CHEM/PHYS 598 Course Outline
Instructor: Steve Pressé

Time, Place: MWF 2:00-2:50 PM, PSH552

Grading and Logistics: 15% 1 Midterm, 50% Problem Sets, 25% Final, 10% attendance.

> **Midterm:** is set on March 16th (it is based on what we discuss in class and problem sets only).
> **Problem Sets:** The time and date at which each problem set is due will be written on top of each problem set. Expect a problem set about every two weeks. All code must be printed out along with final results generated by your method. In all of your code, comment, comment, comment. Explain every function, method, etc... Some problem sets will be demanding and can best be described as mini-projects, so please start early.
> **Final:** time and place is set by the University. It is based only on what we discuss in class and problem sets. i.e. extra material in texts is excluded.
> **Attendance:** 3 in-class pop quizzes will given at the beginning of class on randomly selected dates. These quizzes will be simple and will not require any preparation beyond a quick review of material presented in the last class. If students miss a quiz, they will need to provide material (e.g. a doctor's note) substantiating the reason for the absence.
> **Bonus points:** I will select the best problem set solution for each problem set and provide these as solutions to the rest of the class. There is a bonus of 3 points every time your problem set is selected. There is no maximum number of times you can be selected. If you are selected, I will ask you to retype your solutions and provide all of your code within a week. I may also ask you to provide a 20 minute presentation walking us through the coding portion of your problem set every time your solution is picked.

Coding language: To teach a course in data analysis requires coding. Yet having coding prerequisites would exclude a vast majority of students in physics/chemistry. The ideal programming language would be Matlab. But to teach Matlab/programming from scratch along with data analysis would be bundling two courses into one. Instead, here we opt for Mathematica (students can download both Matlab and Mathematica from https://myapps.asu.edu) which is easier to use. Having written Mathematica code, students can more easily move to Matlab based on their needs after the course is complete if they wish.

Course Content: Data analysis courses that go beyond teaching elementary topics such as fitting residuals are rarely offered to students in the physical sciences. Thus, data analysis, much like programming, is something often learned and improvised “on the job". Yet, with an explosion of experimental methods generating large quantities of data, we believe that students would benefit from a clear presentation of methods of data analysis many of which are straightforward to implement and would raise our community standard for how data is currently being treated.

It is often not realistic to expect graduate or undergraduate students alike to take a course covering topics of statistics relevant to their discipline in the physical sciences without taking multiple prerequisites needed to follow the material presented in such courses.

Our goal here is therefore to provide an introduction to exciting new developments in data science, machine learning and statistics in a language accessible to physical scientists willing to learn the necessary programming and mathematics.

To be more precise, this course is appropriate for upper-level undergraduates in the physical sciences. Calculus and basic linear algebra is assumed and while higher-level mathematics are presented in a self-contained fashion, an elementary course in programming and prior knowledge of undergraduate level ordinary differential equations is helpful.

This course places equal emphasis on explaining the logic of existing methods and implementation. It places little emphasis on formal proofs. It is meant as a self-contained single semester course in data analysis, statistical modeling and inference providing students with the right language and logic to be able to solve data analysis problems based on data they themselves collect (if already involved in research) and understand the limitations of the methods they chose to employ.

The course begins with a survey of stochastic models to motivate the problem of parameter learning from the data. This leads to a discussion of frequentist and Bayesian tools of analysis. Along the way, we introduce computational techniques including Monte Carlo methods. Once basic computational tools are established, we discuss statistical models such as the HMM, mixture models and other models as time allows.

Office Hours: WF – 13:00-14:00 or by email request - Steve Pressé (Instructor), spresse@asu.edu, PSF 350

Required texts:

1) Sivia and Skilling, Data Analysis: A Bayesian Introduction, Second edition, 2011.

For the following, full pdfs are available online. Warning: these below are difficult texts and the goal of the class is to have texts like these become accessible:

2) Gelman et al., Bayesian Data Analysis, Third Edition, 2014.
3) Bishop, Pattern Recognition and Machine Learning, 2006.
4) Fox, Bayesian Nonparametric Learning of Complex Dynamical Phenomena, Ph.D. Thesis, 2009.

Course Topics

* denotes topics that we cannot cover in one semester. If time allows, we will pick 1-2 of these based on class interest.

Topic I: Intro to Modeling and Inference
Random variables and sampling;
Probability interpretations, Frequentist interpretation, Bayesian interpretation;
Manipulating probabilities, Motivating data-driven modeling and inference.

Topic II: Markov and other Stochastic Processes
Markov processes and the transition matrix;
Bernoulli, Poisson and birth-death processes;
Weiner, Ornstein-Uhlenbeck and other processes;
Fokker-Planck and diffusion equation, Master equation;
Model and observation likelihoods.

Topic III: Intro to Bayesian Methods
Bayes' theorem, priors, posteriors and conjugate priors;
Parametric priors and Gaussian distributions, Exponential distribution family, Discrete/Categorical/Multinomial family of distributions;
Hierarchical and graphical models;
Model selection/Model averaging, Outliers, Change point detection;
Motivating non-parametric priors, Dirichlet distribution and parametric density estimation; Dirichlet process and stick-breaking.

Topic IV: Intro to Computational Statistics
Introduction to computational statistics; Gibbs sampling, Markov chains;
Metropolis and Metropolis-Hastings sampling;
Metropolis within Gibbs and related schemes;
EM introduction.

Topic V: Time-Independent Models
Mixture models and clustering, Finite mixture models;
EM algorithm, *Infinite mixture models; *Latent feature models; *Infinite latent factor models; *Indian buffet process;
*Binary variables:Bernoulli-Beta process, *Gaussian process.

Topic VI: Time-Dependent Models I
Hidden Markov models, EM, Viterbi algorithm, Bayesian HMM;
Forward/Backward;
*Infinite hidden Markov models, *hierarchical Dirichlet process;
*Weaklimit: approximate iHMMsampling, *Beam sampling: exact iHMM sampling.

Topic VII: Time-Dependent Models II
*Continuous state space models, *Filtering in continuous state space models;
*Kalman filter, *Extended Kalman filter, *Unscented Kalman filter, *Particle filter;
*Stochastic linear dynamic switch, *Hierarchical Dirichlet-Stochastic linear dynamic switch.

Topic VIII: Time-Dependent Models III
*HMM variations, *Setting transition matrix priors, *Sticky-HMM, *Infinite sticky HMM;
*Left-to-right HMM, *Dependent Dirichlet process mixture models;
*Factorial HMM, *Infinite factorial HMM.

Academic Conduct and Integrity:

Students are entitled to receive instruction free from interference by other members of the class. An instructor may withdraw a student from the course when the student's behavior disrupts the educational process per Instructor Withdrawal of a Student for Disruptive Classroom Behavior. The Office of Student Rights and Responsibilities accepts incident reports from students, faculty, staff, or other persons who believe that a student or a student organization may have violated the Student Code of Conduct.

Academic honesty is expected of all students in all examinations, papers, laboratory work, academic transactions and records. The possible sanctions include, but are not limited to, appropriate grade penalties, course failure (indicated on the transcript as a grade of E), course failure due to academic dishonesty (indicated on the transcript as a grade of XE), loss of registration privileges, disqualification and dismissal. For more information, see http://provost.asu.edu/academicintegrity. Additionally, required behavior standards are listed in the Student Code of Conduct and Student Disciplinary Procedures, Computer, Internet, and Electronic Communications policy, and outlined by the Office of Student Rights & Responsibilities. Anyone in violation of these policies is subject to sanctions.

https://clas.asu.edu/resources/academic-integrity-resources
https://clas.asu.edu/resources/disruptive-behavior

Student and Faculty Conduct:

Title IX is a federal law that provides that no person be excluded on the basis of sex from participation in, be denied benefits of, or be subjected to discrimination under any education program or activity. Both Title IX and university policy make clear that sexual violence and harassment based on sex is prohibited. An individual who believes they have been subjected to sexual violence or harassed on the basis of sex can seek support, including counseling and academic support, from the university. If you or someone you know has been harassed on the basis of sex or sexually assaulted, you can find information and resources at https://sexualviolenceprevention.asu.edu/faqs.

As a mandated reporter, we are obligated to report any information we become aware of regarding alleged acts of sexual discrimination, including sexual violence and dating violence. ASU Counseling Services, https://eoss.asu.edu/counseling, is available if we wish discuss any concerns confidentially and privately.

Accessibility Statement:

In compliance with the Rehabilitation Act of 1973, Section 504, and the Americans with Disabilities Act as amended (ADAAA) of 2008, professional disability specialists and support staff at the Disability Resource Center (DRC) facilitate a comprehensive range of academic support services and accommodations for qualified students with disabilities. Qualified students with disabilities may be eligible to receive academic support services and accommodations. Eligibility is based on qualifying disability documentation and assessment of individual need. Students who believe they have a current and essential need for disability accommodations are responsible for requesting accommodations and providing qualifying documentation to the DRC. Every effort is made to provide reasonable accommodations for qualified students with disabilities. Qualified students who wish to request an accommodation for a disability should contact the DRC by going to https://eoss.asu.edu/drc, calling (480) 965-1234 or emailing DRC@asu.edu. To speak with a specific office, please use the following information:
ASU Online and Downtown Phoenix Campus
University Center Building, Suite 160
602-496-4321 (Voice)
Polytechnic Campus
480-727-1165 (Voice)
West Campus
University Center Building (UCB), Room 130
602-543-8145 (Voice)
Tempe Campus
480-965-1234 (Voice)