

Syllabus

Stat 212A, Fall 2015: Topics in Selective Inference

Instructor: Will Fithian

Course Description

Survey of the theory and methods of selective inference. As large, open-access data sets play an increasingly central role in driving scientific discovery, the scientific method must evolve to support credible open-ended exploration. To meet this challenge, statisticians must develop methods for rigorously answering questions suggested by the data. The problem of selectivity manifests itself in diverse ways, and this course will reflect a corresponding diversity of solutions. Topics covered include: classical multiple testing, false discovery rate and false coverage-statement rate control, empirical Bayes methods, and inference after model selection.

Course Outline (Tentative)

- Hypotheses, families, selection, and type I error rates
- Comparing normal means
- Methods based on p -values
- Stepdown procedures and the closure principle
- The false discovery rate and the Benjamini-Hochberg procedure
- False discovery rate under dependence
- False coverage-statement rate (FDR for CIs)
- Empirical Bayes perspective on FDR
- Local FDR
- Post-selection inference in regression
- Post-selection inference via conditional inference
- Replicability and meta-analysis
- If time: high-dimensional inference, Bayes and selective inference, other topics by request?

References No textbook covers all of the above topics, and some are not covered by *any* existing textbook. For the classical segment of the course (first few weeks), *Multiple Comparison Procedures* by Hochberg and Tamhane and *Multiple Comparisons: Theory and Methods* by Hsu are good references. Efron's *Large-Scale Inference* is an excellent reference for the empirical Bayes segment.

There is a list of references online on the course web page, which I will fill in through the course of the semester.

Prerequisites Stat 210 or equivalent. Concepts from 205 will come up occasionally but are not ubiquitous. Please speak with me if you are unsure of your background.

Administrative

Office Hours Evans 301, Tuesdays 2–3:30pm, Wednesdays 10:30am–12pm, or by appointment.

Email wfithian@berkeley.edu, for administrative questions. For content-related questions, office hours are a much better venue than email.

Web Site www.stat.berkeley.edu/~wfithian/courses/stats212a

Course Work

Homework 3 ± 1 problem sets reinforcing course material, 30% of final grade.

Project Proposal 2-page proposal of a project topic, due Monday October 12, 10% of final grade.

Final Project To be presented in class, 60% of final grade. You can

- read and report on a paper
- solve a new selective inference problem
- analyze one of the techniques we cover, or relationships between them
- contribute to a software package

I suggest you visit my office hours to discuss possible projects.

Late Assignments Due dates will be strictly enforced (zero for late work), but the lowest homework grade will be dropped.

Academic Integrity When solving the problem sets you are welcome to work with each other or consult articles or textbooks online, but you should write up the problem by yourself. If you collaborate or use online resources, you should list your collaborators and cite the resources you used.

Questionnaire

I'd like to get to know you better, so please write on an index card:

- Your name
- Nickname (i.e. what you want me to call you)
- Department
- Email address (if you want to receive course emails)
- Most advanced math/stat courses you've taken (up to Stat 210)
- Why you're interested in this subject matter