Lisa Yan and Jerry Cain
CS 109

Section #6
October 27-28, 2020

Section 6: Samples

Based on the work of many CS109 staffs

1 Warmups

1.1 Sums of Random Variables

For each X and Y below, let X and Y be independent.

1. Let $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$. What is μ and σ for $X + Y \sim \mathcal{N}(\mu, \sigma)$?
2. Let $X \sim \text{Uni}(0, 1)$ and $Y \sim \text{Uni}(0, 1)$. What is the PDF for $X + Y$?
3. In general, two random variables X and Y, what is the PDF f of $X + Y$?

1.2 Sample and Population Mean

Computing the sample mean is similar to the population mean: sum all available points and divide by the number of points. However, sample variance is slightly different from population variance.

1. Consider the equation for population variance, and an analogous equation for sample variance.

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2 \qquad (1)$$

$$S_{biased}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \qquad (2)$$

   S_{biased}^2 is a random variable which estimates the constant σ^2. Is $E[S_{biased}^2]$ greater or less than σ^2?

2. Write the equation for $S_{unbiased}^2$ (known simply as S^2 in the slides). This is known as *Bessel's correction*.

1.3 Sum of I.I.D Random Variables

What is the distribution (with name and parameter(s)) of the average of n i.i.d. random variables, X_1, ..., X_n, each with mean μ and variance σ^2?

2 Problems

2.1 Variance of Hemoglobin Levels

A medical researcher treats patients with dangerously low hemoglobin levels. She has formulated two slightly different drugs and is now testing them on patients. First, she administered drug A to one group of 50 patients and drug B to a separate group of 50 patients. Then, she measured all the patients' hemoglobin levels post-treatment. For simplicity, assume that all variation in the patient outcomes is due to their different reactions to treatment.

The researcher notes that the sample mean is similar between the two groups: both have mean hemoglobin levels around 10g/dL. However, drug B's group has a **sample variance** that is 3 $(\text{g/dL})^2$ **greater** than drug A's group. The researcher thinks that patients respond to drugs A and B differently. Specifically, she wants to make the scientific claim that drug A's patients will end up with a significantly different spread of hemoglobin levels compared to drug B's.

You are skeptical. It is possible that the two drugs have practically identical effects and that the observed different in variance was a result of chance and a small sample size, i.e. the **null hypothesis**. Calculate the probability of the null hypothesis using bootstrapping. Here is the data. Each number is the level of an independently sampled patient:

Hemoglobin Levels of Drug A's Group ($S^2 = 6.0$):
13, 12, 7, 16, 9, 11, 7, 10, 9, 8, 9, 7, 16, 7, 9, 8, 13, 10, 11, 9, 13, 13, 10, 10, 9, 7, 7, 6, 7, 8, 12, 13, 9, 6, 9, 11, 10, 8, 12, 10, 9, 10, 8, 14, 13, 13, 10, 11, 12, 9

Hemoglobin Levels of Drug B's Group ($S^2 = 9.1$):
8, 8, 16, 16, 9, 13, 14, 13, 10, 12, 10, 6, 14, 8, 13, 14, 7, 13, 7, 8, 4, 11, 7, 12, 8, 9, 12, 8, 11, 10, 12, 6, 10, 15, 11, 12, 3, 8, 11, 10, 10, 8, 12, 8, 11, 6, 7, 10, 8, 5

How would this calculation be different if you were interested in looking at the statistical significance of the difference in sample mean? Or the 95th percentile?

2.2 Medication Quantities

Megha has a health condition that requires unpredictable amounts of medication. Every day, there is a 20% chance that she feels perfectly fine and requires no medicine. Otherwise, she needs to take a dose of medication. The necessary dose is equally likely to be any value in the continuous range 1 to 5 ounces. How much medicine she needs on any given day is independent of all other days.

Megha's insurance will fully cover 90 ounces of medicine for each 30-day period. What is the probability that 90 ounces will be enough for the next 30 days? Make your life easier by using Central Limit Theorem.