Syllabus: Fitting Models to Data: Maximum Likelihood Methods and Model Selection (Biol. 304/404), Spring 2011

Course Description

This course will introduce students to maximum likelihood methods for fitting models to data and to ways of deciding which model is best supported by the data (model selection). Along the way, students will learn some basic tenets of probability and develop competency in R, a commonly used statistical package. Examples will be drawn from ecology, epidemiology, and potentially other areas of biology. The second half of the course is devoted to in-class projects, and students are encouraged to bring their own data. This course satisfies a laboratory requirement for the biology BA/BS, a quantitative laboratory requirement for the biology BS, and a systems elective requirement for the systems biology BS. Offered as BIOL 304, BIOL 404.

Prerequisites: MATH 121, 122 or MATH 125, 126

Instructor

Robin Snyder
office: 510 Millis
phone: 368-8838
email: res29@case.edu

Teaching Assistant

Rob Meklemburg
office: 5th floor of Millis, probably 506B
phone: TBD
email: revbobmek@gmail.com

Office hours

TBD. During the semester I am happy to meet with anyone enrolled in this course. I would like especially to meet with students with disabilities who are registered with the Coordinator of Disability Services (368-5230) and who may need individual arrangements.

Textbook

Ecological Models and Data in R, Benjamin M. Bolker, Princeton University Press, 2008.

Tentative course timeline

- 1/11, 13, 18: Pistachio Lab! Learning R. Work through lab 1. Start reading ch. 2.

- 1/20, 25: Exporatory data analysis. Work through lab 2. Finish ch. 2 and start reading ch. 3.
- 1/27: Deterministic models (modeling something we're interested in as a function of other quantities). Work through lab 3. Start reading ch. 4.
- 2/1, 3: Stochastic models (modeling the noise around the value predicted by the deterministic model). We'll cover an introduction to the rules of probability and look at a variety of useful distributions. Do lab 4. Start reading ch. 6.
- 2/8: Maximum likelihood. Finally! Ch. 6 through the end of section 6.3, excluding the Bayesian section. Start lab 6.
- 2/10: ML. Section 6.4.1
- 2/15: ML. Section 6.5
- 2/17: Derivative-free vs derivative-based optimization: ch. 7 through section 7.2.
- 2/22: Model selection and criticism. Sections 6.6.2, 6.6.5, 6.6.6.
- 2/24: flex time — time series methods?
- 3/1: flex time — time series methods?
- 3/3: flex time — time series methods?
- *Spring break*
- 3/16–4/22: team projects.

Course format

The first half of this course will combine occasional mini-lectures with exercises to be worked on in class. I plan to devote the great majority of class time to these exercises, so that you will have the benefit of my presence and your classmates' as you work. However, the lack of formal lectures means that you will need to do a lot of work on your own. **You will need to spend time outside of class working on the exercises. You will need to spend time outside of class reading the book before we get to that material in class.**

As a way of encouraging you to read the book (did I mention you'll need to **read and work outside of class**?), there will be small, frequent homework assignments on the reading — I may ask you to answer some questions, I may ask you to explain the material as to an intelligent 12-year-old — something that will reassure me that you're keeping up and help you justify the time spent reading[1]. **If you're confused by the reading and want me to go over something in class, these assignments are a good place to tell me that.** My occasional mini-lectures will be driven in part by what people tell me they need me to go over.

In the second half of the course, teams will work on projects. You may reproduce and hopefully extend the results of a published paper. You may work on your own model of data you've found online. You may work on data of your own, if you have something suitable. I value cooperation and strongly encourage students to help each other: part of your grade will depend on your active participation in class and your willingness to help others. In lieu of a final exam, each student will submit a paper presenting their model and analysis. Those enrolled in the graduate version of this class will be expected to do a project of their own.

Assessment

Class participation: 20%
Homework: 15%
Quizzes: 15%
Modeling project: 50%

- *Project*: The anaylsis will count for 75% of your grade and the quality of the writeup (including grammar, sentence structure, organization, etc.) will count for 25%. The paper should be as long as it needs to be (and no longer, please!) to adequately present the problem, your analysis, and your conclusions. Papers are typically 6–9 pages. I will deduct 10 points for every day a paper is late.

- *Participation*: Class participation will be evaluated by your active work to understand the techniques we are studying and keep to our agreed schedule and by your willingness to help other students. More specifically:

  - On the day each lab is due, I will email you a short exercise or two that uses skills from that lab, due (electronically) by midnight the following day. Think of these as low-stress, take-home quizzes that help me to verify that you're keeping up.
  - As described below, I will ask you to evaluate how your team is working each week. These scores will not directly determine your grade. However, if your teammate consistently states that you are not contributing your share or that you are not interacting well, despite warnings from me, it will

[1] I mean, aside from self-preservation: if you don't understand the material by the time you do your project, you're doomed.

lower your participation grade. I will also, of course, have the opportunity to observe the dynamics of your team directly.

- *Teamwork*: I will ask you to email me a simple evaluation form each week. The email should be headed "Team Evaluation". Rate your contribution, your teammate's contribution, and the overall functioning of your team on a scale from 1 to 3. A score of 1 means "needs improvement," a score of 2 means "good but not exceptional," and a score of 3 means "very good." Please also note any students from other teams who have been helpful to you.

  For example, if you have worked very hard, your teammate has also worked hard, but the two of you have not interacted very much, your email might consist of the numbers 3, 3, 1.

  In contrast, if your teammate has been making very helpful suggestions, you have worked very hard, but your teammate hasn't been able to do much this week, your email might consist of the numbers 3, 1, 3.

  If you didn't do much during the week, your teammate worked very hard, and the two of you didn't really discuss anything or work on anything together, your email might consist of the numbers 1, 3, 1.

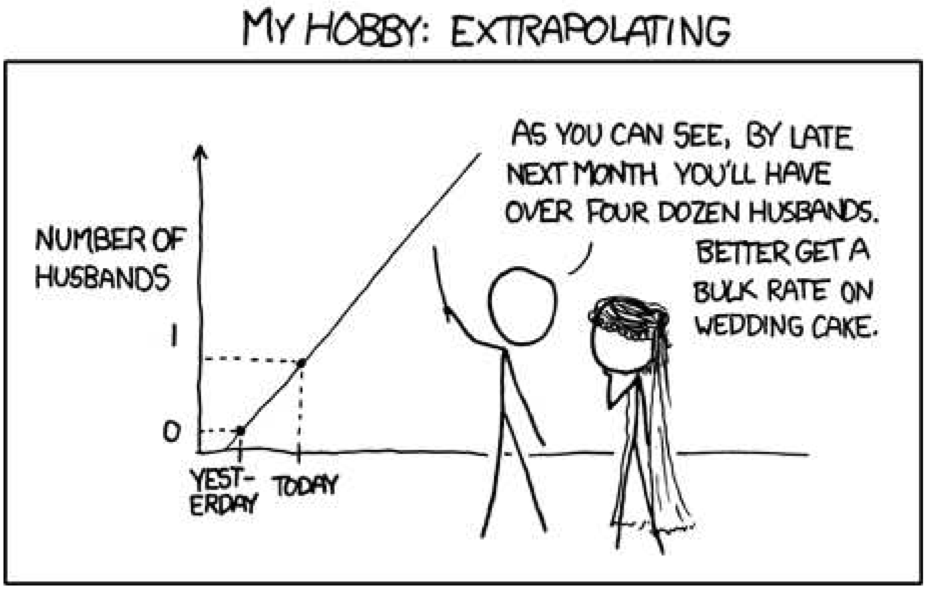
  Evaluations will be due every Friday morning by 10 AM.

  If the two team members are unable to work well together, I may need to intervene. If necessary, I may re-assign students to different teams, but I hope this will not be necessary.

Class culture

I hope that you will work closely together. Instead of doing the labs largely on your own, speaking to your teammate only when you have a question, I encourage you to work as pairs, possibly off of a single laptop. The classroom should be noisy! I also encourage you to talk to other teams if you're having trouble. Real-world problems have no professors and no TAs. They do have colleagues.

Along the same lines, students in this class have a wide variety of mathematical and statistical backgrounds. I expect those of you with stronger backgrounds to act as teachers to those with weaker backgrounds. That doesn't mean telling someone exactly what to do — not much learning results that way. Rather, guide each other toward understanding.


MY HOBBY: EXTRAPOLATING
NUMBER OF HUSBANDS
1
0
YEST-ERDAY
TODAY
AS YOU CAN SEE, BY LATE NEXT MONTH YOU'LL HAVE OVER FOUR DOZEN HUSBANDS.
BETTER GET A BULK RATE ON WEDDING CAKE.