Syllabus

1 Logistics

- Instructor: Tselil Schramm (tselil@stanford.edu).
- Course website: www.tselilschramm.org/statstheory/stats300b-winter26.html.
- Teaching Assistant: Yash Nair
- Lectures: Mondays & Wednesdays 10:30-11:50 in 540-108.
- Office Hours: listed on Canvas.

The best way to contact me is by email. Please be sure to include "STATS 300B" in the subject line.

Prerequisites: Students will need a background in real analysis (MATH 171), probability theory (STATS 310A), and a solid command of linear algebra. Some finite-sample hypothesis testing theory (e.g. STATS 300A) is helpful, but not essential. If you have not taken the courses listed but believe that you have sufficient background to take the course, please contact me.

2 Overview

This course covers core topics in asymptotic statistics and high dimensional statistics. Our goal will be to understand (through a hard-line mathematical lens) properties of the outputs of learning algorithms (a.k.a. estimators) computed from a large quantity of (usually i.i.d.) data. We will explore the following questions: How do estimators behave (when we have access to ample i.i.d. data)? How can we quantify their success? What makes some estimators better than others? En route to answering these questions, we will introduce a variety of useful tools in concentration of measure and analysis.

Topics The course is roughly partitioned into the following units:

1. **Asymptotic statistics**. Asymptotic statistics is the study of the behavior of statistics the large-sample limit. Though it is usually arduous to understand the precise law of a statistic in n independent samples, in the large sample limit as n tends to infinity, the law of many natural statistics simplifies drastically, in that it is **well approximated** by the Normal distribution (or other distributions).
    - **Limiting behavior of statistics and estimators** We begin with some technical tools that form the basis for asymptotic statistics: notions of stochastic convergence, the law of large numbers, the central limit theorem, and the delta method. We then study the limiting behavior of maximum likelihood estimators (MLE) and other M-estimators (solutions to optimization problems) on i.i.d. data. The punchline is that under some "niceness" conditions, the law of such estimators is normal in the limit as the number of samples goes to infinity.
    - **Testing, estimation, and optimality**. Understanding the asymptotic law of functions of our data allows us to derive asymptotic confidence intervals for estimators, and to understand the

power and level of hypothesis tests. We will discuss the efficiency of estimators and what makes some estimators preferable to others. Topics include Fisher information, the Crámer-Rao bound, and the local asymptotic minimax theorem.

2. **High-Dimensional statistics and Learning Theory** We next go beyond the asymptotic setting, building a theory that incorporates the dependence of rates on dimension and complexity of loss functions. The emphasis shifts from determining the precise law of limiting statistics, to instead obtaining reasonable non-asymptotic confidence intervals.
    - **Uniform convergence and model complexity**. Topics include concentration inequalities, matrix concentration, uniform convergence, Rademacher complexity, VC dimension, Gaussian processes, chaining. The mathematical tools introduced here apply equally well in the more classical asymptotic theory.
    - **Computational Efficiency**. In many cases, the MLE is not computable by polynomial-time algorithms (unless P = NP). We will discuss a few of the algorithms/heuristics commonly employed to overcome this, and analyze (using our STATS 300B toolkit) the performance of these heuristics in some contexts where they provably succeed.

The course schedule may be found on the website.

3 Materials & Resources

Course website. The course website is www.tselilschramm.org/statstheory/stats300b-winter26.html. There you will find the course schedule, a list of texts and resources, and additional relevant readings.

Scribe notes on Overleaf. Each lecture will be scribed by a student, and the scribe notes will be available for all to edit on Overleaf. You will be added to the Overleaf project in the first week of the quarter.

Scribe notes from the 2024 and 2025 editions of the course, available on the course website, may also be helpful.

Canvas. Problem sets, problem set solution keys, Tselil's handwritten lecture notes, and any other files will be made available on Canvas. The scheduling interface for TA problem set appointments will appear on Canvas. Problem set grades will also be posted on Canvas.

Gradescope. Exams will be graded on Gradescope. You will receive a Gradescope invitation in the first week of the quarter.

Ed. We will use Ed as an online class forum, where you may ask questions and discuss with your fellow students. You can access our Ed forum through Canvas.

4 Coursework & Evaluation

Problem Sets (20%) You will be assigned a problem set every week. Collaboration is encouraged, but you are responsible for your own understanding.

- **Feedback:** We'll release a solution key a couple of days after we release the problem set. You should then verify the correctness of your own solutions using the key plus the AI tool of your choice. Note that Gemini is free for all Stanford students.
- **Assessment:** You will make a 10-minute appointment with our course TA, who will choose a problem uniformly at random from the problem set and then have a discussion with about the problem. You will be given a pass/fail grade, with a pass given if you demonstrate that you understand the solution. Each problem set will be weighted equally towards your problem set grade, except that we will drop up to two failing problem set scores.
- **Late work:** We will not allow make-ups as a matter of course policy. You may miss up to two weeks with no penalty, as we are dropping two failing scores.

Scribe Notes (5%) You will be responsible for the scribe notes of one lecture. Students will sign up (possibly in pairs) to scribe each lecture. Students have 48 hours to type up and post the scribe notes. A set of scribe notes from last year already exists; it is okay, and even encouraged, to use last years' notes as a starting point, and fulfill your assignment by updating/improving last years' notes rather than starting from scratch.

The lecture notes will be available to edit by all on Overleaf; you're encouraged to add to and improve the notes for any week, regardless of whether it is the week you are responsible for. Overleaf allows us to see who made which edits; your contributions across all lectures of the class will be counted.

In a lecture, the instructor usually only writes a subset of what they are communicating on the board. Good scribe notes are not just a latex transcript of what the instructor wrote; they fill in the gaps. For this reason, producing scribe notes requires a thorough understanding of the material, which often entails some additional reading of the supplementary texts.

Midterm Exams (20% each) There will be two in-person midterm exams, one at the end of week 4 and another at the end of week 7.

Final Exam (35%) The final will be in-person during finals week. Students must be present in-person for the final.

Ed participation (up to 5% bonus) Students may receive up to 5% extra credit for meaningful participation on Ed. Meaningful participation constitutes the asking or answering of questions in a way that has a positive impact on the learning of other students in the class.

5 Policies

The Honor Code. It is expected that you and I will follow Stanford's Honor Code in all matters relating to this course. You are encouraged to meet and exchange ideas with your classmates while studying and working on problem sets, but you are individually responsible for your own work and for understanding the material. You are not permitted to copy or otherwise reference another student's homework or computer code.

Late Work Policy. Late work will not be accepted. Instead, your lowest two problem set scores will be dropped.

Accommodations. I am happy to provide accommodations, understanding that they may be necessary for student success. Students who may need an academic accommodation based on the impact of a disability must initiate the request with the Office of Accessible Education (OAE). Students should contact the OAE as soon as possible since timely notice is needed to coordinate accommodations.

Course Privacy Statement. As noted in the University's policy on recording and broadcasting courses, students may not audio or video record class meetings without permission from the instructor (and guest speakers, when applicable).