Data Science 5030: Understanding Uncertainty

University of Virginia, Fall 2025

Terence Johnson (trj2j@virginia.edu, 103 102 Cresap Rd.)

Course description: "Provides an in-depth exploration of probabilistic and statistical methods used to understand, quantify, and manage uncertainty. Learn foundational concepts in probability and statistics, simulation techniques, and modern approaches to parameter estimation, decision theory, and hypothesis testing. Topics include parametric and nonparametric methods, Bayesian and frequentist paradigms, and applications of uncertainty in real-world problems."

In this course, we pick up the threads of your math and probability education, and provide a foundation for later developments in machine learning and artificial intelligence, grounded in data. You will take multiple classes on machine learning, so the goal is not to compete with those courses, but to complement and foreshadow what will happen in them. We want to ensure you are ready to be quantitative thinkers and thoughtful modelers for the rest of your career. The course is slightly recursive, with elements building on themselves. This is intentional.

Grading:

1. Recital: Generative modeling from Markov Chains (12%, 9/23)
2. Exam: Traditional blue book exam on probability theory (22%, 11/6)
3. Bayesian Project: Detecting and blocking potential fraudulent banking transactions (12%, 11/25)
4. Frequentist Project: High dimensional inference with genomics data (12%, 12/9)
5. Final Presentation: Portfolio optimization, or expand on the Bayesian/Frequentist projects (12%, Dec 15/9:00-12:00 pm or Dec 16/2:00-5:00 pm)
6. Assignments: Each class will include two to five exercises that can be done as a group (30%)

Lecture Delivery: Lectures will be delivered in-person. All assignments will be distributed and collected through GitHub and Canvas.

Meeting Times: Tuesday/Thursday, 9:30–10:45 or 11:00 – 12:15, SDS 306

Teaching Assistants:

9:30 - 10:45: Oai Tran, dzn7nf@virginia.edu
11:00 - 12:15: Eirik Steen, zxc6hs@virginia.edu

Office Hours: TBA, still working this out with TAs

Course Materials: Course materials will be available through Canvas/GitHub.

Textbook: There is no required textbook. Some books I referenced while building the course are: *Probability and Random Processes* by Grimmett and Stirzaker, *Bayesian Data Analysis 3* by Gelman et al, *Statistical Inference* by Casella and Berger, *Asymptotic Statistics* by van der Vaart,

All of Statistics by Wasserstein, *Econometrics* and *Probability and Statistics for Economists* by Hansen, *Computer Age Statistical Inference* by Efron and Hastie, and *Probability* by Durrett.

Prerequisites: Student must be enrolled in the MSDS program.

Who is this class for? Everyone is welcome to participate in this class, of any age, culture, gender, language or geographic heritage, learning and physical abilities, political or social beliefs, race or ethnicity, religious or spiritual beliefs, sex, and social or economic class. Conversely, everyone participating in this class is expected to respect the dignity and humanity of their peers. Please feel free to approach or email your TA's or instructor with instructions about how you would like to be addressed, including adjustments to or pronunciation of names or preferred pronouns.

Academic Integrity: (Wording suggested by the administration) "I trust every student in this course to fully comply with all of the provisions of the University's Honor Code. By enrolling in this course, you have agreed to abide by and uphold the Honor System of the University of Virginia. All graded assignments must be pledged, including homework and exams. All suspected violations will be forwarded to the Honor Committee. Please let me know if you have any questions regarding the course Honor policy. If you believe you may have committed an Honor Offense, you may wish to file a Conscientious Retraction by calling the Honor Offices at (434) 924-7602."

Generative Artificial Intelligence: First, cite AI code appropriately. Second, make sure it actually runs and does what you intend. Third, recognize that you are not learning how to code, but instead learning how to query an AI system upon which you will become entirely dependent.

Absences, late work, extra credit: I don't take attendance and you don't have to explain your absence from class. If your work will be submitted late, please contact me in advance to explain. There is no extra credit.

Students with Disabilities: If you are entitled to any accommodations, please get in touch with me as soon as possible.

Recommended Technology: GitHub account, Miniconda for package/environment management, WSL2 if a Windows user, VS Code or PyCharm. We'll use NumPy, Pandas, Matplotlib/Seaborn or Plotly, PyMC, and perhaps Scikit-Learn or PyTorch.

Class Schedule: This is the first time the course has been offered, and the probability that we stick to this schedule is approximately zero. The topics at the end would be nice to cover, but our goal is to get through "Classical Hypothesis Testing."

Date	Unit	Topic
08/26/25	Non-Parametric Analytics	Syllabus, Working with data
08/28/25		ECDF, KDE
09/02/25		Conditioning
09/04/25		Resampling, Cross Validation
09/09/25		Dynamics
09/11/25	Reading Day	
09/16/25	Probability Theory	Probability and Random Variables
09/18/25		Distributions and Densities
09/23/25		RECITAL
09/25/25		Multivariate Distributions
09/30/25		Conditional Probability, The Likelihood, Bayes Rule
10/02/25		Maximum Likelihood Estimation
10/14/25		Asymptotics and Simulation
10/07/25		Markov Chains
10/09/25		Steady State Distributions
10/14/25	Bayesian Estimation	Bayesian Estimation
10/16/25		Conjugate Priors and Posteriors
11/04/25	Election Day	
11/06/25		EXAM
11/11/25		MCMC and PyMC
11/13/25	Frequentism	Sampling Distribution
11/18/25		Bootstrap
11/20/25		Classical Hypothesis Testing
11/25/25		M Estimators
11/27/25	Thanksgiving	
12/02/25	Topics	Optimization
12/04/25		Decision Theory
12/09/25		Markov Decision Processes