PROB 140

Probability for Data Science

Fall 2022

WEEK 8 STUDY GUIDE

The Big Picture

We establish a principal concept central to variance calculations. 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 technique called *probability generating functions* 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 10/10	Tue 10/11	Wed 10/12	Thu 10/13	Fri 10/14
	Lectures	Sections	Lecture	Sections
HW 7 Due HW 8 (due Mon 10/17)				HW 8 party 3PM - 5PM
Lab 4 (due Mon 10/17)				Lab 4 party 10AM - 12 noon
Work through Sec 13.1, skim Sec 13.2 and 13.3	Work through Ch 13	Skim Section 14.1, 14.2	Skim Chapter 14	Work through Chapter 14

Reading, Practice, and Class Meetings

Book	Topic	Lectures: Prof. A.	Sections: GSIs	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 10/11 Variance of a sum: - Covariance and main properties - Sums of independent random variables - Handling dependence	Wednesday 10/12 Ch 13: - Ex 1, 11, 13	Ch 13 - 2, 3, 4, 6, 15
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 10/13 - Our first generating function: a math technique for understanding distributions - The CLT and some consequences	Friday 10/14 Ch 15: - Ex 1, 5, 3, 10	Ch 14 - 2, 3