Statistics 992: Statistical Phylogenetics

Professor: Bret Larget
Office: 4390 CSSC (Computer Science and Statistics Center)
Phone: 262-7979
E-Mail: larget@stat.wisc.edu
Web: http://www.stat.wisc.edu/~larget/

Office Hours: Wednesday 9:45 – 12:00 noon, and by appointment
Class Hours: MWF 8:50 – 9:40 A.M.
Room: 1289 CSSC (Computer Science and Statistics Center)
Textbook: *Inferring Phylogenies*, by Felsenstein

Prerequisites: I intend to make the course accessible to both graduate students in Statistics and in the biological sciences.

Statistics students should have completed (or be taking concurrently) Statistics 610 or a more advanced course. I will not assume any prior knowledge in the biological sciences.

Biology students should have completed (or be taking concurrently) any two courses in statistics or probability (such as Statistics 571/572), preferably at the graduate level.

Topics: We will supplement Felsenstein's book *Inferring Phylogenies* with readings in the primary scientific literature as well as with occasional lecture notes.

The course will include these topics: (1) mathematical description of phylogenetic trees, (2) the estimation of phylogenetic trees from aligned DNA sequence data using maximum likelihood, parsimony, distance, and Bayesian methods, (and supporting probability and statistics topics including likelihood, continuous-time Markov chains, the parametric and nonparametric bootstrap, and Markov chain Monte Carlo), (3) comparisons of statistical properties of different phylogeny estimators, (4) the comparison of the bootstrap and Bayesian posterior probabilities for assessing uncertainty in phylogeny estimation, (5) the estimation of phylogeny from genome arrangement data, and (6) additional topics as time permits. Possible additional topics include statistical tests of tree topology, model selection, and statistical models of coevolution.

Course Structure: I anticipate spending about 1/3 of the course time as a seminar where we will discuss readings, usually on Mondays, but possibly spilling over onto other days. Discussion homework will include reading the material prior to the time of discussion and submission of questions about the reading (in advance!) that will guide the discussion. The remaining class time will be for lectures.

Course Goals: My objective for statistics students is to provide a sufficient background in statistical phylogenetics so that students are positioned to begin exploring their own research questions if they desire to continue beyond the course. My objective for biology students is to provide a much deeper understanding of the statistical background underlying widely-used methods of phylogenetic inference than they have at the beginning of the course. My objective for both groups of students is to provide the opportunity to learn from one another and to gain experience in cross-disciplinary communication and collaboration.

Student Evaluation: Students may take the course for one, two, or three credits and will be evaluated on a corresponding number of areas, equally weighted. For example, students taking the course for one credit may choose to be evaluated on the basis of discussion homework only. Grades for students taking the course for three credits will be the average of grades from each of the three areas.

Discussion homework: I will assign weekly reading from the textbook or other sources. Students are expected to read the assignment prior to the class discussion time, normally on Mondays. One portion of the homework is to send to me by midnight on the preceding Saturday via e-mail two questions regarding the reading. I will organize these questions and distribute them to everyone at the beginning of class on Mondays to use as a basis for discussion. Participation in these discussions is also a basis for evaluation. In addition, I may ask students to prepare a five minute summary occasionally during the semester to help facilitate the beginning of a discussion.

Written homework: Written homework will be assigned approximately every other week. Assignments will include calculations and small derivations as well as small coding exercises.

Project: A project (preferably with a cross-disciplinary group of two or three students) will explore a topic of the course in greater depth. A project that addresses a question through a simulation study with a written report and summary of conclusions would be sufficient, but I am willing to entertain other possibilities.