ENG EK381 Probability, Statistics, and Data Science for Engineers
Summer 2025, 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. Jeffrey Carruthers (`jbc@bu.edu`)

Course Websites:

Main:	`https://curl.bu.edu/ek381`
Alternate website:	`https://probstatdata.bu.edu`

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:

We will have 15 minute test at the end of every lecture (except for last one, which will be the final exam). Each is worth 5 percent. There will be twenty one of these. Your best 20 count. However, if you wish to take the final exam, the final exam can make up for your seven worst (or missing) scores.
You can miss up to seven tests for any reason and still get an A in this course.
Of the 100 points available on the end-of-lecture tests, these will be distributed as follows:

- todays lecture and videos
- current chapter
- previous chapter
- comprehensive
- lab-based

pts covered during the last couple of lectures.
Labs: Prior semesters included a programming problem on every homework. The goal was to learn how basic statistical concepts can be implemented by writing algorithms and applying them to large datasets. In the Fall 2024 semester, we switched the focus from writing algorithms to exploratory data analysis. We will use built-in functions to analyze datasets, generate plots, and draw conclusions. The goal is build statistical skills and intuition, especially by visualizing data.
Python: Many large-scale data analysis and machine learning algorithms use Python as the underlying programming language, and we will do the same in EK381. Note that we are not asking you to learn how to program in Python, but rather how to invoke specific built-in functions and interpret the results. Of course, it will be important to follow correct syntax, but examples of each method will always be provided at least once.
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/`