Model selection for phylogenetic mixtures via linear invariants

A. Kedzierska[1,2]   M. Casanellas[1]
anna.kedzierska@crg.es   marta.casanellas@upc.edu

J. Fernàndez-Sànchez[1]
jesus.fernandez.sanchez@upc.edu
1. Dpt. Matemàtica Aplicada I , Universitat Politècnica de Catalunya
2. Center for Genomic Regulation, PRBB, Barcelona

24.09.2010

Abstract

Phylogenetic data comprise the aligned sequences of contemporary taxa. It is of main interest in phylogenetic inference to estimate the parameters of the model: the underlying relalationship depicted in the form of a phylogenetic tree and the continuous parameters, that is the evolutionary distances between two consecutive nodes in the tree. In practice the data usually arises from different tree topologies resulting in the mixed model for a given alignment. The existing approaches require prior knowledge of the model in order to perform the inference step. However, the model selection step calls for the input tree as well. This contradictory argument questions the validity of the estimates. In addition, none of the methods incorporates the inherent mixed nature of the data.

In this work we propose an approach for model fitting of algebraic models on m-tree mixtures. It is well known that the probabilities of patterns observed at the leaves satisfy specific equalities if the tree evolves under certain models. Our model selection method is based on non-phylogenetic linear invariants, that is linear equations on the joint distribution at the leaves, that vanish on any mixture of the tree topologies under a given model.

References

[1] Joseph Felsenstein *Inferring Phylogenies.* Syst. Biol., **53**(4) (2003), 669–670

[2] Jan Draisma and Jochen Kuttler. *On the ideals of equivariant tree models.* Math. Ann., **344**(3) (2009), 619–644

[3] Elizabeth S. Allman and John A. Rhodes. *Phylogenetic ideals and varieties for the general Markov model.* Adv. Appl. Math., **40**(2) (2008), 127-148