

USC Marshall School of Business
DSO 699: Special Topics in Data Sciences and Operations
Fall 2023 – “Modern Statistical Inference” (3 Units)

Professor: Adel Javanmard

Meeting Time: Fridays 2pm - 5pm, 3 Units

Office: Bridge Hall 300 A

Email: ajavanma@usc.edu

Office Hours: *By appointment only*

Course Description

This is a Ph.D. level lecture course exposing students to modern ideas in statistical theory. While the traditional theory assumes that one has access to many observations (large sample size) and the number of variables (features) is fixed, in the modern big data era we are able to collect fine grained information on each individual that allows us to fit feature-rich models. Our emphasis will be on statistical inference for such high-dimensional settings where there may be many or significantly more variables than observations. We will cover ideas to gauge reliability of statistical methods and the reproducibility of findings. As an example suppose that you develop a machine learning system to make personalized predictions (such as risk score for a patient). Consider the following questions:

- How certain should we be about predictions made by your algorithm?
- How certain should we be about the discovered associations between the variables and the response? Are they statistically significant?
- How fair is your prediction with respect to minority subgroups in the data?

This class will serve as a rapid introduction to current topics in statistical learning with a focus on theory and methodology. The course will cover:

- Testing problems in high dimensions: Bonferroni’s method, Fisher’s test, higher criticism
- Multiple testing problems: family-wise error rate (FWER), procedures for controlling FWER, false discovery rate (FDR), procedures for controlling FDR, online control of FDR
- Conformal prediction, conformalized quantile regression and its applications
- Conditional randomization test
- Gaussian Comparison inequalities: Slepian’s inequality, Gaussian interpolation, Gordon’s theorem
- Applications of Gaussian inequalities for analyzing statistical behavior of M-estimators

Learning Objectives

Upon successful completion of this course, students will be able to:

- Speak comfortably about statistical significance, p-values, confidence intervals.
- Evaluate statistical inference for high-dimensional settings, where there may be many or significantly more variables than observations, to improve the reliability of statistical methods and the reproducibility of findings, including any discovered associations between the variables and the response.
- Describe and analyze various statistical tests and their applications: Bonferroni’s global test, Fisher’s test, chi-square test, higher criticism

- Identify multiple testing problems: family-wise error rate (FWER), procedures for controlling FWER, false discovery rate (FDR), procedures for controlling FDR, online control of FDR.
- Explain conformal prediction, conformalized quantile regression and its applications
- Perform conditional randomization test to properly account for confounding factors
- Prove Gaussian Comparison inequalities: Slepian's inequality, Gaussian Interpolation and Gordon's theorem
- Apply Gaussian inequalities to derive precise asymptotic characterization of the statistical behavior of M-estimators.

Required Materials

We will not follow a text book, but as a PhD class, you are highly encouraged to consult outside sources to supplement your learning. The following references may be useful for background readings:

1. Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction by B. Efron, IMS Monographs.
2. Testing Statistical Hypotheses, Third edition, E. L. Lehmann, Joseph P. Romano, Springer Science & Business Media
3. Vershynin, Roman. High-dimensional probability: An introduction with applications in data science. Vol. 47. Cambridge university press, 2018.

I will post a list of papers for the final projects on the blackboard.

Prerequisites and/or Recommended Preparation:

The goal of this course is to provide students with a rigorous theoretical background to help them with statistical inference tasks in their research or pursue their own research in these areas. This course is **not** meant as a first course in Statistics or Probability theory – students are expected to have some familiarity with concepts like random variables, expectation, and conditional probability, hypothesis testing, p-values, false positives/negatives, confidence intervals.

Students are expected to have graduate working knowledge on statistics and probability concepts. Most importantly, the course language is “**proof language**”. Students are expected to have **ability to write mathematical proof and do numerical experiments with programming in Matlab, Python or R.**

Course Notes:

The majority of the course will be lecture based. A rough list of topics and outline of material is at the end of the syllabus, however, depending on the speed of the class and discussion, these topics are open to change. The precise dates of homework and exams, however, will **not** change without substantive notice.

Grading Policies:

Your final course grade will be assessed as follows:

<u>Assignment</u>	<u>% of Total Grade</u>
<u>Participation/Discussion</u>	<u>5%</u>
<u>Homework</u>	<u>30%</u>
<u>Midterm Exam</u>	<u>25%</u>
<u>Final Project/Presentation/Report</u>	<u>40%</u>
<u>Total</u>	<u>100%</u>

Homework Assignments

There will be three homework assignments (each 10%). Assignments should be written in Latex with a copy of the code included (if applicable). Late homeworks are not acceptable.

Final Projects

In the second half of the course, I will release a pool of references (recent papers on topics we cover in the course). Students should select one of these references (each reference can be chosen by only one student). Students are required to make a presentation to discuss the paper, and also submit a **5 page latex document** summarizing the paper. The reports should in particular discuss the main idea of at least one of the theorems in the paper. The final projects are evaluated based on the quality of slides, report and the presentation.

Midterm Exam:

The midterm exam will be in-class based on the materials covered in the first half of the course. There will be no-make up for the midterm exam.

Policy on Group Work

Discussion of homework assignments is permitted and encouraged; however, each student is required to prepare and submit his/her own solutions, including program/computer output, individually and separately. Duplication of homework solutions and codes prepared in whole or in part by someone else is not acceptable and is considered plagiarism. Collaboration of any sort on exam is prohibited and will result in a zero for the exam. Any suspicion of cheating will be reported and investigated by USC. Please see the "Academic Integrity and Conduct" section below for further details.

Class Attendance & Participation

Students are expected to attend all lectures and more importantly actively contribute to lectures and discussions. Your comments and feedbacks on the course materials, and the pace of the course, or my teaching style are very welcome especially that this is the first time the course is offered.

If you are not comfortable bringing up your questions in the class, take advantage of the many opportunities to speak with me one-on-one. I am always accessible by e-mail, and will be more than happy to speak with you before or after class. Note: if your question requires a conversation rather than a short answer, email is not the best way to go – please talk to me after class or schedule an appointment with me to talk in office hours.

ADDITIONAL INFORMATION

Academic Conduct:

Plagiarism – presenting someone else’s ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, “Behavior Violating University Standards” policy.usc.edu/scampus-part-b. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, policy.usc.edu/scientific-misconduct.

Students will be referred to the Office of Student Judicial Affairs and Community Standards for further review, should there be any suspicion of academic dishonesty. The review process can be found at: <https://sjacs.usc.edu/>. Failure to adhere to the academic conduct standards set forth by these guidelines and our programs will not be tolerated by the USC Marshall community and can lead to dismissal.

Discrimination, sexual assault, and harassment are not tolerated by the university. You are encouraged to report any incidents to the *Office of Equity and Diversity* or to the *Department of Public Safety*. This is important for the safety of the whole USC community. Another member of the university community – such as a friend, classmate, advisor, or faculty member – can help initiate the report or can initiate the report on behalf of another person. *Relationship and Sexual Violence Prevention and Services (RSVP)* provides 24/7 confidential support.

Support Systems:

Technical Support

Blackboard, Zoom, Slack, Google Drive, Office 365, Respondus, etc.
<https://keep-teaching.usc.edu/students/student-toolkit/>
USC IT: 213-740-5555
Marshall IT: 213-740-3000 or helpdesk@marshall.usc.edu

Counseling and Mental Health - (213) 740-9355– 24/7 on call

<https://studenthealth.usc.edu/counseling/>

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

National Suicide Prevention Lifeline - 1 (800) 273-8255 – 24/7 on call

suicidepreventionlifeline.org

Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-9355(WELL), press “0” after hours – 24/7 on call

<https://studenthealth.usc.edu/sexual-assault/>

Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

Office of Equity and Diversity (OED)- (213) 740-5086 | Title IX – (213) 821-8298

equity.usc.edu, titleix.usc.edu

Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants. The university prohibits discrimination or harassment based on the

following *protected characteristics*: race, color, national origin, ancestry, religion, sex, gender, gender identity, gender expression, sexual orientation, age, physical disability, medical condition, mental disability, marital status, pregnancy, veteran status, genetic information, and any other characteristic which may be specified in applicable laws and governmental regulations. The university also prohibits sexual assault, non-consensual sexual contact, sexual misconduct, intimate partner violence, stalking, malicious dissuasion, retaliation, and violation of interim measures.

Reporting Incidents of Bias or Harassment - (213) 740-5086 or (213) 821-8298

https://usc-advocate.symplicity.com/care_report/

Avenue to report incidents of bias, hate crimes, and microaggressions to the Office of Equity and Diversity | Title IX for appropriate investigation, supportive measures, and response.

The Office of Disability Services and Programs - (213) 740-0776

dsp.usc.edu

Support and accommodations for students with disabilities. Services include assistance in providing readers/notetakers/interpreters, special accommodations for test taking needs, assistance with architectural barriers, assistive technology, and support for individual needs.

USC Campus Support and Intervention - (213) 821-4710

<https://ucsa.usc.edu/>

Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity at USC - (213) 740-2101

diversity.usc.edu

Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call

dps.usc.edu, emergency.usc.edu

Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-120 – 24/7 on call

dps.usc.edu

Non-emergency assistance or information.

COURSE CALENDAR

This calendar is tentative and its details are subject to change depending on the pace of the class.

Week	Topics	Deliverables/ Due date
1	Course Motivation/ overview of statistical inference, hypothesis testing and confidence intervals	Aug 25
2	Global testing, Bonferroni's global test, Fisher's combination test, chi-square test, optimality Properties	Sep 1
3	Simes test, Kolmogorov-Smirnov test, higher criticism statistics, Sparse mixtures	Sep 8
4	Multiple testing problem, FWER, methods for FWER control	Sep 15
5	False discovery rate (FDR), Benjamin-Hochberg procedure, FDR control under dependence	Sep 22
6	Online control of False discovery rate	HW1 is due Sep 29
7	Midterm Exam	Oct 6
8	Fall Recess (no class)	Oct 13
9	Conformal Prediction, Conformalized Quantile Regression	Oct 20
10	Fairness via equalized coverage, Conformal prediction under covariate shift	Oct 27
11	Model-X methodology: Conditional Randomization test, Holdout randomization Test	HW2 is due (Nov 3)
12	Normality of MLE and debiasing procedures	Nov 10
13	Gaussian Comparison inequalities (Slepian's and Gordon's inequalities)	Nov 17
14	Thanksgiving Holiday (no class)	Nov 24
15	Final Project Presentations	HW 3 is due Dec 1

Project reports are due on Dec 8th, 2023