PROB 140

Probability for Data Science

Spring 2022

WEEK 6 STUDY GUIDE

The Big Picture

We develop an algorithm that uses a Markov chain to simulate a probability distribution on an intractably large outcome space. We then move on to the variability in distributions, necessary for assessing the accuracy of estimates.

- *Monte Carlo* methods use simulation to address problems that are intractable by math or by complete enumeration.
- *Markov Chain Monte Carlo* (MCMC) can be used to simulate probability distributions on intractably large outcome spaces, even when the normalizing constant of the distribution can't be calculated.
- The *standard deviation*, familiar to you from Data 8 as a measure of the spread in a data distribution, is defined as a measure of spread in the distribution of a random variable.
- *Variance*, which is the mean squared error and the square of the standard deviation, has better computational properties.

Week At a Glance

Mon 2/21	Tue 2/22	Wed 2/24	Thu 2/25	Fri 2/26
	Lecture	Sections	Lecture	Sections
Homework 6 (Due Mon 2/28)	**HW 5 due at noon** **Lab 2 due at noon**			
Lab 3 (Due Mon 2/28)				Lab 3 Party 3pm to 5pm
Holiday	Work through Sec 11.2 and 11.3 carefully	Complete Lab 3	Skim Ch 12; prepare for midterm	Work through Ch 12; prepare for midterm

Reading, Practice, and Live Sessions

Sections	Topic	Lectures: Prof. A.	Sections: GSIs	Optional Additional Practice
Ch 11	**Markov Chain Monte Carlo (MCMC)** - 11.2 solves the code-breaking problem with a tiny alphabet, by complete enumeration - 11.3 develops a general Markov chain Monte Carlo method that can be used to solve the problem with a large alphabet	Tuesday 2/22 - Recap of the code breaking problem, with a tiny alphabet - Using MCMC to solve the problem with a large alphabet	Wednesday 2/23 Ch 11: - Proposal transition matrix in Lab 3 - Common type of Markov Chain problem	None Prepare for midterm
Ch 12	**Variance and Standard Deviation** - 12.1 has the basics of SD and variance; much of this should be an easy read - 12.2 connects variance and prediction - 12.3 shows how expectation and variance can be used to bound the tails of a distribution - 12.4 has examples of distributions with heavy tails, for students interested in economics, natural language processing, etc	Thursday 2/24 SD and variance: - Definition, alternative computational method, examples - Use in prediction - Tail bounds	Friday 2/25 Midterm review	

There are no exercises at the end of Chapter 10 because the methods of Chapter 11 make many problems easier to solve.