Homework 2

Due date: 09/21/2024

September 10, 2024

# Introduction

This exercise sheet contains a series of problems designed to test and enhance your understanding of the topics covered in the course. Please ensure that you attempt all problems and provide detailed solutions where necessary. If you have any questions or need clarification, feel free to reach out your TA.

# Exercises

**Exercise 1: Bootstrap Visualization** Navigate to this [web app](https://seeing-theory.brown.edu/frequentist-inference/index.html) and choose the bootstrap section.

1. Set the distribution to Normal, choose a sample size, and draw a sample. Resample once and describe what happens. **[5pts]**

When I resample once, the software takes the average and puts it in a histogram.

1. Hit the ”Resample 100 times” button a few times. Describe the resulting distribution. Describe the steps the web app is taking using the terms from class (hint: the word ”bootstrap” or ”bootstrapped” appears in all those terms). **[5pts]**

The app creates a bootstrap sample by sampling with replacement from the original sample. After, it calculates the bootstrap estimate by calculating the estimate of the sample. This is repeated several times resulting in the bootstrap estimated sampling distribution which describes the resulting distribution of the bootstrap estimates.

1. Select a different distribution (not Normal or Student’s t). Before you begin, should the shape of your resulting distribution match your initial distribution? What shape should it be? Repeatedly resample and describe the resulting distribution. Is the shape what you expected? **[5pts]**

For this example, I chose the exponential distribution. The shape of my resulting distribution should match my initial distribution. It should slope downwards. The initial distribution is more like a normal distribution than an exponential distribution. This was not the shape that I initially expected.

1. Adjust the sample size from your preferred distribution and repeat the procedure. Compare what happens with a small sample size (*n <* 8) to a larger sample size (*n >* 15). **[5pts]**

A smaller sample size will yield a wider spread than a larger sample size.

## Exercise 2: Estimated Sampling Distributions

For the following exercise, use the provided data on NBA 2023 season stats which includes the following variables:

* Player - Player Name (str): Name of the player.
* Team - Team Abbreviation (str): The team whose this player is playing for this season in abbreviated term.
* Age - Age (float): The Age of the player.
* GP - Games Played (float): The total games that the player has played in this season.
* W - Wins (float): Total of Games won while the player has played.
* L - Losses (float): Total of Games lost while the player has played.
* Min - Minutes Played (float): Total Minutes the player has played for this season.
* PTS - Points (float) : Total Points made by the player.

1. Let’s look the variable ’PTS’—the number of points players scored over the course of the season. Calculate the sample mean and standard deviation. Draw a histogram with the mean and mean plus or minus 1.96 standard deviations indicated by vertical lines. **[5pts]**

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated



A graph of a graph

Description automatically generated with medium confidence

1. Using your sample, calculate a 95% confidence interval for the population mean. Calibrate it using a bootstrap estimated sampling distribution. Create a histogram of your bootstrap estimated sampling distribution and annotate it with vertical lines to indicate the mean and the mean plus or minus 1.96 standard deviations. **[10pts]**

A screenshot of a computer code

Description automatically generated

A screenshot of a computer code

Description automatically generated

A screenshot of a computer program

Description automatically generated



A graph with a line in the middle

Description automatically generated with medium confidence

1. Calculate a 90% confidence interval and 99% confidence interval using your bootstrap estimated sampling distribution. Compare these intervals with your 95% CI. **[5pts]**

**A screenshot of a computer code

Description automatically generated**

1. Using the plug-in method, calculate a 95% confidence interval for the mean of points. How does this interval compare to your bootstrap estimated version? **[5pts]**

**A screenshot of a computer program

Description automatically generated**

In the plug-in version, the lower and upper bounds of the 95% CI are very similar to that of the bootstrap estimated version.

1. Repeat (a), (b), and (d) with the alternate sample. Compare the sample means, standard deviations, and 95% confidence intervals calibrated via bootstrap and plug-in method (note: you need to make a new estimated sampling distribution for your new sample! and your plug-in estimates should be different) **[15pts]**

This time, I made the sample sizes larger.

Original:

A number with numbers on it

Description automatically generated with medium confidence

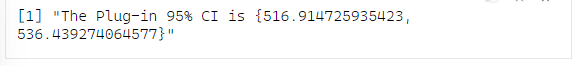
A close up of numbers

Description automatically generated

New Version:

A number on a white background

Description automatically generated



Here, we see all of the values get slightly smaller. Meaning that for all the values except the plug-in method, they are all getting closer to their true value in the population. For the plug-in method, the interval got slightly smaller, this could simply reflect the increase in sample size.

## Exercise 3: Comparing Sample and Population

Noting your interest in NBA stats, a benevolent data analyst sends you data for the full population. That is, they send you a dataset like the one you’ve been working with, but with a row for every player in the league the 2023 season. We can use the population to check our work in the previous problem.

1. Calculate the mean number of points scored in the population. Do any of the approximated intervals cover? Which ones? **[5pts]**

A white rectangular object with a black stripe

Description automatically generated

All of the bootstrap estimates and plug-in intervals cover the population mean.

1. Create a sampling distribution with 10,000 iterations to compare to your estimated sampling distribution (note: pause and think first about what your sample size should be). Make a histogram and compare it to your estimated sampling distribution. If you could use this sampling distribution to perfectly calibrate a 95% confidence interval, what width would it be? Compare this width to the intervals you estimated with your sample. **[10pts]**

A graph of a bar chart

Description automatically generated

This histogram has a more normal distribution compared to the estimated sampling distribution. The 95% confidence interval would be 77.84 in width around the mean which was found to be 523.73.

A number with numbers on it

Description automatically generated with medium confidence

The interval here is much wider than those my estimated sample.

## Exercise 4: Random Variables

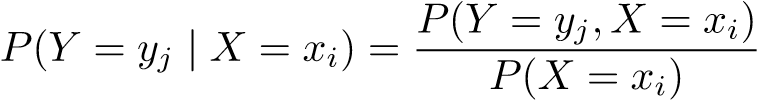
Your received a magical genie lamp for your birthday. It may contain three different types of genie: a golden genie that grants 3 wishes, a silver genie that grants 2 wishes or a bronze genie that grants just one wish.

1. Create a random variable *Y* for the number of wishes. Define the sample space and the output of the random variable *Y* . Is it a discrete or continuous? Why? **[5pts]**
2. The probability of having a golden, silver, and bronze genie are , , and  respectively. What is the expected value of *Y* . **[5pts]**
3. On friday, the number of wishes is squared! So, no matter what genie you get, you havethe usual number of wishes squared. Compute the expected value for the random variable describing the random of wishes on friday. **[5pts]**

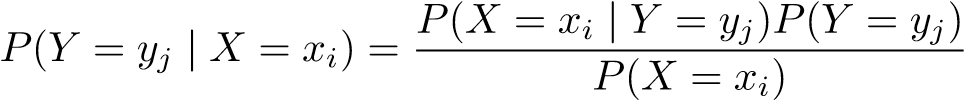
## Exercise 5: Bayes’ Rule

Given two random variables *X* and *Y* and being *xi* one of the *nX* elements in the sample space of the random variable *X*, and *yj* one of the *nY* elements in the sample space of the random variable *Y* , answer to the following questions:

(a) In class, we introduced the concept of conditional probability *P*(*Y* = *yj* | *X* = *xi*). Explain this concept in word and with a practical example. **[5pts]** (b) The Bayes’ rule states that:

 (1)

We can rewrite the Bayes’ rule in the following way:

 (2)

Explain in words the meaning of *P*(*Y* = *yj,X* = *xi*) and *P*(*X* = *xi*) and *P*(*Y* = *yj*).

Using Equation 1, show that Equation 2 is true and explain your steps.

(*Hint: start from Equation 1 and rewrite P*(*Y* = *yj,X* = *xj*) *using Equation 1.)* **[10pts]**

(c) Explain the not explained passages (\*1, \*2, \*3, \*4) of the proof of the Law of Iterated Expectations. **[10pts]**

*nX*

*E*[*E*[*Y* | *X*]] = X*E*[*Y* | *X* = *xi*]*P*(*X* = *xi*) \*1

*i*=1

*nX nY*

= XX*yjP*(*Y* = *yj* | *X* = *xi*)*P*(*X* = *xi*) \*2

*i*=1 *j*=1

*nX nY*

= XX*yjP*(*X* = *xi* | *Y* = *yj*)*P*(*Y* = *yj*) \*3 *Hint: use Equation 2.*

*i*=1 *j*=1

*nX nY*

= XX*yjP*(*Y* = *yj*)*P*(*X* = *xi* | *Y* = *yj*) *moving P*(*Y* = *yj*) *close to yj*

*i*=1 *j*=1

*nY nX*

= X*yjP*(*Y* = *yj*)X*P*(*X* = *xi* | *Y* = *yj*) *taking out the terms that do not involve yj*

*j*=1 *i*=1

*ny*

= X*yjP*(*Y* = *yj*) \*4

*j*=1

= *E*(*Y* )

# Submission Instructions

Please submit your completed exercises by **19 September** through **gradescope**. Ensure that your solutions are well-organized, clearly written, and include all necessary calculations and explanations. Questions about submission should be directed to your TA.

# Helping Resources

To better assist you in the completion of this exercise sheet, we suggest you to review the following material:

* **Lecture 2** - covering sampling distributions and perfectly calibrated confidence intervals (using population)
* **Lecture 3** - covering estimated sampling distributions, bootstrapping, and calibration of confidence intervals;
* **Lecture 4** - covering random variables, probability, expected values, and properties of expected values;
* **Lab** - practicing sampling, iterated sampling procedures, and probability