PROB 140

Probability for Data Science

Fall 2020

WEEK 3 STUDY GUIDE

The Big Picture

This week is largely about two of the principal families of distributions: the binomial and the Poisson.

- Random samples often result in random counts, such as the number of voters who favor a candidate. 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. The Poisson is our first distribution on infinitely many values.
- Randomizing parameters can have dramatic effects on dependence and independence. A Poisson number of i.i.d. success/failure trials has beautiful and powerful properties.

Week At a Glance

Mon 9/7	Tue 9/8	Wed 9/9	Thu 9/10	Fri 9/11
	Instructor's Session		Instructor's Session	
		GSIs' Sessions		GSIs' Sessions
	Checkpoint Week 3 (Due Wed 9/9)	**Checkpoint Week 3 Due**		
	HW 2 Due HW 3 (Due Mon 9/14)			HW 3 Party 6-7PM
	Lab 1B Due Lab 2A (Due Mon 9/14)		Lab 2A Party 6-7PM	
	Read 6.1, 6.2, skim 6.5	Read Chapter 6	Skim 7.1, 7,2	Read Ch 7

Reading, Practice, and Live Sessions

<table>
<tr><th>Sections</th><th>Topic</th><th>Live Sessions: Prof. A.</th><th>Live Sessions: GSIs</th><th>Recommended Practice</th></tr>
<tr><td>Ch 6</td><td>Binomial and its relatives
- 6.1 is about a fixed number of i.i.d. success/failure trials; the number of successes has a binomial distribution
- 6.2 has examples you should read
- 6.3 extend the binomial to the multinomial case where each trial has several possible outcomes, not two
- 6.4 compares the number of successes when sampling with replacement (binomial) and the number of successes when sampling without replacement (hypergeometric), and shows when the two are almost the same
- 6.5 examines the shape of the binomial histogram, and identifies the mode, by studying odds ratios
- 6.6 uses odds ratios to show that under some conditions the binomial has a Poisson limit</td><td rowspan="2">Tuesday 9/8

- Straightforward but important observations about success counts
- Deeper dive into the math to explain what we see
- An approximation, leading to a new class of distributions

Checkpoint is based on Chapter 6

Thursday 9/10

Poissonization:
- Beautiful calculations with surprising results
- Pay attention to the math because you’ll need the methods again</td><td rowspan="2">Wednesday 9/9
- Ch 6 Ex 2
- Lab 2A Part I: a new look at total variation distance
- Ch 6 Ex 5

Friday 9/11
Poisson approximation and Poissonization:
- Ch 6 Ex 10
- Ch 7 Ex 2
- Ch 7 Ex 8
Relations with other exercises</td><td>Chapter 6
1, 4, 11, 12</td></tr>
<tr><td>Ch 7</td><td>Poissonization
- 7.1 has properties of the Poisson distribution
- 7.2 asks the same questions as 6.1, but with a Poisson number of trials
- 7.3 extends this to trials with more than two categories, analogous to 6.3</td><td>Chapter 7
3, 4, 5</td></tr>
</table>

Lab 2 is about measuring the quality of the Poisson approximation in 6.6.