Mathematics 660: Advanced Mathematical Modeling
(Theory of Phylogenetics)
Spring 2016

Instructor: John Rhodes

Office: 102 Chapman, 474-5445

E-mail: `j.rhodes@alaska.edu`

Office Hours: M F 11:30-12:30, T 2:00-3:00, and by appointment

Web page: `http://www.dms.uaf.edu/~jrhodes/M660.html`

Prerequisites: Flexible. Useful background material includes linear algebra, probability and statistics, and some biology. However, a motivated student proficient in one area can pick up the basics as we proceed. In particular, if requested I'm willing to give (out-of-class) tutorials on mathematical tools that biologists may not be familiar with.

Credit Hours: 3.0

Textbook: <u>Lecture Notes: The Mathematics of Phylogenetics</u>, by E. S. Allman and J. A. Rhodes. A version from the last offering of this class is posted on the class web page, but it will be edited and a new version posted chapter by chapter as the term progresses.

Class Meetings: MWF 10:30-11:30, in Bunnell 313 at UAF, videoconferencing at UAA Gordon Hartlieb Hall 101E

Exams: Midterm: Friday, March 11, in class;
Final: Tuesday, May 3, 10:15am-12:15pm

Course overview and goals:

Phylogenetics is concerned with inferring evolutionary relationships among a collection of organisms, such as representatives of different species. Most commonly these days, inference is from DNA or protein sequence data, though historically morphological characteristics were used. Although phylogenetic problems arise from biology (and linguistics, with DNA replaced by the words in a language), they are rather quickly abstracted to mathematical ones. The techniques used to address them involve a mix of mathematics, statistics, and computer science, but are informed by an understanding of what is biologically plausible.

The early mathematical approaches to analyzing phylogenetic data were primarily combinatorial in nature, with a strong algorithmic flavor. Although these methods were quickly adopted to provide new biological insights, and are still important, they lacked a firm statistical basis. Making evolutionary inferences

in a statistical framework requires formulating an explicit mathematical model of the evolutionary process. Although such a model will at best be a caricature of the true process, by capturing the most important features of the process it can lead to better inference procedures. Understanding the limitations, both biological and mathematical, of phylogenetic models is necessary to use them wisely.

This course will develop all the main approaches to phylogenetic inference, including parsimony and distanced-based methods, but emphasizing the model-based methods of maximum likelihood and Bayesian analysis. Throughout, the emphasis will be on understanding the mathematics behind these, so that strengths and weaknesses of each are highlighted. While software implementing various approaches will be mentioned, the goal is not to make you a proficient user of any package, but rather to make the packages something other than 'black boxes' which analyze data for you. It should also point you toward some of the future developments that are likely to occur as genomic scale data becomes more readily available.

Phylogenetics is interdisciplinary; biology, computer science, statistics, and mathematics all play important roles. The students in this class come from several of these fields, and can contribute by providing differing disciplinary perspectives. No one person is likely to have the 'right' background across the board. What is more important is a willingness to contribute and ask questions.

Mechanics of the course:

Lectures The course will be delivered as a lecture, but I prefer to keep it informal. I am happy to be interrupted with questions and comments at any point. It is particularly important to speak up if you do not understand some background material. You'll be doing a favor to other students who are too shy to ask.

Some students will be taking the course remotely, through videoconferencing. In my experience, this tends to discourage students from speaking up —whether they are in the room with me, or not. Please try to fight this. Asking question during lectures will benefit everyone.

OIT has said they will be recording lectures, so if you miss class, you should be able to watch it anyway. Check the class web page for any handouts and homework assignments.

Textbook There really is no textbook that is ideal for this course — published ones tend to give either too little mathematical explanation to develop real understanding, or are written to be read by professional statisticians and/or mathematicians. While I will recommend some of these during the semester, we will follow a set of course notes that a collaborator and I have been developing. As these notes are edited and extended every time we use them, I will post them on the webpage in sections. (If you prefer a print copy, it is up to you to make it.)

Homework You are expected to read and assimilate assigned sections of the course notes. Readings may sometimes include material dealt with only cursorily, or not at all, in lectures.

Homework problems will usually be assigned daily, but only collected each Wednesday (due in class, but accepted until 5pm). If you are at UAF, I prefer you give me hard-copy homework solutions; remote students can of course send work in by e-mail, in the form of a pdf.

Homework problems will be posted on the course web page as they are assigned. Those students who are based in the mathematical sciences will have additional problems not required for those based in biology.

Late homework that has not been approved ahead of time will not be accepted. *There will be no exceptions to this* other than for a genuine emergency (e.g., a death in the family, documented illness, etc.).

I encourage you to work with others on the homework, *but to write up the solutions independently.* In writing up your work, you should present your arguments in such a way that an intelligent, but ignorant, person can understand them. In particular complete sentences and a logical presentation are expected.

Examinations Both the midterm and final examinations will be in-class, closed-book exams. For missed examinations that are not approved in advance, no make-up exams will be given, except in case of emergencies.

Project Each student will complete a project in this course, which can be tailored to particular interests. You may work on these singly or in pairs.

What makes a good project will depend on your primary field (Biology, Mathematics, Computer Science, etc.), and the background with which you enter this course. Possibilities include delving more deeply into a topic introduced in class by reading and presenting appropriate research papers, developing your own ideas for a variation on some phylogenetic method, writing software to implement some calculation, exploring a real data set thoroughly through many different phylogenetic approaches, or using simulated data to investigate the limits of some of these approaches. I am open to other ideas as well – the only thing ruled out is performing a 'straightforward' phylogenetic analysis using standard software.

Soon after the spring break, you will submit for approval a brief description of the project you would like to undertake.

Grades Your performance will be evaluated based on 15% homework, 5% class participation, 30% midterm exam , 35% final exam, 15% project. Course grades will be determined according to the following cutoffs:

$$A \geq 90\%, \quad B \geq 80\%, \quad C \geq 70\%, \quad D \geq 60\%.$$

I reserve the right to move the cutoff points downward if particular exams turn out to be unexpectedly difficult. Note that you are not in competition with your peers – everyone in the class may get an A, or everyone may get an F.

University and Department Policies Your work in this course is governed by the UAF Honor Code. The Department of Mathematics and Statistics has specific policies on incompletes, late withdrawals, and early final exams which can be found at
<center>http://www.dms.uaf.edu/dms/Policies.html.</center>

If you have any disabilities that your instructors should know about, bring them to our attention soon so that we can work with the Office of Disability Services to set up any necessary accommodations.

Theory of Phylogenetics
MATH 660: Advanced Mathematical Modeling

Tentative Schedule

Week	Dates	Topics
0/1	Jan 15, 20	Molecular evolution, combinatorics of trees
2	Jan 25	Parsimony, unweighted and weighted
3	Feb 1	Splits, clades, consensus trees and supertrees
4	Feb 8	Distance algorithms
5	Feb 15	Probabilistic models of DNA mutation
6	Feb 22	Probabilistic models (cont.)
7	Feb 29	Phylogenetic distances
8	Mar 7	Maximum likelihood; Midterm (Friday)
	Mar 14	Spring break
9	Mar 21	Tree space
10	Mar 28	Rate-variation models, codon and protein models, Covarion models, Lewis model, Mixture models
11	Apr 4	Model testing, bootstrapping, consistency, long branch attraction
12	Apr 11	Bayesian methods
13	Apr 18	Species tree/gene tree concordance and the coalescent model
14	Apr 25	Species tree/gene tree inference methods