PROB 140

Fall 2022
WEEK 2 STUDY GUIDE

Probability for Data Science

The Big Picture

We continue to develop the basic toolkit: the rules of probability, how to work with numerical random quantities and collections of events.

- In Data 8 you saw a *statistic* defined as a number that you compute based on a sample. The more general concept is that of a *random variable*, which is a function on the outcome space.
- *Distributions* describe how probability is spread over a set of values. Every random variable has a distribution; pairs and larger groups have *joint distributions*.
- There is a formula for the chance of the union of overlapping events, with a famous application.
- If there is a complicated dependence structure, you might not be able to calculate exact or even approximate chances. Sometimes the best you can do is find *bounds* for a chance.
- *Symmetry* in random permutations and simple random samples greatly simplifies calculations.

Week At a Glance

Mon 8/29	Tue 8/30	Wed 8/31	Thu 9/1	Fri 9/2
	Lecture	Sections	Lecture	Sections
HW 1 Due HW 2 (Due 12 NOON Tue 9/6)				Lab Party 10-1 HW 2 Party 3-5
Lab 1 (Due Tue 9/6)				
Work through Chapter 2, skim Chapter 3	Work through Chapters 3 and 4	Work through Chapter 4, skim Sec 5.1	Work through Sec 5.1 and Sec 5.4, skim Sec 5.1 and Sec 5.2	Work through Ch 5

Reading, Practice, and Class Meetings

Book	Topic	Lectures: Prof. A.	Sections: GSIs	Optional Additional Practice
Ch 3	Random variables - 3.1 has the definition - 3.2 defines the distribution of the random variable, and shows how to find probabilities of events based on the random variable - 3.3 shows how random variables can have two kinds of equality	**Tuesday 8/30** - The key ideas in Chapters 3 and 4, focusing more on the math than the code	**Wednesday 8/31** - Exercises selected to help with assignments: Chapter 3 Ex 3, 5ab Chapter 4 Ex 3, 4ab	Chapter 3 4, 7
Ch 4	Pairs of random variables - 4.1 is the two-variable version of 3.2: joint distributions, and finding probabilities - 4.2 has examples you should study - 4.3 shows how to extract the behavior of one random variable from the combined behavior of two - 4.4 shows how to update chances for one random variable given the value of another - 4.5 looks at how joint distributions help us understand dependence and independence; note the acronym "iid"			Chapter 4 Do as much as you can of all five exercises.

Chapters 3-4 aren't difficult technically, but they contain many basic concepts and essential terminology.

(continued below)

Reading, Practice, and Class Meetings

Book	Topic	Lectures: Prof. A.	Sections: GSIs	Optional Additional Practice
Ch 5	- 5.1: Simple bounds for the chance of an overlapping union, to be used when we can't find the chances of the overlaps - 5.2: The exact chance of a union, overlapping or not. This is called the *inclusion-exclusion* formula and requires that we be able to find the chances of the overlaps. - 5.3: One of the most famous applications of inclusion-exclusion is to *fixed points* of a *random permutation*, also known as *matches* - 5.4: Summary of results on symmetry in random permutations and simple random sampling, some of which we'll have used earlier	**Thursday 9/1** - Some discussion of bounds and symmetry - Main focus on inclusion-exclusion and the matching problem	**Friday 9/2** - Bounds, symmetry, and inclusion-exclusion: Chapter 5 Ex 3, 1, 9, 12; comparisons with other exercises	Chapter 5 5, 6, 10, 13

Chapter 5 is more technically detailed than Chapters 3-4.