

Psychology 205C  
Computational Statistics with Matlab

**Instructor:** Mark Steyvers (mark.steyvers@uci.edu)

**Location/time:** SBSG2200, Fri 9-11:50am

**Overview**

Computational statistics is an approach in statistics that focuses on computer intensive procedures to transform raw data into knowledge. The idea is to rely on computer algorithms to solve problems in statistical analysis that would be too difficult to solve using analytic techniques by themselves. By using a computational approach, the researcher can often approximate solutions to complex problems without being forced to make unrealistic assumptions (e.g. normality and independence) required for some analytic techniques.

One area in computational statistics is concerned with issues of data visualization and explorative data-analysis. Often, this approach is used at the start of data analysis when the researcher does not fully understand yet what is even contained in a large and high-dimensional datasets. In this course, we will spend little time on this area of computational statistics.

Instead, we will spend most of our time on the part of computational statistics that is concerned with Monte Carlo approaches in latent probabilistic models, also referred to as Bayesian networks, belief nets, and graphical models. The goal here is to develop precise probabilistic models for the data that can be used to explain and describe the data as well as make predictions about future data. It can often be challenging to find the posterior distributions over the variables of interest. A number of computational procedures have been developed to approximate posterior distributions using sampling techniques. One such approach is known as Markov chain Monte Carlo.

In this course, the goal is to give students a practical, programming-based introduction to latent probabilistic modeling using Matlab. The course will focus on sampling techniques such as Metropolis-Hasting and Gibbs sampling, both forms of Markov chain Monte Carlo. Time permitting, we will also discuss sequential Monte Carlo techniques such as particle filters, and non-parametric techniques such as the Chinese Restaurant Process.

**Course format**

The class will be taught without lectures. Instead, the emphasis will be on in-class programming exercises using MATLAB. Students will be required to attend a weekly three hour laboratory session and completing assigned programming exercises in class. *Students will need a laptop* preferably with MATLAB already installed (including the statistics toolbox). If Matlab is not installed, installation software might be provided. There will be no required textbook. Instead, course notes will be provided that contain brief introductions to the weekly topics along with a number of programming exercises and questions.

## **Grading**

There will be no exams in this course. Over the course of the quarter, students are required to submit the answers to exercises. Often this involves showing the output of a Matlab program to demonstrate that the program performs accurately, displays the correct result, etc. Grading will be based on level of participation (attendance, completion of requirements in class exercises), and quality of completed answers.