



Quartet Methods for Phylogenetic Inference Under the Coalescent

The advent of rapid and inexpensive DNA sequencing technologies has necessitated the development of computationally efficient methods for analyzing sequence data for many genes simultaneously in an evolutionary framework. The coalescent process is the most commonly used model for linking the underlying genealogies of individual genes with the global species-level phylogenetic tree, but inference under the coalescent model is computationally daunting in the typical inference frameworks (e.g., the likelihood and Bayesian frameworks) due to the dimensionality of the space of both gene trees and species trees. By viewing the data arising under the phylogenetic coalescent model as a collection of site patterns, the algebraic structure associated with the probability distribution on the site patterns can be used to develop computationally efficient methods for inference via phylogenetic invariants. In this talk, I will describe how identifiability results for four-taxon species trees based on site pattern probabilities can be used to build a quartet-based inference algorithm for trees of arbitrary size. I will also show how a composite likelihood approach based on quartets can be developed to obtain estimators of the branch lengths within the tree that are consistent and asymptotically normal. I will demonstrate the performance of the methods by applying them to both simulated and empirical data. Because these methods are derived in a fully model-based framework, they are promising approaches for computationally efficient, model-based inference for the large-scale sequence data available today.

Laura Kubatko

Departments of Statistics and
Evolution, Ecology, and
Organismal Biology
Co-Director,
Mathematical Biosciences Institute
The Ohio State University

Virtual Colloquium / 3:30 to 4:30 p.m. Eastern Time (U.S. and Canada) / Wednesday, September 16, 2020