

Stat 41521: Topics in Distribution-free Inference — Autumn 2023 Syllabus

Course description This course will focus on the recent field of distribution-free inference, which seeks to provide verifiable statistical guarantees without assumptions on the distribution of the data. Methods in this area include holdout set methods, cross-validation type methods, and conformal prediction. The course will cover theoretical advances and practical methodologies, theoretical hardness results, and open problems in the field.

Course info

- Instructor: Rina Barber, rina@uchicago.edu
- Class times: Tue/Thu 9:30–10:50, Jones 303
- Office hours: Fri 12–1, Jones 214 (starting Week 2 – no OH during Week 1)
- The main course information and materials will be on Ed Discussion. We will also use Canvas to share files if needed.
- There is no textbook, but we will provide resources and references to go along with each topic covered in the class. These resources will include papers, tutorials, videos of talks, etc, and will be posted on Ed Discussion.

Topics The topics covered in the course will depend on student interest. A tentative list:

- The distribution-free inference framework: goals, definitions, targets of inference
- Data exchangeability
- Distribution-free predictive inference methods: holdout set methods for prediction; conformal prediction; jack-knife+ and CV+
- Optimality results and hardness results for distribution-free predictive inference
- Weighted exchangeability; weighted versions of distribution-free inference methods, with applications to various problems in statistics
- Distribution-free prediction in different settings: streaming data ; relaxing exchangeability ; localized methods
- Computational cost and computational tricks for distribution-free prediction
- Distribution-free inference beyond the prediction problem: alternative measures of predictive risk; inference for regression; calibration

Grading Students may choose to take the class Pass/Fail or for a quality grade. All students registered for the course are expected to contribute to one group presentation (details below). In addition, students taking the course for a quality grade are expected to submit work for a substantial portion of the homework exercises, and are encouraged to participate in online discussion on Ed (details below).

Homework exercises Homework exercises will be posted each week on Ed Discussion. These exercises are intended to be challenging (some will be at the level of current research problems in the field). It is not necessary to submit answers to every problem, but most problems should be attempted, and students are encouraged to submit partial solutions or ideas for problems even if the answer is not complete. The problems will mostly be open-ended, so there may be multiple possible solutions. After solutions are due, students are encouraged to then share their solutions in a public post on Ed and to engage in discussion.

Presentations of empirical work Students will work in groups of size 2–3 to present an empirical exploration of a topic covered in the class. The goal for a presentation is to explore some interesting empirical aspects of methods covered in class (e.g., what types of data make a method succeed or fail, or, what types of empirical behavior of the method are or are not explained by the theory?). The presentation itself should be ≈ 15 minutes, but the group should plan to spend substantial time developing ideas, running simulations, etc, to prepare.