DATA 140

Probability for Data Science
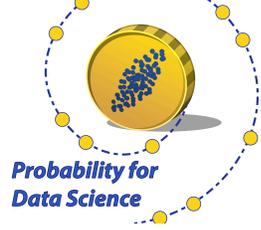
WEEK 7 STUDY GUIDE

The Big Picture

We study a quantity involved in finding variances of sums (and also the correlation between two variables). We then move to random variables with a continuum of values, via one of the most important theorems in probability.

- *Covariance* helps calculate variances of sums and can be normalized to become *correlation*.
- General properties of variance and covariance help us calculate the variances of the main distributions.
- We know how to find expectations and variances of sums of random variables. To find the distribution of a sum, we can use partitioning as before. But a more abstract math concept of a *probability generating function* lets us quickly calculate distributions of sums in special cases.
- Many of the simulations in Data 8 are evidence of the *Central Limit Theorem* in action: the distribution of the sum of a large i.i.d. sample is roughly normal. We use this to construct confidence intervals for the population mean.

Week At a Glance

Mon 3/2	Tue 3/3	Wed 3/4	Thu 3/5	Fri 3/6
Regular OH 10AM - 3PM in Warren 101B	Lecture	Sections	Lecture	Mega sections
Lab 4 Due 5PM Lab 5 (due 5PM Monday March 9)			Lab 5 party 3-5 PM in Warren 101B	
HW 6 Due 5PM HW 7 (due 5PM Monday March 9)				HW 7 party 2-5 PM in Evans 330
Skim Chapter 13	Work through Chapter 13	Work through Chapter 13	Work through Sections 14.1-14.2, skim the rest of Ch 14	Work through Chapter 14

Reading, Practice, and Class Meetings

Book	Topic	Lectures: Prof. A	Sections: TAs	Optional Additional Practice
Ch 13	**Covariance** - 13.1-2 define covariance and establish its main properties - 13.3 covers the important special case of sums of independent variables - 13.4 covers variances of dependent sums - 13.5 compares dependent and independent sums via a *correction factor*	Tuesday 3/3 Variance of a sum: - Covariance and main properties - Sums of independent random variables - Handling dependence	Wednesday 3/4 Ch 13 - Ex 1, 11, 15	Ch 13 - 2, 3, 4, 6, 13
Ch 14	**Sums and the CLT** - 14.1-14.2 cover an abstract math method for understanding probability distributions; 12.2 finds exact distributions of i.i.d. sample sums. - 14.3 states the Central Limit Theorem and formally defines the normal curve - 14.4 shows how to work with the normal curve in Python; **this is for you to read by yourself** - 14.5-14.6 cover the distribution of the i.i.d. sample mean, and hence the use of the sample mean in confidence intervals	Thursday 3/5 - Our first generating function: a math technique for understanding distributions - The CLT and some consequences	Friday 3/6 Ch 14 - Ex 1, 4, 5, 6	Ch 14 - 2, 3, 7