ENG EK381 Probability, Statistics, and Data Science for Engineers
Fall 2023, College of Engineering, Boston University
Course Information

Motivation: Engineers 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 gives 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-level electives in machine learning, data analytics, and 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 prerequisite.

Course staff:

Instructors:	Prof. Vivek Goyal (`goyal@bu.edu`) Section A1, Tuesdays and Thursdays, 3:30pm–5:15pm, LSE B03 Prof. Hua Wang (`wangh@bu.edu`) Section A2, Mondays and Wednesdays, 3:30pm–5:15pm, PHO 117 Prof. Joshua Kays (`jkays@bu.edu`) Section A3, Tuesdays and Thursdays, 11:00am–12:45pm, PHO 117
Graduate Student Teachers (GSTs):	Andres Chavez Armijos (`aschavez@bu.edu`), Poyraz Durgun (`rdurgun@bu.edu`), Jasmine Hu (`xingyuhu@bu.edu`), Kao Kitichotkul (`rkitich@bu.edu`), Haochen Wan (`lhwan@bu.edu`)

Instructors and GSTs will have weekly office hours. See the Blackboard site for days and times.

Undergraduate Teaching Fellows:	Miriam Bounar (Lead UTF), Jason Calalang, Ahona Dev, Eileen Duong, Mete Gumusayak, Viraj Jaswani, Abdulrahman Kobayter, Dilek Aylin Manav, Jiahe Niu, Vittoria Sama, Athena Wang
Homework Help:	Wednesdays and Thursdays, 7:00pm–9:00pm, PHO 117

Course websites:

Main ("Blackboard"):	`https://learn.bu.edu/`
Alternate website:	`https://probstatdata.bu.edu`
Discussion board:	`https://piazza.com/bu/fall2023/ek381`
Homework submission:	`https://gradescope.com/`

Piazza discussion: When you post to Piazza, you benefit from the collective knowledge of your classmates and instructors. Posts are not anonymous to the course staff. *Please use the discussion board for all conceptual, logistical, and other course-related issues. Emailing instructors should be reserved for issues of a personal nature.*

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

Videos: This is a "flipped" class meaning that, 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 welcome via the discussion board, 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. 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 welcome via discussion board, including pointing out typos.

Grading:

3 Exams (equally weighted)	85%
Homework	10%
In-class quizzes	5%
Letter grade bonus	`if` (quiz average $\geq$ 85%)
	`and` (reasonable attempts for all programming problems)
	`then` (+1/3 letter grade)

Exams: Midterm exams are scheduled for **Thursday, October 5**, and **Thursday, November 2**, **6:30pm–8:30pm**. The final exam date is to be determined. **Students must be present for exams during the scheduled time slots, except under exceptional circumstances that must be discussed with your section's instructor at least ten days in advance.** Unapproved exam absences will result in a zero grade for the missed exam.

Homework: Homework will typically be due weekly on Friday at 7:00pm, with all submissions made on Gradescope. Homework may be submitted 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 discuss with your section's instructor 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. After HW2, incorrect use of Gradescope may result in complete loss of credit, i.e., questions without properly associated pages will not be graded. While this may seem harsh, we will not allow sloppiness to increase the workload of UTFs.
- *Drop lowest score:* To allow for some flexibility, we will drop your lowest homework score and lowest quiz score at the end of the semester.
- *Problem solving in lecture:* 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 sessions:* In each week with homework due, Undergraduate Teaching Fellows will be available to help you Wednesday and Thursday from 7:00pm to 9:00pm in PHO 117. We encourage you to join these sessions to discuss your assignment with other students in the class and ask questions to the UTFs when you get stuck.

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.

Suggested 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 instructors 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)
   - 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 and Variance
   - Important Families of Continuous RVs
   - Conditioning a Continuous RV by an Event
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
   - Correlation Coefficient
   - Jointly Gaussian RVs
   - Random Vectors
   - Gaussian Vectors
6. **Detection**
   - Binary Hypothesis Testing
   - Maximum Likelihood (ML) and Maximum a Priori (MAP) Decision Rules
   - Likelihood Ratio
   - Vector Observations
7. **Estimation**
   - Scalar Estimation
   - MMSE Estimation
   - LLSE Estimation
   - Vector Estimation
8. **Limit Theorems**
   - Sequences and Sums of RVs

- Independent and Identically Distributed (i.i.d.) RVs
- 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
    - Pet Classification Challenge

11. **Markov Chains**
    - Finite Markov Chains
    - Steady-State 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 an instructor immediately. The penalty for academic misconduct at BU is severe. For further information on the BU Academic Code of Conduct, visit the following website: `https://www.bu.edu/academics/policies/academic-conduct-code/`

- *Incomplete grades:* The purpose of an incomplete grade is to allow a student *who has nearly completed the course* and who has a legitimate interruption in the course to complete the remaining material in another semester. 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. 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.

- *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/`