APHY 470: Statistical methods with applications in science and finance (2019)

Instructor: Sohrab Ismail-Beigi, Becton 307, sohrab.ismail-beigi@yale.edu
Time and location: Tue/Thu 1pm-2:15pm in Mason 107, does **not** observe reading period
Makeup lectures: Weds Feb 20th and Apr 3rd 8:00pm-9:50pm; please reserve these days since they will be used if a regular class is canceled due to unforeseen reasons
Exams: first exam March 5th, second exam April 30th; exams are 1.25 hours long
Other required work: weekly problem sets
Instructor office hours: Wed & Fri 11am-noon in Becton 307 (or by email appointment)
Study halls: Thursday 8pm-10pm Becton 408 : unsupervised reserved room for you to meet you fellow classmates, discussion problem sets, etc.
Required textbook: none; instructor will distribute notes and selections of texts
Recommended for selected reading:

- "An Introduction to Econophysics: Correlations and Complexity in Finance" by Mantegna and Stanley (ISBN 978-0521039871)
- "Options, Futures, and Other Derivatives" by John C Hull (ISBN 978-0132777421)

Prerequisites:

- ENAS 194 or equivalent (ordinary and partial differential equations)
- MATH 222 or equivalent (linear algebra)
- Familiarity with a programming language for solving numerical problems: e.g., ENAS 130 or equivalent (introduction to computing for engineers and scientists); the main point is that you are able to program and solve numerical problems in some computer language and create digital plots of your computed results; the lecture notes and problem set solutions use MATLAB, but in the past students have used Python, Mathematica, etc. A prior programming course is suggested but not required; having prior programming experience of some sort is a must. We will not be using any "fancy" software programming constructs: loops, if-thens, arrays and matrices, indexing, and basic plotting functions are the main constructs that are needed. Talk to the instructor during the first week of classes if you need guidance.

Recommended (but not required) prior courses:

- One year of introductory physics (e.g., PHYS 200/201) recommended
- Prior introductory level course in statistics and/or probability recommended (a 100 level Statistics course, Applied Math 364b, Math 241a or 242a or 251b, or Physics 420)

Problem sets: Problem sets will be posted on Thurdays or Fridays on Canvas. They are due by noon on the second Monday after they are assigned. They are due in Prof. Ismail-Beigi's mailbox on the 2nd floor of Becton Center (the top mailbox of the two stacked ones labeled with his name). Homework handed in after that are out of 50% until solutions are posted online at which time they will no longer be accepted. Solutions will be posted on Canvas within a few days of the problem set being due. Problem sets are a <u>crucial</u> part of the course and learning process. To receive full credit, you must show the logic and steps and not simply produce the final answer out of thin air. Problem set statements will be posted on Canvas about a week before they are due. I will drop the one problem set with lowest score to determine the problem set contribution your grade.

Grading: Course grade will be determined by student performance on the problem sets and the exams. Problem sets will count for 70% and each exam for 15% of the total course score.

Maintaining academic integrity: Maintaining academic integrity is crucial. For problem sets, students are encouraged to collaborate in groups, but must submit their own independent work. Exams are meant to gauge individual understanding. Cheating and/or plagiarism are not tolerated. You can read more about these issues and also good practices for citation of prior work at http://ctl.yale.edu/writing/using-sources

This course teaches how one can model, analyze and understand primarily natural (e.g., physical, chemical, biological) as well as some human (e.g., financial) phenomena that, due to their complexity, must be modeled using probabilistic models. As part of this endeavor, the course will introduce important methods used in statistical physics both in terms of analysis (formulae and mathematics) as well as numerical implementation and modeling. The main methods are stochastic random processes (random walks), Monte Carlo, and analysis of covariance with associated linear algebra. The course has a strong focus on introducing, teaching, and using Monte Carlo (i.e., stochastic) methods with actual numerical examples and data which prior students found was an important benefit and unique point of attraction.

Provisional plan:

1. Jan 15,17: Course logistics, introduction to probability
2. Jan 22,24: Probability distributions, variance, characteristic functions, cumulants
3. Jan 29,31: Estimation of uncertainty, central limit theorem, elements of statistics
4. Feb 5,7: Monte Carlo: introduction, simple examples,
5. Feb 12,14: Monte Carlo: Markov chain sampling, sampling from a desired distribution
6. Feb 19,21: Monte Carlo: importance sampling, more examples
7. Feb 26,28: Monte Carlo: more examples
8. Mar 5,7: First exam (March 5th), no class (March 7th)
9. Mar 26,28: Random walks: introduction, Brownian motion, time-dependent variance
10. Apr 2,4: Random walks: higher dimensions, self-avoiding walks, numerical methods
11. Apr 9,11: Random walks: Weiner processes and continuous random walks
12. Apr 16,18: Variance and covariance matrix, principal component analysis, correlation
13. Apr 23,26: Dimensionality reduction of data from principal component analysis, review
14. Apr 30: Second exam