USC Marshall School of Business

DSO 699: Special Topics in Data Sciences and Operations

Spring 2021 – Foundations of Stochastic Modeling for Prescriptive Analytics

Professor: Vishal Gupta
Meeting Time: T/Th 10 am – 11:30 am
Zoom Meeting Link: Prepopulated on Blackboard site
Office: See Zoom and Slack
Email: None! Use Slack only.
Office Hours: *By appointment only*

Course Description

This is a Ph.D. level lecture course covering the core probability theory and stochastic analysis necessary for modeling and analysis in data-driven optimization, a.k.a., prescriptive analytics. As examples, consider the following questions:

- Suppose we fit a parametric demand model to some transaction data, and then optimize revenue to find a price. How suboptimal do we expect our price to be? Would doubling the amount of data substantively help?
- What if we fit a non-parametric demand model to our data instead? (How do we even do this?!?) When will this be better or worse than the parametric model?
- Suppose we have access to large collection of patient health data. Can we use this data to design personalized treatment plans for patients with specific diseases? How? How much data do we need before we are sure our plans aren't unintentionally hurting our patients?
- Suppose we design a custom, blended ``estimation and optimization" methodology for prescriptive analytics. Can we say anything about the statistical properties of our approach? Is it actually better than state of practice?

The goal of this course is to provide students with a rigorous theoretical background to enable them to answer questions like these and pursue their own research in these areas. This course is **not** meant as a first course in probability theory – students are expected to have some familiarity with concepts like random variables, expectation, and conditional probability. Similarly, it is **not** a formal treatment of measure-theoretic probability. Rather, the focus is on the probability tools most commonly used to analyze algorithms in data-driven optimization, machine learning, and personalization. In this sense, the course is also application focused with special attention to the common problems in these fields.

Learning Objectives

By the end of the course, student should be able to

- Speak comfortably about asymptotics of random variables with standard big "Oh-p" notation, and formalize such statements using relevant notions of convergence of random variables.
- Use standard tail bounds and concentration results to analyze randomized algorithms and data-driven methods. In particular, they should be able to analyze "stable" data-driven optimization algorithms.
- Prove fundamental results in causal inference and policy learning.
- Prove fundamental results about sample average approximation (empirical risk minimization) including generalization guarantees and uniform laws of large numbers. In particular, students will analyze algorithms using metric entropy and VC-dimension.
- Analyze algorithms for contextual stochastic optimization based on data-driven estimates of conditional expectation

Required Materials

There is no required textbook for the course. Lecture notes and recordings will be distributed through Blackboard/Slack.

That said, as a PhD Class, you are highly encouraged to consult outside sources to supplement your learning as necessary. Some works I personally recommend:

- Measure Theoretic foundations of Probability and Conditional Expectation
  - *Probability with Martingales* by David Williams. This book is beautiful. If you want a good, formal, elegant treatment of measure theoretic probability (and especially conditional expectation), this is the book to read. Most of this is beyond the scope of our course.
- Asymptotic of Random Variables
  - *Asymptotic Statistics* by Van der Vaart (Esp Chapt. 1-2.5). This is a classic book and a good reference on these basics.
- Tail Bounds and Concentration of Measure
  - *High-Dimensional Statistics: A Non-Asymptotic Viewpoint* by Martin Wainright. This treatment in this book is my favorite. I highly recommend it because it is fairly intuitive.
  - *Concentration Inequalities: A Nonasymptotic Theory of Independence* by Boucheron, Lugosi and Massart. (Esp. Chapt 1-2) This book is a bit more technical/terse than the Wainright book above, but has also some other techniques results that are worth learning for more advanced students. It also serves as an excellent reference.
- Empirical Process Theory
  - Pollard's Iowa Notes (Esp. Chapt. 1-7) - Available here: http://www.stat.yale.edu/~pollard/Books/Iowa/Iowa-notes.pdf . My treatment will largely follow this presentation with some deviations around the development of pseudo-dimension.
  - *Asymptotic Statistics (listed above)* (Chapt. 19): This is very terse, but covers the basics.
  - *Concentration Inequalities (listed above)*. This book gives a more "classical probability" treatment of empirical processes. Consequently, it covers ``more cases" than the Pollard treatment but is also more technical and a little less "user friendly." I recommend it as a reference or for more advanced students, but not as a first read.
  - Lecture Notes: There are also TONS of lecture notes online on this topic. Here are some that I like and use: https://www.shivani-agarwal.net/Teaching/E0370/Aug-2011/ and https://ambujtewari.github.io/teaching/LearningTheory-Spring2008/

Office Hours/Contacting Me:

I will not be using email for this course. We will attempt to use Slack to replace email. Hence, I will NOT respond to any emails sent to me about course materials. Please instead send me a Slack Message. Slack is available (for free) to all USC students and you should automatically be enrolled in the class channel. If you aren't, send me a slack message and I will add you. Familiarize yourself with slack and how to use it on IT's website: https://cio.usc.edu/digital-campus-slack/

Occasionally, you may need to contact me about a private matter. In that case, please use the direct message feature. If upon reading your message, I deem it should be public, instead, I might ask you to resend to the public class slack channel so that all students can benefit from the question.

I will do my best to respond promptly to slack messages. Some common emoji's we will use in this course are listed on blackboard and our first pinned Slack message.

1-on-1 office hours are available at any time by appointment. Slack Me and we'll schedule a time.

Please keep in mind that Slack is as much part of the academic environment of this course as is class-time. Hence, please keep language professional (but fun!). You know how to be a good citizen. Just do what you know.

Prerequisites and/or Recommended Preparation:
There are no formal prerequisites for the course but students should be familiar with basics of probability theory (random variables, expectation, probability density functions, moment generating functions, combinatorial probability). Measure theoretic probability is **not** required. Any students concerned about their background ability should reach out to the instructor to discuss their particular situation.

Course Notes:
All meetings for this course will be online over ZOOM. Please log-into class using your USC Zoom account. Note, if you're a Marshall student or have a personal zoom account, you may have multiple log-ins. If you use an account other than your USC zoom account, you will be placed in a "Waiting Room" upon entry. To avoid this, please use your USC zoom account. I will only allow participants in from the waiting room BEFORE class starts. So if you're late or if your internet drops-out, you won't be able to re-join. It's better in general to just use your USC account.

As an experiment, the course will be using its Slack channel extensively to distribute materials, lecture slides, make announcements. Part of the rationale is I would like to encourage discussion among students to coordinate working on homework together, sharing materials, and in general building their academic community. All that said, if Slack turns out to be too difficult to use, we will transition back to Blackboard.

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:

There will be several graded deliverables for the class:

- Homework: Assignments will be approximately every two weeks. These are meant to be challenging and proof-based. You are welcome to work in groups on homework (see below for group-work policy). Some homework may involve coding small simulations. **You cannot receive an A for the course, however, unless you turn in every homework.**

- Midterm: There will be one take-home midterm (24 hours). You must complete this alone, i.e., not in groups.

- Final Exam: There will be a one take-home final exam (24 hours). Again, you must complete this alone, i.e., not in groups.

In addition, participation in discussion and "Zoom-Ettiquette" and Slack participation will contribute to your final grade.

Policy on Group Work
Group discussion is STRONGLY encouraged throughout this class with other students in the class. Throughout your PhD, your peers will always be your best resource. Use them. You may collaborate with other students on ANY of the above deliverables.

However, you MUST always write up your own assignments individually and separately. (Thus, you can talk about a paper together, or even get a peer to read through your report and give you feedback, but you must incorporate that feedback on your own.) Please also list the names of students you collaborated with on the deliverable under your name, with a brief description of their contribution (if you deem it necessary).

For example, on my homework project, I might write:

Collaborated with: John Snow (Problem 1 and 2), Sansa Stark (Problem 3), Tyrion Lannister (entire assignment)

Grading Breakdown:

Assignment	% of Total Grade
Participation/Discussion	10%
Homework	30%
Midterm	30%
Final Exam	30%
Total	100%

Synchronous session recording notice

All sessions of the course will be recorded and provided to all students enrolled in the course (and officially auditing non-USC students) via Slack (or, failing slack, BB). Consequently, it is also important that students respect USC's policy and do NOT share any of the course content outside the course. This includes recordings, lecture notes, or other materials. For clarity, from SCampus:

SCampus Section 11.12(B)

Distribution or use of notes or recordings based on university classes or lectures without the express permission of the instructor for purposes other than individual or group study is a violation of the USC Student Conduct Code. This includes, but is not limited to, providing materials for distribution by services publishing class notes. This restriction on unauthorized use also applies to all information, which had been distributed to students or in any way had been displayed for use in relationship to the class, whether obtained in class, via email, on the Internet or via any other media. (SeeSection C.1 Class Notes Policy).

Zoom Etiquette Expectations

Online-learning brings some additional challenges outside the usual environment. I expect students to bring the same curiosity, engagement, and professionalism that they would normally in an in-person class. To that end, students are required

- To ensure they are in a quiet, private place to attend class. They should be able, if called upon, to unmute and participate in discussion.
- To ensure they have adequate technology and internet access to attend class without disturbance. If you have particular technology concerns, please reach out to me ASAP to discuss.
- To be familiar with Zoom: How to share screen, present, mute/unmute audio, raise your hand, ask for a coffee-break, etc.
- To keep their cameras on during class-time and mute their microphones when not speaking
- To participate in breakout rooms and discussion.

In summary, you know how to be a good student and a good citizen. Just do what you know.

ADDITIONAL INFORMATION

USC Statement on Academic Conduct and Support Systems

Explanation - This section, or an enhanced version, is required by the University. You are free to enhance the content as you deem necessary within the structure of the following.

Academic Conduct:

Students are expected to make themselves aware of and abide by the University community's standards of behavior as articulated in the Student Conduct Code. 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" https://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, http://policy.usc.edu/scientific-misconduct.

Support Systems:

It is important to recognize that distance learning is hard. The online platform is difficult, but being isolated, especially in a process as challenging as Ph.D., is also hard. Please look out for one another. If you are feeling overwhelmed, reach out. You may always reach out to me or to your classmates. In other circumstances, you might feel more comfortable reaching out to one of the resources below.

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.

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 is committed to making reasonable accommodations to assist individuals with disabilities in reaching their academic potential. If you have a disability which may impact your performance, attendance, or grades in this course and require accommodations, you must first register with the Office of Disability Services and Programs (www.usc.edu/disability). DSP provides certification for students with disabilities and helps arrange the relevant accommodations. Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to me (or to your TA) as early in the semester as possible. DSP is located in GFS (Grace Ford Salvatori Hall) 120 and is open 8:30 a.m.–5:00 p.m., Monday through Friday. The phone number for DSP is (213) 740-0776. Email: ability@usc.edu.

USC Campus Support and Intervention - (213) 821-4710
https://uscsa.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

The details of the course calendar are subject to change depending on the pace of the class.

Sessions 1-5: Measure Theoretic Foundations and Common Random Variables

- Measure-Theoretic definitions (in-brief)
- Common Random Variables (Bernouli, Rademacher, Binomial, Poisson, Normal, Multivariate Normal) and their properties
- Poisson Approximation and "generating" independence

Session 6-12: Conditional Expectation with Applications to Contextual Stochastic Optimization and Causal Inference

- Formal Definitions
- Bayesian Estimators
- (Kernel) Estimates of Conditional Expectation
- Contextual Stochastic Optimization
- Policy Evaluation and Causal Inference

Session 13 – 17: Asymptotics of Random Variables

- Modes of Convergence
- Weak and Strong Laws, CLT
- Delta method

Session 18 – 22: Concentration of Measure

- Tail Bounds
- Sums of Independent Random Variables
- Functions of Independent Random Variables
- Algorithmic Stability and other Applications

Session 23 – 28: Uniform Laws of Large Numbers via Empirical Processes

- Discretization and Packing
- Chaining and Entropy Integrals
- VC-Dimension and Metric Entropy
- Generalization Guarantees for ERM with Application to SAA

Homework

- HW 1: 2 Feb
- HW 2: 18 Feb
- HW 3: 4 March
- HW 4: 16 March
- HW 5: 6 April
- HW 6: 20 April
- HW 7: 29 April

Midterm Exam: Week of 16 March (Take Home)

Final Exam: 5 May (Take Home)