DSC 381: Probability and Simulation Based Inference for Data Science
Fall 2021

Course overview

Students will learn the foundations of probability and inferential statistics, through both basic theoretical underpinnings and simulation examples. The concepts and skills that are taught in this course will provide students with the tools to understand many different models and approaches used in statistical learning and inference. Topics include basic probability calculus, random variables, probability functions and densities, useful inequalities, sampling distributions of statistics and confidence intervals for parameters, hypothesis testing, properties of estimators, maximum likelihood estimation, and exponential families. Students will learn the basic principles of statistical inference not through formulas and memorization, but by the basic theoretical underpinnings of the process through simulation and practical examples.

Textbooks

For the probability part of the course we use the textbook
M. Mitzenmacher and E. Upfal, "Probability and Computing", Cambridge University Press (CUP), ISBN 9781108110723. The e-book sells for $56, and hardcopy for $70.
This is a required textbook. The 1st edition is also fine (of the material covered in this class, it is missing only §9). Cheaper used copies are available on-line.

For the statistics part, you are expected to use some material from an applied statistics textbook. The recommended textbook is Lock, Lock, et. al., *Statistics: Unlocking the Power of Data*, Wiley. (3rd ed. 2020, 2nd ed. 2016, 1st ed. 2012) The book is not required. Any edition is acceptable. No homework is assigned from it. Online access to the current edition, with additional helpful videos and complete solutions for all odd-number problems, can be purchased for $70 (with a grace period of 14 days before paying.) See also the handout "Prerequisite Information", linked in edX (under "Week 0", "Handouts").

Instructor
Mary Parker (she/her)

Teaching Assistants (TAs):
 Korawat Tanwisuth (he/his), Yanxin Li (she/her), and Fuheng Cui (he/his)

Contact the instructor or TAs through the course email address: DSC381@austin.utexas.edu

Communication and Piazza:
This course is quite large, with close to 500 students. We will be using Piazza heavily. Please ask questions about course content and syllabus through piazza only. (Email messages about content and syllabus questions will not be answered because those are useful to many students in the course.) The TAs and the instructor will be very active on Piazza. But we also urge you to do your best to answer each other's (and your own) questions. Trying to explain a concept you have learned is the best way to understand it more deeply. So, you will help yourselves, and also your classmates!

Office Hours:

Professor Parker: Online office hours as scheduled on course page.
TA: Office hours on zoom (times TBA) and by appointment

Prerequisites

Calculus and knowledge of basic applied statistics methods or willingness to work on these from one of the recommended elementary statistics books during the course.

Computing

Some homework problems will need the use of a statistical programming language like R. Example code will be given in R. There are many excellent tutorials available on-line, for example the (free) edX course "Data Science: R Basics", by Rafael Irizary. If you do not yet know and use R, please this or a similar tutorial to acquire a basic working knowledge. However, R is not strictly required. You could use any other programming language with good statistics support.

In the statistics part of the course, examples and solution keys are prepared using StatKey, http://www.lock5stat.com/StatKey/. which provides basic statistical inference using randomization methods.   It accompanies the recommended statistics textbook (Lock et al.) and is freely available at on the web. Instructions for its use are provided.

Grading Policy

The final grade will be a weighted average of your homework, mid-term, and final scores. The weightings are: Homework: 30%, Quizzes: 30%, Midterm 1: 20%, Midterm 2: 20%.
The following cutoffs will be used for course
grades: A: 93-100, A-: 90-92,   B+: 87-89, B: 83-86, B-: 80-82,
grades: C+: 77-79, C: 73-76, C-: 70-72,   D+: 67-69, D: 63-66, D-: 60-62,   F: 0-59

Homework: Each weekly homework assignment will count equally in the final homework grade. The central goal of the homework assignments is to ensure that students are practicing using the main concepts presented in the class. Substantial effort in practicing these in order to be able to use these concepts in later courses, so these make up almost 1/3 of the course grade. Homework problems will be posted approximately weekly (as we complete a major topic or chapter in the lecture).

Quizzes: Quizzes are short (usually 3 or 4 problems similar to shorter homework problems) so each problem counts a high percentage of the grade. They are timed assignments (30 minutes.) Be sure to understand how to work all the homework problems before taking the quiz.. Quizzes have a window of seven days in which to take them.

Midterms: Midterm exams have approximately 8 – 12 problems, which will be similar in scope and difficulty to medium-length and short homework problems. They are timed assignments (120 minutes.) You will have a window of seven days in which to take the exams.

Rules for Quizzes and Exams: All quizzes and exams are open book (any book is fine), and open-notes (your course notes, lecture notes from edX). Calculator is ok (including R or any other statistics program on a computer). However, on quizzes and exams please no on-line resources and no help from anyone else (in person or on line.)

Rules for getting help on homework.   See the FAQ page portion of the syllabus.

Discussion of availability of solutions, grading, etc.   See the FAQ page portion of the syllabus.

Calendar and Topics

- Learning material for the week: Available Wednesday of the previous week at 11:00 UTC
- Homework for the week: Available Sunday of the week. Due Wednesday of the following week at 11:00 UTC
- Quizzes and Exams: Available Thursday of the week. Due Thursday of the following week. at 11:00 UTC

(UT has a 3-day holiday in Week 14, resulting in modified due day for HW 12.)

Course Calendar and Outline of Topics
MU §x refers to chapters in the (required) textbook by Mitzenmacher and Upfal.

Before class begins: Aug. 23 – Aug. 24
Read through the Week 0 materials

Week 1: Aug 25 – 27
Events and probability (MU §1), lectures 1.1–1.11z
Start working on HW 1
Complete HW 0   (Orientation HW) (ungraded)
Complete Quiz 0 (Orientation Quiz) (ungraded)

Week 2: Aug 30 – Sept 3
Events and probability (continued), lectures 1.12-1.13;
HW 1 is assigned (due Wed. Sept 8)
Discrete random variables (MU §2), lectures 2.1-2.7

Week 3: Sept 6 – 10
Discrete random variables (continued), lectures 2.8-2.18
Thur Sept 9: Quiz 1 assigned (covers Weeks 1 and 2)
HW 2 is assigned (due Wed Sept 15)

Week 4: Sept 13 – 17
Moments & Deviations (MU §3.1-3.3) , lectures 3.1-3.9
HW3 is assigned (due Wed Sept 22)
Thur Sept 16: Quiz 1 is due.
Thur Sept 16: Quiz 2 is assigned (covers Week 3)

Week 5: Sept 20 – 24
Introduction to Statistical Inference, lectures 4.1-4.5
HW 4 is assigned (due Wed Sept 29)
Thur Sept 23: Quiz 2 is due

Week 6: Sept 27 – Oct 1
Cautions, lectures 5.1 and 5.2
Use of Simulation 6.1-6.2
HW 5 is assigned (due Wed Oct 6)
Thur, Sept 30: Midterm is assigned (covers all material through Week 5)

Week 7: Oct 4 – 8
Moment generating functions (MU §4.1-4.2.1), lectures 7.1-7.3
Continuous random variables (MU §8.1 – 8.3.1), lectures 8.1-7
HW 6 is assigned (due Wed Oct 13)
Thur Oct 7: Midterm exam is due (Covers all material through Week 5)
Thur Oct 7: Quiz 3 available (covers weeks 5 and 6)

Week 8: Oct 11 – 15
Continuous random variables (continued), lectures 8.8-15
HW 7 is assigned (due Wed Oct 20)
Thur Oct 14 Quiz 3 due

Week 9: Oct 18 – 22
Normal distribution (MU §9), lectures 9.1-7
HW 8 is assigned (due Wed Oct 27)
Thur Oct 21 Quiz 4 available (covers weeks 7 and 8)

Week 10: Oct 25 – 29
Inference with Simulation: Details, lectures 10.1-5
HW 9 is assigned (due Wed Nov 3)
Thur Oct. 28 Quiz 4 is due

Week 11: Nov 1 – 5
Inference with Theoretical Distributions, lectures 11.1-6
HW 10 is assigned (due Wed Nov 10)
Thur Nov 4: Quiz 5 available (covers Weeks 4, 5, 10, and 11)

Week 12: Nov 8 – 12
Chi-Squared and ANOVA Tests, lectures 11.7-10
HW 11 is assigned (due Wed Nov 17)
Thur Nov 11: Quiz 5 is due

Week 13: Nov 15 – 19
Theory: Estimation, lectures 12.1-6
 HW 12 is available Fri. Nov. 12 and due Tue Nov 23 (one day earlier than usual to avoid holiday)

Week 14: Nov 22 – 23 (UT holiday Nov 24 – 29)
Theory: Exponential Families, lectures 12.7-9
HW 13 is available Fri Nov. 19 (due Wed Dec 1)
Quiz 6 is assigned Tue. Nov 22 (covers Weeks 12 and 13) (due Thur Dec 2)

Week 15: Nov 29 – Dec 3
Thur Dec 2 Quiz 6 is due
Fri Dec 3: Midterm 2 is available

Week 16: Dec. 5-10
Fri Dec 10 Midterm 2 is due

Thur Dec 16 Course ends
- edX course is archived, with learning materials remaining available, but not assignments.
- Piazza Discussion Board is archived, so no new posts can be made

Program Requirements
For questions about program requirements and credit, please contact the MSDS Program Coordinator at msdsgradcoordinator@utexas.edu.

Academic Dishonesty and Policies on Cheating
Faculty are committed to detecting and punishing all instances of academic dishonesty and will pursue cases of academic dishonesty in accordance with university policy. Academic dishonesty, in all its forms, is blight on our entire academic community. All parties in our community (professors, staff, and students) are responsible for creating an environment that educates outstanding engineers, and this goal entails excellence in technical skills, self-giving citizenry, and ethical integrity. Industry wants engineers who are competent and fully trustworthy, and both qualities must be developed day by day throughout an entire lifetime. Details about what constitutes academic dishonesty can be found at the following URL: UT Dean of Students Office (http://www.utexas.edu/depts/dos/sjs/academicintegrity.html). All cheating will be reported directly to the college/university. Unless explicitly indicated in an assignment, you must do your homework, projects and exams individually. You are welcome and encouraged to discuss material with your colleagues, when and where it is appropriate, but copying, stealing papers, etc. are considered dishonest and will be prosecuted.

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 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. We are 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. Allegations of Scholastic Dishonesty will be dealt with according to the procedures outlined in Appendix C, Chapter 11, of the General Information Bulletin, http://www.utexas.edu/student/registrar /catalogs/.
The University of Texas at Austin provides, upon request, appropriate academic adjustments for qualified students with disabilities. For more information, contact the Office of the Dean of Students at 471-6259, 471-4241 TDD or the College of Engineering Director of Students with Disabilities at 471-4321.

Documented Disability Statement

Any student with a documented disability who requires academic accommodations should contact Services for Students with Disabilities (SSD) at (512) 471-6259 (voice) or 1-866-329-3986 (videophone). Faculty are not required to provide accommodations without an official accommodation letter from SSD. Please notify us as quickly as possible if the material being presented in class is not accessible (e.g., instructional videos need captioning, course packets are not readable for proper alternative text conversion, etc.). Contact Services for Students with Disabilities at 471-6259 (voice) or 1-866-329-3986 (video phone) or reference SSDs website for more disability-related information: https://diversity.utexas.edu/diversity