PROB 140
Probability for Data Science

Fall 2021
WEEK 2 STUDY GUIDE

The Big Picture
We continue to develop the basic toolkit: how to work with collections of random variables 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 of random variables have *joint distributions*, from which you find the chance of any event determined by the random variables.
- 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 8/30	Tue 8/31	Wed 9/1	Thu 9/2	Fri 9/3
	Instructor's Session		Instructor's Session	
		GSIs' Sessions		GSIs' Sessions
HW 1 Party 12-2PM **HW 1 Due** HW 2 (Due Tue 9/7)				
Lab 1A due Lab 1B (Due Tue 9/7)				Lab 1B Party 3-6PM
Skim Chapter 3 and Chapter 4	Read/watch Chapter 3 and Chapter 4	Read/watch Sec 5.1 and Sec 5.4	Read/watch Sec 5.2 and Sec 5.3	Fill any holes you left in Chapter 1-5

Reading, Practice, and Live Sessions

Sections	Topic	Live Sessions: Prof. Sahai	Live Sessions: GSIs	Recommended 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	Tue 8/31 - The key ideas in Chapters 3 and 4, focusing more on the math than the code		Chapter 3 2, 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"		Wed 9/1 - Conditioning and Bayes: points to notice - Random variables and equality Chapter 2 Ex 2 Chapter 3 Ex 5, 6 Chapter 4 Ex 2	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. You'll need the code for Lab 1B.

Sections	Topic	Live Sessions: Prof. A.	Live Sessions: GSIs	Recommended Practice
Ch 5	- 5.1: Simple bounds for the chance of an overlapping union - 5.2: The exact chance of a union, overlapping or not (requires the chances of all 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	Thu 9/2 - Some discussion of bounds and symmetry - Main focus on inclusion-exclusion and the matching problem	Fri 9/3 - Bounds, symmetry, and inclusion-exclusion: Chapter 5 Exercises 3, 1, 9, 12; comparisons with other exercises	Chapter 5 5, 6, 10, 13, 14

Chapter 5 is more technically detailed than Chapters 3-4. You'll need it (as well as Chapter 3-4) for HW 2.