ATS 655: Objective Analysis
HW2

Tyler Barbero

18 February 2022

0. Time Management

Estimate of Time to Completion: 10 hrs
Maximum Allotted Time to Completion: 15 hrs
Actual Time to Completion: 7 hrs
Collaborators: Amanda Bowen, James Larson, Andrey Marsavin, Ivy Glade, Jack Cahill, Spencer Jones, Anindita Chakraborty, Olivia Sablan

Problem 1: Testing composite-averages using bootstrapping and the t/z-test

a. What was the average pressure in 2014 ($\bar{P}$)? What was the average pressure when it rained ($\bar{P}_R > 0$)?

$\bar{P} = 846.3$ hPa and $\bar{P}_R > 0 = 847.0$ hPa

b. Using a t/z-test, employ a standard hypothesis test to determine whether the local pressure is anomalously high or low during times when it is precipitating (R ¿ 0). Describe all of your steps. Do you think that a t-test/z-test is appropriate in this situation?

1. Confidence level $\alpha = 0.05$. Use a two-tailed test because the pressure is high or low.
2. H_o: pressure is not anamolously (low or high) when it is raining.
3. The z-score, which assumes the following assumptions: each pressure value is independent from one another and the underlying distribution is normally distributed.
4. The critical z value z_c: +-1.96 for the 95% confidence interval.

Using a z-test, we compare the average pressure when it is raining to the pressure over 2014

$$z = \frac{\bar{P}_{R>0} - \bar{P}}{\sigma_{\bar{P}}/\sqrt{N}}$$

and plug in the values into the formula, I calculate z = 2.439, thus we reject the null hypothesis, that pressure is not anomalously low or high when it is raining. I think the z-test is appropriate in this situation versus a t-test because we have sample size $N > 30$. I also plotted the PDFs of the original pressure and pressure when it is raining (not shown here). The total pressure PDF looks similar to a Gaussian distribution but skewed slightly to the left, so I think the z-test is a good appropriate.

c. Instead of the t/z-test, use bootstrap sampling to determine whether the local pressure is anomalously high or low during times when it is precipitating. How does your answer compare with your results from Part (b)? What does this tell you about your conclusion in Part (b)?

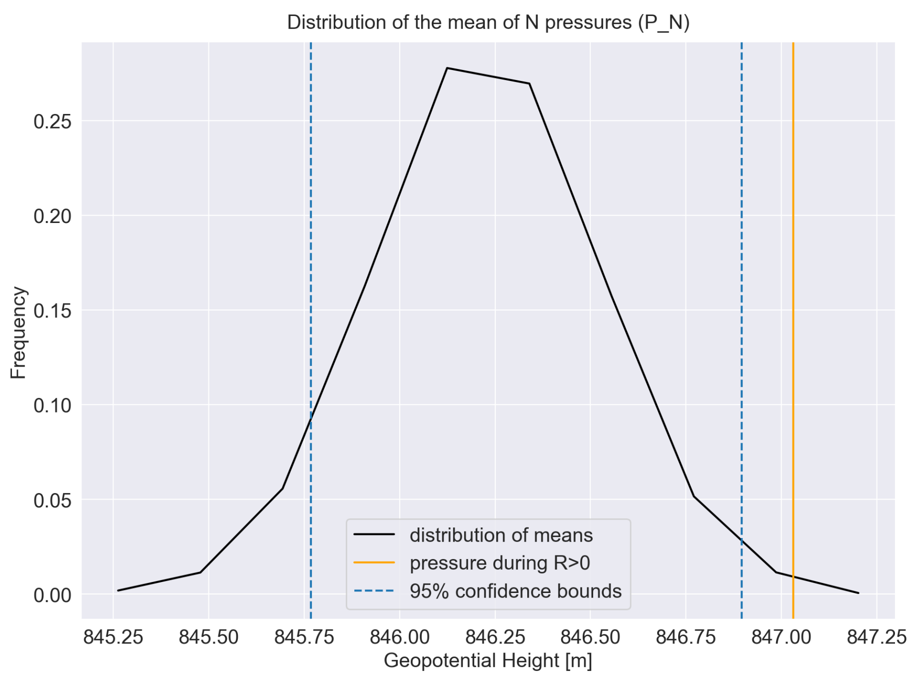

Figure 1: Plot for Problem 1c.

By doing bootstrapping, we are taking N subsamples of the original data, taking the mean of those subsamples and plotting the distribution of the N average pressures during rain events. This way we are not assuming anything

about the underlying distribution of the original data. Along with this we can plot the 95% confidence interval of the sample mean distribution and the mean of the pressures when it is raining. This plot below tells us the probability by CHANCE of getting our $P_{R>0}$, is very rare given the orange line is outside confidence interval. Therefore we reject the H_o again.
Using a bootstrap example, we come to the same conclusion than with the z-test. By utilizing the central limit theorem for this sample of N=100, we can approximate the sample means distribution as a normal distribution and therefore our answer is the same as the z-test where we assumed a normal distribution and independent events. This tells me that my conclusion in part b was a good approximation.

Problem 2: Bayesian vs Frequentist

a. Use hypothesis testing, can you reject the null hypothesis that the air is actually pristine air from the west (H0 : = 0)? Use a two-tailed confidence interval of 95%.
I calculate z-score by

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{N}}$$

where I got z = 1.3. This falls between the bounds of of the z_c values of +-1.96. So, we do not reject the null hypothesis H_o.

b. Use Bayes' Theorem to compute the probability that the air came from the west over your 100-sample period.

$$P(West|\bar{C}_{N_{100}} = 0.13) = \frac{P(\bar{C}_{N_{100}} = 0.13|West) * P(West)}{P(\bar{C}_{N_{100}} = 0.13|West) * P(West) + P(\bar{C}_{N_{100}} = 0.13|East) * P(East)}$$

We cannot actually solve for $Pr(C_{N_{100}}) = 0.13$ therefore we reformat the problem into $Pr(West|C_{N_{100}} = 0.13\delta <= \bar{C}_{N_{100}} <= 0.13 + \delta)$

$$P(West|\bar{C}_{N_{100}} = 0.13) = \frac{P(0.14 > \bar{C}_{N_{100}} > 0.13|West) * P(West)}{P(0.14 > \bar{C}_{N_{100}} > 0.13|West) * P(West) + P(0.14 > \bar{C}_{N_{100}} > 0.13|East) * P(East)}$$

We have or can calculate all of these values with the information given. I get that the probability the air came from the west using Bayes Theorem is **0.27 or 27%**.

c. Explain your results and what they imply about the two different methods. Be sure to specifically discuss the behavior in the limits of large and small γ.
We see that the Bayesian approach has a 100% frequency of getting the correct answer at each extreme and a dip in the middle, while the Frequentist has a very high (95% frequency) of getting the correct answer at γ=1.0 and a lower frequency at γ=0.

It makes sense that the Bayesian approach has a higher frequency of getting the correct answer closer to the extremes of γ because we have more information (we are more sure of which direction the wind is coming from). At $\gamma = 0.2$ for example, we are more sure that the wind is coming out of the east. Similarly, at $\gamma = 0.8$, we are most sure that the wind is coming out of the west, where if $\gamma = 0.5$, we are less sure where the wind is coming from. Mathematically, a γ closer to the extremes will output a smaller denominator and a higher probability value for Bayes approach, than if $\gamma = 0.5$.

In the frequentist approach, we see that even for a $\gamma = 1$ that the frequency of guessing the correct direction is not 100%, rather it is 95% due to our 95% confidence interval, such that we reject anything in the extreme 5% of the distribution. At $\gamma = 0$, we see a 50% probability of getting the correct answer due to the underlying distribution of the wind.

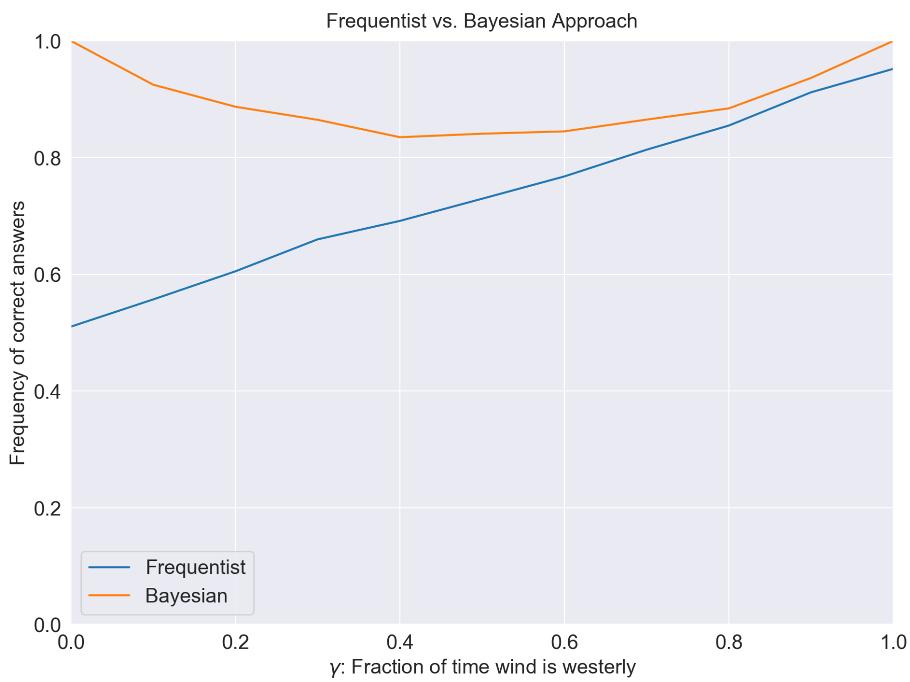

Figure 2: Plot for Problem 2c.