Modeling Molecular Evolution: Markov models of DNA mutation on trees

Elizabeth Allman
University of Southern Maine

Thursday, February 10, 2005
L02 Carson Hall, 4:00 pm
(Tea 3:30 pm Math Lounge)

Abstract

The problem of tracing evolutionary relationships between species has long interested scientists. In recent years, as DNA sequences have become readily available, researchers in the areas of biology, mathematics, statistics, and computer science have been working to develop new techniques of phylogenetic tree construction using this data source.

In a model-based method of phylogeny reconstruction, one assumes that a particular probabilistic model governs the mutation process. Statistical techniques such as maximum likelihood can then reconstruct phylogenetic relationships fairly reliably. For any model-based method of phylogenetic inference to be accurate, however, it is essential that the model 'fits' the data. One way to begin to address the question of model fit is by identifying the polynomial relationships that patterns in the sequence data must satisfy. Such constraints are known as *phylogenetic invariants*.

This talk will give a brief overview of models of molecular mutation, highlighting a mixture model known as the general Markov plus invariable sites model (GM+I). In the latter part of the talk, we explain our work as a problem in 'algebraic statistics', where we exploit the observation that a model of molecular evolution gives rise to a parameterization of an affine algebraic variety. Finally, we explain how this viewpoint has led to results including a proof of the identifiability of parameters for the GM+I model (which is necessary to establish the consistency of maximum likelihood methods) and suggests possible improvements to estimates of parameters of interest to biologists.

This talk should be accessible to graduate students.