ENG EK381 Probability, Statistics, and Data Science for Engineers
Spring 2021, College of Engineering, Boston University
Course Information

Motivation: It is becoming increasingly clear that the next generation of engineers will need to draw upon concepts and techniques from probability and statistics to tackle the challenges posed by uncertain, complex systems and large, high-dimensional data sets. This course is intended to give all engineering students a strong foundation in probability and statistics as well as an introduction to ideas from data analytics and machine learning. Any student that successfully completes this course will be well-prepared to take upper-electives in machine learning, data analytics, random processes, as well as any other course that draws heavily upon probabilistic reasoning.

BU Hub Outcomes: Students who successfully complete this course will have satisfied:

- *Quantitative Reasoning II:* Students will learn how to translate complex engineering problems into formal probabilistic questions as well as the techniques needed to evaluate these statements analytically and algorithmically. They will gain intuition for statistical thinking, its modern application to large data sets, as well as the limitations of this approach and associated risks.
- *Critical Thinking:* Students will learn where and why "common-sense" intuition does not match up with formal reasoning through probability and statistics. They will develop the toolkit needed to assess modern engineering problems from the probabilistic lens and make and defend the validity of arguments, including their own.

Course Goals (ENG): Provide students with:

- A background in the foundations of probability theory.
- Intuition for probabilistic concepts and reasoning.
- Experience with standard probability distributions and their application to modeling engineering systems and data sets.
- An understanding of statistics and the application of statistical techniques to data sets.
- An appreciation for the applications of probability in the design and analysis of modern engineering systems and processes.
- An understanding of hypothesis testing and optimal decision rules.
- Experience with numerical software for generating and analyzing pseudorandom quantities.
- A basic understanding of modern methods in data science and machine learning.

Course Outcomes (ENG): As an outcome of completing this course, students will:

- Understand the basic foundations of probability theory.
- Develop intuition for probabilistic concepts and gain experience with principled probabilistic thinking and computation.
- Become familiar with standard probability distributions for discrete and continuous random variables.
- Acquire intuition for multiple random quantities and their relationships.
- Understand the role of probability for modeling and analyzing complex engineering systems.
- Understand the basic principles and techniques of statistics, and gain experience with their application to data sets.
- Gain experience using numerical software to simulate random events.
- Understand the principles of optimal hypothesis testing and detection.
- Be exposed to basic methods in data analytics and machine learning and their applications.
- Acquire the preparation needed to succeed in upper-level courses that rely on probability as a pre-requisite.

Course Staff:

Instructors:	Prof. Bobak Nazer Course Content and Logistics Email: `bobak@bu.edu`
	Prof. Vivek Goyal Section A1, Tuesdays and Thursdays, 3:30-5:15pm, LAW AUD Office Hours: TBD Email: `goyal@bu.edu`
	Prof. Laura Lewis Section A2, Tuesdays and Thursdays, 1:30pm-3:15pm, BCN 208 Office Hours: TBD Email: `ldlewis@bu.edu`
	Prof. Ray Nagem Section A3, Mondays and Wednesdays, 10:10-11:55am, FLR 152 Office Hours: TBD Email: `nagem@bu.edu`
Graduate Teaching Fellows:	Yue Dai (`ydai0@bu.edu`), Unay Dorken (`udorken@bu.edu`), David McIntyre (`dpmc@bu.edu`), Deja Rogers (`dejar14@bu.edu`), Yuping Wang (`yupingw@bu.edu`), Zili Wang (`zw2445@bu.edu`)
Undergraduate Teaching : Fellows:	Kylee Anders, Eugenia Angelopoulos, Jun Young Choi, Guadalupe Garcia (*Lead UTF*), Harsh Hemlani, James Maher, Justin Sadler, Kyra Scott, Laith Shamieh, Esther Ye Homework Help Sessions: Wednesdays and Thursdays, 7-9pm, Zoom

Course Websites:

Main:	`http://learn.bu.edu/`
Discussion Board:	`http://piazza.com/`
Homework Submission:	`http://gradescope.com/`

Piazza: This semester we will be using Piazza as a discussion board. The system is highly catered to getting you help quickly and efficiently from both the course staff and your fellow classmates. Unlike previous semesters, all posts must be made under your own name, which will be visible to both the instructors and to your classmates.

Prerequisites: ENG EK103 Computational Linear Algebra.
Corequisites: CAS MA225 Multivariate Calculus.

Videos: This is the second semester that we are running a "flipped" class. Specifically, prior to each lecture, you will be required to watch videos made by Profs. Nazer and Castañón that introduce the main concepts and work through some basic examples. Lectures will be devoted towards working out more complex examples, developing high-level intuition, connecting concepts to engineering applications, and tying the math to computation and real datasets. Keep in mind that, without watching the videos, you might easily get lost in lecture. Any feedback on the videos is quite welcome via Piazza, including pointing out typos.

Textbook: The required text for EK381 is a set of (free) lecture notes that are under development by Profs. Castañón and Nazer. Draft PDFs of individual chapters are available on the course website. These lecture notes contain more details and examples than we are able to cover in class directly, and we highly recommend that you read along as the course progresses. Any feedback is quite welcome via Piazza, including pointing out typos.

Grading:

3 Exams (equally weighted)	85%
Homework and Quizzes	15%
Bonus	`if` (Homework Score $\geq$ 90%) `then` (+1/3 letter grade)

Homework: Homework will be due weekly with a typical due date of Friday at 11:00pm via Gradescope. Homework may be submitted via Gradescope up to 24 hours after the deadline with a 25% grade penalty. (Note that if any problems are submitted past the deadline, the entire homework will be counted as late.) Requests for later submissions and/or extensions will not be entertained (except under exceptional circumstances, which must be discussed with your section's instructor).

- *Collaboration policy:* While you may discuss homework problems with other students for clarifying your understanding, you must *independently solve and write* your own solutions. Please get in touch if you are not sure whether the extent of your collaboration with other students is acceptable.

- *Homework Submission Guide:* See the course website for a detailed guide, including how to upload PDFs to Gradescope and assign questions to pages. (Unassigned questions will not be graded.)

- *Drop Lowest Score:* To allow for some flexibility, we will drop your lowest homework score at the end of the semester.

- *Problem-Solving in Lecture:* Since you will be required to watch videos outside of class time, some part of lecture time will be devoted to solving problems that *appear on the homework*. The corresponding problems will be noted on the homework, so that you can save them for lecture if you wish.

- *Homework-Help Zoom Sessions:* Every Wednesday and Thursday from 7-9pm, Undergraduate Teaching Fellows will be available on Zoom. We encourage you to join these sessions at `https://bostonu.zoom.us/j/99547038916?pwd=b2dsU2hBYXYvQnZIWlZlcTB1WnExQT09` to discuss your assignment with other students in the class and ask questions from the UTFs when you get stuck.

Online Quizzes: Along with each homework, you will have a short online quiz due on Wednesday at 11:59pm. The questions are chosen to walk you through the basic concepts. For each quiz, you are allowed 3 attempts and your grade will be the highest of the 3 grades. After each attempt, you will get feedback on which questions you got wrong. It is highly recommended that you take time between attempts to work out correct answers (rather than just guessing randomly). To allow for some flexibility, we will drop your lowest quiz score at the end of the semester.

Supplementary Resources: Beyond the course materials available through the website and the textbook, there are many other resources that are useful for learning probability. In a separate information sheet (titled Course Resources), you can find pointers to video lectures, additional textbooks and lecture notes as well as exercises with solutions. You are encouraged to look through these resources to find the best combination of materials that work for your learning style.

Ideal Study Habits: The material in this class can be quite challenging if you do not approach it in the right way. In a separate document (titled Ideal Study Habits), we have compiled some suggestions (from the instructor and past students) on how to organize your time and effort to do well in this class.

Core Topics. Over the course of the semester, we will cover the following *concepts*. This is how we will organize our thinking about probability throughout the semester (as opposed to formulas):

1. **Foundations of Probability**
    - Set Theory
    - Probability Axioms
    - Conditional Probability
    - Independence
    - Counting
2. **Discrete Random Variables**
    - Probability Mass Function (PMF) and Cumulative Distribution Function (CDF)
    - Average and Expectation,Variance and Standard Deviation
    - Functions of Discrete Random Variables (RVs) and their Expectations
    - Important Families of Discrete RVs
    - Conditioning a Discrete RV by an Event
3. **Continuous Random Variables**
    - Probability Density Function (PDF) and Cumulative Distribution Function (CDF)
    - Expectation of Continuous RVs, Variance and Standard Deviation
    - Important Families of Continuous RVs
    - Conditioning a Continuous RV by an Event
    - Functions of Continuous RVs
4. **Pairs of Random Variables**
    - Pairs of Discrete RVs
    - Pairs of Continuous RVs
    - Independent RVs
    - Expected Values of Functions of Pairs of RVs
    - Conditional Expectation
5. **Second-Order Analysis**
    - Covariance and Correlation
    - Jointly Gaussian RVs
    - Random Vectors
6. **Detection**
    - Binary Hypothesis Testing
    - Performance Analysis and Receiver Operating Characterisic
    - Generalizations: Bayes Cost, M-ary Hypotheses, and Vector Observations
7. **Estimation**
    - Maximum a Priori (MAP), Maximum Likelihood (ML), and Minimum Mean-Squared Error (MMSE) Estimation

- Linear Least-Squares Error (LLSE) and Vector Estimation

8. **Limit Theorems**
   - Sequences and Sums of RVs, Independent and Identically Distributed (i.i.d.) RVs
   - Basic Inequalities
   - Law of Large Numbers, Central Limit Theorem

9. **Statistics**
   - Basic Statistics
   - Confidence Intervals
   - Hypothesis Acceptance

10. **Data Science and Machine Learning**
    - Parametric and Non-Parametric Machine Learning
    - Classification
    - Dimensionality Reduction
    - Cross-Validation and Performance Evaluation
    - Pet Classification Challenge

11. **Markov Chains**
    - Finite Markov Chains
    - Steady-State Behavior
    - Infinite Markov Chains and Transient Behavior

General Policies:

- *Academic misconduct:* The student handbook defines academic misconduct as follows:

  `Academic misconduct occurs when a student intentionally misrepresents his or her academic accomplishments or hurts other students' chances of being judged fairly for their academic work.`

  This basic definition applies to EK381. If you are ever in doubt as to the legitimacy of an action, please talk to me immediately. The penalty for academic misconduct at BU is severe. For further information on the BU Academic Code of Conduct, visit the following website: `http://www.bu.edu/academics/policies/academic-conduct-code/`

- *Make-up Exams:* There will be no make-up exams. If there is a legitimate reason for missing an exam, then the scores of other exams will be used appropriately to compensate for the missed exam. If there is no legitimate reason provided for missing an exam, a grade of zero will be assigned for the missed exam.

- *Final Exam:* Students must be present for the final exam during the assigned time slot, except under exceptional circumstances that must be discussed with the instructor at the beginning of the semester.

- *Incomplete grades:* Incomplete grades will not be given to students who wish to improve their grade by taking the course in a subsequent semester. An incomplete grade may be given for medical reasons if a doctor's note is provided. The purpose of an incomplete grade is to allow a student *who has essentially completed the course* and who has a legitimate interruption in the course, to complete the remaining material in another semester. Students will not be given an opportunity to improve their grades by doing extra work.

- *Drop dates:* Students are responsible for being aware of the drop dates for the current semester. Drop forms will not be back-dated.

- *COVID 19 & BU Community Health Expectations:* Masks are required and face coverings must be worn over the mouth and nose at all times when in public spaces on campus, including classrooms. Students should be prepared to show proof that they are compliant with health attestations and testing in order to attend class. All students are expected to follow all university guidelines with respect to daily symptom checks, testing, social distancing, and mask wearing when they leave their dorm or home. For a detailed description of official BU policies regarding COVID, please visit: `http://www.bu.edu/dos/policies/lifebook/covid-19-policies-for-students/`

- *Inclusion:* This class is to be a place where you will be treated with respect, and it welcomes individuals of all ages, backgrounds, beliefs, ethnicities, genders, gender identities, gender expressions, national origins, religious affiliations, sexual orientations, ability – and other visible and non-visible differences. All members of this class are expected to contribute to a respectful, welcoming, and inclusive environment for every other member of the class.

- *Accommodations for Students with Documented Disabilities:* If you are a student with a disability or believe you might have a disability that requires accommodations, requests for accommodations must be made in a timely fashion to Disability & Access Services, 25 Buick St, Suite 300, Boston, MA 02215; 617-353-3658 (Voice/TTY). Students seeking academic accommodations must submit appropriate medical documentation and comply with the established policies and procedures `http://www.bu.edu/disability/accommodations/`