DATA 140

Probability for Data Science

Spring 2024

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. This gives us the tools to study some fundamentally important families of distributions.

- If events have 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.
- There is a formula for the exact chance of the union of overlapping events, with a famous application.
- Distributions on a large finite number of values can be approximated by distributions on infinitely many values; a fundamentally important example of this is introduced.
- Random samples often result in random counts. The distribution of the count depends on the method of sampling.
- If the sample is a fixed number of i.i.d. success/failure trials, the distribution of the number of successes is *binomial.* The shape of the distribution can be understood by using consecutive odds ratios.
- In some situations, the binomial distribution is well approximated by a *Poisson* distribution, introduced earlier.

Week At a Glance

Mon 1/22	Tue 1/23	Wed 1/24	Thu 1/25	Fri 1/26
	Lecture	Sections	Lecture	Mega sections
Lab 1 Due 5 PM Lab 2 (Due Mon 1/29)			Lab 2 Party 9 AM - 11 AM	
HW 1 Due 5 PM HW 2 (Due Mon 1/29)				HW 2 Party 2 PM - 4 PM
Finish working through Ch 4; Skim Ch 5	Work through Ch 5	Finish working through Ch 5; skim Ch 6	Work through Ch 6	Fill any holes you left in working through Ch 4-6

Reading, Practice, and Class Meetings

Book	Topic	Lectures: Prof. A.	Sections: TAs	Optional Additional Practice
Ch 5	- 5.1: Simple bounds for chances of unions and intersections - 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*; this can be approximated by a distribution on infinitely many values - 5.4: Summary of results on symmetry in random permutations and simple random sampling	Tue 1/23 Highlights of Ch 5	Wed 1/24 - Lab 2 Part 1: a new look at the TVD - Chapter 5 Ex 9, 12; also 1 if there is time	Chapter 5 1, 5, 6, 10, 13
Ch 6	- 6.1: In a fixed number of i.i.d. 0/1 trials, the number of successes has a *binomial* distribution - 6.2: Examples you should read - 6.3 extends the binomial to the *multinomial* case where each trial has several possible outcomes - 6.4 compares the number of successes when the sampling is with replacement (binomial) and when the sampling without replacement (*hypergeometric*) - 6.5 uses odds ratios to study the shape of binomial histograms, and finds the mode - 6.6 uses odds ratios to show that under some conditions the binomial has a *Poisson* limit	Thu 1/25 Highlights of Ch 6	Fri 1/26 - Ch 6 Ex 2, 4, 10, 11	Chapter 6 1, 5, 9, 12