these would have to be specialized to include the ordering. Then, the next step would be to modify Fitch's algorithm to take into account the similarities when penalizing. A starting point is to use this information when there are multiple assignments for a particular node, resulting in the same parsimony score. Then, one can compare how this approach compares with maximum likelihood assignments (which are extremely time consuming), and explore the effects of different similarity measures between the levels.

Data

The project will be predominantly illustrated with simulated data. Possibly data sets attached to the `ape` or `phytools` packages can be explored.

References

- [1] M. W. Fitch. Defining the course of evolution: Minimum change for a specific tree topology. *Systematic Zoology*, 20:406–416, 1971.