PROB 140



Probability for Data Science

Fall 2021

WEEK 8 STUDY GUIDE

The Big Picture

We move to random variables with a continuum of values, via one of the most important theorems in probability.

- 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.
- The exact distributions or 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.
- The normal is a continuous curve that acts a probability distribution. We formally define the *density* of a random variable with a continuum of values, and extend the concepts of cdf and expectation to this situation.
- Along with the normal, we study two major distribution families: the uniform and the exponential.

Week At a Glance

Mon 10/11	Tue 10/12	Wed 10/13	Thu 10/14	Fri 10/15
	Instructor's Session		Instructor's Session	
		GSIs' Sessions		GSIs' Sessions
HW 7 Party 12-2PM	HW 7 Due HW 8 (Due Tue 10/19)			
	Lab 5A and 5B (Due Tue 10/19)			Lab 5 Party 3-6PM
Skim Sections 14.1-14.2	Read Sections 14.1-14.2, skim the rest of Ch 15	Read Ch 14, skim Section 15.1	Read Sections 15.1-15.3	Read Chapter 15 (you can postpone 15.5 till next week)

Reading, Practice, and Live Sessions

Sections	Topic	Live Sessions: Prof. Sahai	Live Sessions: GSIs	Recommended Practice
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 probability theory of the i.i.d. sample mean, and hence the use of the sample mean in confidence intervals	Tuesday 10/12 - Our first generating function: a math technique for understanding distributions - The CLT and some consequences	Wednesday 10/13 Ch 14: - Ex 1, 4, 5	Ch 14 - All exercises not covered in section
Ch 15	Random Variables with Densities - 15.1-15.2 define a "continuous" probability histogram, and generalize the concept of density from Data 8 histograms - 15.3 covers expectation (including variance) and has examples including the uniform distribution family - 15.4 covers the exponential distribution family - 15.5 shows how to do calculus in SymPy, for your next lab	Thursday 10/14 Random variables on a continuum of values: extending all previous concepts to this case, and recognizing a benefit of the continuous world: single points affect probability calculations	Friday 10/15 Ch 15: - Ex 7, 2, 6	Ch 15 - 1, 3, 5, 9, 10