

SDS 387: Linear Models
Department of Statistics and Data Sciences
University of Texas at Austin
Fall 2024

Instructor: Alessandro (Ale) Rinaldo, Professor of Statistics and Data Sciences
E-mail: alessandro.rinaldo@austin.utexas.edu
Office: WEL 5.216G
Office: hours by appointment

TA: Khai Nguyen
E-mail: khainb@utexas.edu
Office: hours TBA

Lectures: Tuesday and Thursday, 9:00am - 10:30am, PMA 5.112

Webpage: ???

COURSE DESCRIPTION

SDS 387 is an intermediate graduate course in theoretical statistics for PhD students, covering two separate but interrelated topics: (i) stochastic convergence and (ii) linear regression modeling. The material and style of the course will skew towards the mathematical and theoretical foundations of common models and methods, designed to provide a foundation for those who wish to pursue research in statistical methods and theory. This is not an applied regression analysis course.

COURSE OBJECTIVES

The objectives of the course are twofold. Successful completion of the course should give students a firm basis from which to navigate literature on a variety of modern statistical methods and theory.

- to describe various modes of stochastic convergence, illustrate their difference, exemplify their uses in various problems and provide tool-box of basic techniques for carrying out an asymptotic analysis of statistical and probabilistic models.
- to provide a foundational understanding of linear modeling, including linear algebra background, the theoretical underpinnings of well-known estimators, tests, model diagnostics, methods for variable selection and dimension reduction and minimax lower bounds. Linear models will be introduced and studied mainly under an “assumption-lean” framework in which minimal assumptions are imposed on the data generating distribution – notably, the regression function will not be assumed to be linear and the model may very well be misspecified. The study of linear models will be the main component of the course.

PREREQUISITES and TARGET AUDIENCE

As this course is a required core course for the Ph.D. in Statistics and Data Sciences, the target audience is graduate students pursuing a Ph.D. in statistics or another closely related field. While there are no formal course prerequisites, students are expected to have graduate-level knowledge of probability, mathematical statistics, and linear algebra.

OUTCOME MEASURES

Evaluation and the course grade will be based on three components:

Homework	70%
Final Project	20%
Participation	10%

COURSE MATERIALS

The lectures will be based on material extracted from various sources.

Stochastic convergence

- T.S. Ferguson (1996). A Course in Large Sample Theory, Chapman.
- A. W. van der Vaart (1998). Asymptotic Statistics, Cambridge University Press
- The notes used in my old [Advanced Probability Overview](#) course, last taught at CMU in 2020.
- E. L. Lehmann and J. P. Romano (2002). Testing Statistical Hypotheses, 4th edition, Springer.

Other recommended references:

- D. Hunter, Notes for a graduate-level course in asymptotics for statisticians, available [here](#)
- R. Serfling (1980). Approximation Theorems of Mathematical Statistics, John Wiley, New York.
- J. Shao(2003). Mathematical Statistics, 2nd edition, Springer.
- A. DasGupta (2008). Asymptotic Theory of Statistics and Probability, Springer.

Linear modeling and linear algebra

- R. Christensen (2020). Plane Answers to Complex Questions: The Theory of Linear Models, 5th Edition, Springer.
- G. A. F. Seber and A. J. Lee (2003). Linear Regression Analysis, Second Edition, Wiley.
- S. Weisberg (2013) Applied Linear Regression, 4th edition, Wiley.
- W. Greene (2017)., Econometric Analysis, 8th edition, Pearson.

- R. Christensen (2019). Advanced Linear Modeling: Statistical Learning and Dependent Data, 3rd edition, Springer.

The following references on minimax theory are also recommended:

- A. Tsybajov, (2009). Introduction to Nonparametric Estimation. Springer.
- M. Wainwright (2019). High-Dimensional Statistics: A Non-Asymptotic Viewpoint, Cambridge University Press.
- Y. Polyanskiy and Y. Wu () Information Theory: From Coding to Learning, Cambridge University Press, forthcoming available [here](#).

CLASS STRUCTURE AND POLICIES

Lectures:. Lectures for this course will be held in person but, if/when needed broadcast via zoom. In the event that lectures are held via zoom, students are expected to attend lectures synchronously, but special arrangements can be made in coordination with the instructor if necessary.

Final Project: The final project involves picking a topic of interest, reading the relevant results in the area and then writing a short review paper (8 pages max) summarizing the key ideas in the area. You may focus on a single paper if you prefer. You are NOT expected to present your own, novel research, but you are welcome to. The paper should include background, statement of important results, and brief proof outlines for the results. If appropriate, you should also include numerical experiments or an application with real data.

Information and guidelines.

- You may work by yourself or in teams of two.
- The goals are (i) to summarize key results in literature on a particular topic and (ii) present a summary of the theoretical analysis (results and proof sketch) of the methods (iii) implement some of the main methods. You may develop new theory if you like but it is not required.
- You will provide: (i) a proposal, (ii) a progress report and (iii) and final report.
- The reports should be well-written.

Timeline.

- Proposal. Due October ???. A one page proposal. It should contain the following information: (1) project title, (2) team members, (3) precise description of the problem you are studying, (4) anticipated scope of the project, and (5) reading list. (Papers you will need to read).

- Progress Report. Due November ???. Three pages. Include: (i) a high quality introduction, (ii) what have you done so far, (iii) what remains to be done and (iv) a clear description of the division of work among teammates, if applicable.
- Final Report. Due December ???. The paper should be in NeurIPS format¹. (pdf only). No appendix is allowed. You should submit a pdf file electronically. It should have the following format:
 1. Introduction. Motivation and a quick summary of the area.
 2. Notation and Assumptions.
 3. Key Results. Proof outlines for the results.
 4. Implementation (simulations or real data example.)
 5. Conclusion and/or future work/open questions (if appropriate).

Problem Sets and Homework: Biweekly homework assignments will be a combination of textbook-style exercises and problems taken from the literature. Students are encouraged to work collaboratively with each other on homework problems, but should make a full individual effort before consulting or comparing with peers. **There will be approximately 6 homework assignments.**

Academic Integrity for Homework: You are responsible for ensuring that any answers you submit for evaluation are the result of your own efforts. Final solution writeups and any requested computer code **MUST** be a result of your own work, and no material may be copied from another student. If you work with other students on the homework problem sets, **you must turn in your own completed assignment** and please be sure to include with your assignment the names of your collaborators.

Sharing of Course Materials is Prohibited: No materials used in this class, including, but not limited to, lecture hand-outs, videos, assessments (quizzes, exams, papers, projects, homework assignments), in-class materials, review sheets, and additional problem sets, may be shared online or with anyone outside of the class unless you have my explicit, written permission. Unauthorized sharing of materials promotes cheating. It is a violation of the University's Student Honor Code and an act of academic dishonesty. I am well aware of the sites used for sharing materials, and any materials found online that are associated with you, or any suspected unauthorized sharing of materials, will be reported to Student Conduct and Academic Integrity in the Office of the Dean of Students. These reports can result in sanctions, including failure in the course.

In particular, use of any assignments provided through previous offerings of the course (e.g., previous semesters' homework or exams) or communication about details of these materials with students who have taken the course previously will be regraded as academic dishonesty in this course.

Class Recordings: In the event that course lectures are held via zoom and recorded, class recordings are reserved only for students in this class for educational purposes and are protected under FERPA. The recordings should not be shared outside the class in any form. Violation of this restriction by a student could lead to Student Misconduct proceedings.

Accessible and Inclusive Learning Environment: The university is committed to creating an accessible and inclusive learning environment consistent with university policy and federal and state law. Please let the instructor know if you experience any barriers to learning so we can work with you to ensure you have equal opportunity to participate fully in this course. If you are a student with a disability, or think you may have a disability, and need accommodations please contact Disability and Access (D&A). Please refer to D&As website for contact and more information: <http://diversity.utexas.edu/disability/>. If you are already registered with D&A , please deliver your Accommodation Letter to me as early as possible in the semester so we can discuss your approved accommodations and needs in this course.