Math 541b: Introduction to Mathematical Statistics
Fall 2022

Instructor: Dmitrii M. Ostrovskii, email: `dostrovs@usc.edu`, office: KAP 406h
Grader: TBA, email: `TBA`, available via zoom
Schedule/Classroom: M/W/F 11-11:50 am, KAP 140
Office hours: Monday 12:30-2:15 pm for the instructor, TBA for the grader

Outline of the course. This course is supposed to complement Math 541a as a graduate-level introduction to Mathematical Statistics. The goal is to understand the key principles behind statistical estimation and testing, and learn how to apply these principles in concrete scenarios.

We shall first briefly recap the main results of Math 541a: the maximum likelihood principle and the method of moments; sufficiency; the basics of asymptotic theory (consistency, optimality, and asymptotic efficiency). Then we shall cover hypothesis testing, including the concept of p-values and connection with interval estimation. Overall, this will take about 2/3 of the term.

In the remainder of the semester, we shall consider some "assortment" of more modern topics: bootstrapping techniques, Bayesian methods such as expectation-maximation and Markov-Chain Monte Carlo; linear and generalized linear models, and, possibly, robust estimation.

Contact. The best way to contact me is by email (**put Math 541b in the subject field**) or in-person during the office hours. If you have a schedule conflict with my office hours, send me an email, and we'll arrange for an alternative time.

Prerequisites. Working knowledge of point estimation (M541a), Probability at the level of M407/505a, Analysis (M425a,b), and Linear Algebra (M471). Measure-theoretic Real Analysis (M525a) and Probability as per M507a are not required, but might prove to be useful on a few occasions. Basic knowledge of convex analysis is also useful, but not required: some background will be given when necessary.

Covered topics. Some topics are optional due to limited time; they are marked with "?".

1. <u>Recap of M541a</u>
   - Brief review of the probability concepts: probability spaces, random variables, expectations, moment-generating function, basic inequalities (Jensen, Markov, Chebyshev, and Chernoff); conditional expectations and independence.
   - Point estimation: consistency, laws of large numbers, central limit theorem, maximum likelihood principle and sufficiency.
2. <u>Hypothesis testing:</u>
   - Problem statement and terminology.
   - Likelihood ratio test (LRT) for two simple hypotheses; the Neyman-Pearson lemma.

- Testing composite hypotheses: uniformly most powerful (UMP) tests, monotone likelihood ratio families and the Karlin-Rubin theorem; locally most powerful tests via the generalized Neyman-Pearson lemma. Neyman's structure; conditioning method.
- UMP unbiased tests in exponential families. Permutation and rank tests; invariance.
- Constency of LRT when testing two simple hypotheses. Asymptotics of the power function and Pittman's efficiency. LRT and Wilks' theorem.
- Connection between interval estimation and hypotheses testing.

3. <u>"Data-intensive" inference:</u>
   - Bootstrap and jackknife.
   - Expectation-maximization (EM) algorithm.

   ? Markov Chain Monte Carlo (MCMC).

4. <u>Topics in modern statistics:</u>
   - Estimation in linear models; finite-sample results.

   ? Generalized linear models (GLMs).

   ? Basics of robust estimation; median-of-means (MoM) methodology.

Homework assignments. There will be 5 to 6 homework assignments, designed in a way to help you understanding the material covered in the course. Here is some advice:

- Always start to work on homework assignments as early as possible. (Well, at least just *look* in there, simply in order to estimate the difficulty of what you'll have to deal with...)
- It's okay to work in a small group. However, always try to solve the problem yourself first—especially those of you who do research or planning to pursue it. If you are blocked, use the office hours to ask a question. Taking notes during a discussion is a useful practice.
- Always write the solution on your own; suspected plagiarism will be mercilessly punished. Please make sure your handwriting is readable (by someone other than yourself, that is).

Exams. The midterm will take place during a lecture after the Fall break—most likely, in the late October to early November—in our regular classroom. The (comprehensive) final exam is scheduled on Wednesday, **December 7**, from 11 am to 1 pm, in our regular classroom.

Grading. 40%/25%/35% for homeworks/midterm/final exam; can be slightly adjusted later.

Attendance. I shall not keep track of attendance: I rely on your motivation. That said, note that the class is in-person, and no online option is offered—unless in exceptional circumstances.

Literature. Your primary reference are the (comprehensive) lecture notes left from the previous iteration of the course; they shall be available online. *Statistical Inference* by Casella and Berger (2nd ed.) is the main reference for the first part of the course. *Testing Statistical Hypotheses* by Lehman and Romano is a good reference for the part on hypotheses testing. The final "miscellanea" topics do not follow a single book, so the lecture notes will be useful.

Other useful references to get a perspective on asymptotic theory of estimation (circa 1980s) are: *A Course in Large Sample Theory* by Ferguson and *Asymptotic Statistics* by Van der Vaart. A comprehensive text on *modern* statistical theory is *Basics of Modern Mathematical Statistics* by Spokoiny and Dickhaus; however, it is way beyond the scope of this course.