Syllabus

The primary aim of this course is to help you think clearly and quantitatively about open questions in statistical inference, and in particular those with applications to Biology. We will provide you with the basic knowledge of the biology and concepts from statistical mechanics and probability theory that have been useful so far in exploring topics in Biology. The aims are both to help you analyze data better, but also to think about experimental design. Not surprisingly, the same statistical techniques used for data analysis are also being used for artificial intelligence. This gives us an opportunity to answer poorly posed questions in neuroscience, psychophysics, population biology, cell biology, and applied math.

The course will have bi weekly problem sets and a final project. A background in linear algebra, Fourier analysis and basic statistics is assumed.

To determine if we have achieved the aims of the course, we will have a final project in lieu of a final exam. The project will be a research project undertaken by groups of students (group size up to 5 students depending on enrollment). There will also be a mid term exam. There will be two lectures and one section each week.

Textbook: Data Analysis, A Bayesian Tutorial by D. S. Sivia
Recommended books: The lectures wild draw from multiple textbooks and primary literature. Some books that will be used are:

Information Theory, Inference, and Learning Algorithms by David MacKay

Maximum Entropy in Action by Buck and Macaulay

Principles of Statistical Inference by Cox

Spikes by Riecke, Bialek et al.

Neuroscience by Purves et al.

Biophysics, W. Bialek

Probability theory: the logic of science, E. T. Jaynes

Staff

Professor Sharad Ramanathan: sharad@cgr.harvard.edu

Office Hours: Thursday, 4-5PM. Northwest Labs 435

TAs: (Sections and Office Hours start 2/6)

Sam Melton: sam.melton@gmail.com

Section: Monday 4:30-5:30 pm, MD 123

Office Hours: Tuesday 1:30-2:30 pm, NW 430

Po-Yi Ho: poyiho@g.harvard.edu

Section: Tuesday 4-5pm, MD 323

Office hours: Tuesday 11:30am-12:30pm, Pierce 407

Allen Lin: allenlin@g.harvard.edu

Section: Friday 2:30-3:30pm, MD 323

Office hours: Friday 3:30-4:30pm, MD 221

Course Outline

Introduction, Information theory

1. Basic concepts.

    b. entropy, information and compression
    c. Application to horse races and living in a fluctuating environment
2. Shannons theorems, Nyquist frequency, Compressive sensing

Basic Probability theory and inference:
a. Introduction to deductive logic and plausible reasoning.
b. Bayes' rule, corollaries, priors.
c. Parameter estimation and model selection.
d. Maximum entropy principles for prior selection.
e. Markov and hidden markov models.

Application of Probability theory as logic to analyze one-dimensional signal trains:
a. Chemotaxis and food search as an inference problem.
b. Sequence analysis and tracing evolutionary history of proteins.
c. Analyzing neural spike trains.

1. Text processing and playing hangman
    d. Analysis of music as a one-dimensional time series.

Non-parameteric methods

1. SVD and PCA
2. Deep Learning
3. Renormalization group and relation to deep learning
4. Renormalization group and building models with limited data.