EE 381 - Probability and Statistic Computing

CONFIDENCE INTERVALS

Professor: Duc Tran

—

Students: Alyssa Faiferlick – Thanh Nguyen

—

Due date: 12/18/2020

<u>Question 1:</u> **The effect on sample size on the confidence intervals**

I. INTRODUCTION

A barrel of a million ball bearings (population size N = 1,000,000) where someone has actually weighed all one million of them and found the exact mean to be $\mu = 100$ grams, and the exact standard deviation to be $\sigma = 12$ grams.

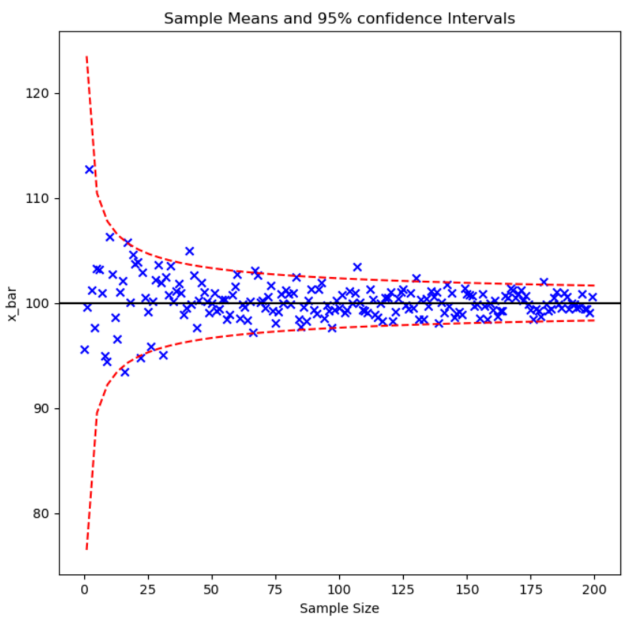
Choosing a sample of size n (n = 1, 2, ..., 200) of bearings from the barrel, weigh them and find the mean of the sample, and calculate the standard deviation of the sample. Then, plot the points $(n, \bar{X}_n)$ using a point marker, and the lower and upper bounds define the 95% and 99% confidence intervals to see how many of the sample means fall outside of the curves.

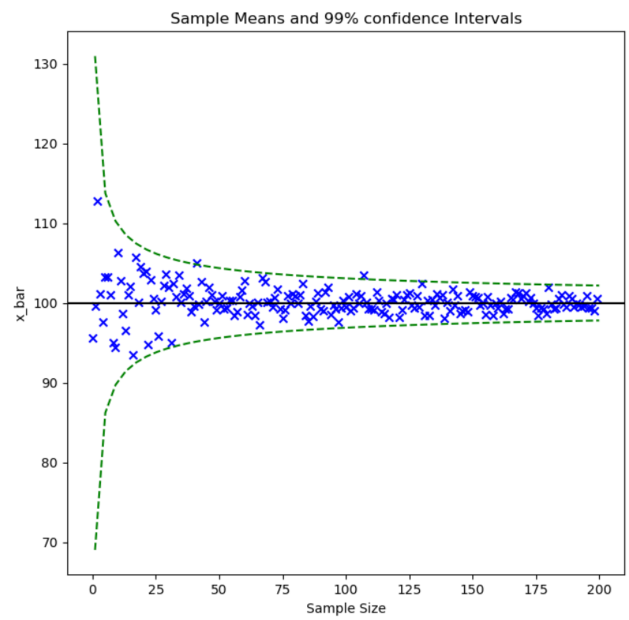
II. PROCEDURE:

- Generate a million ball bearings by using numpy.random.normal(mean, sigma, N) with mean = 100, sigma = 12, and N = 1,000,000
- For n from 1 to 200, pick a sample of size n of bearings from the barrel, weigh them and calculate the mean of each sample.
- Plot the points $(n, \bar{X}_n)$ using a point marker
- For each values of n (from 1 to 200), calculate and plot the values of $\mu \pm z_c \frac{\sigma}{\sqrt{n}}\ with\ z_c = 1.96\ for\ 95\%\ confidence\ interval\ and\ z_c = 2.58\ for\ 99\%\ confidence\ interval.$

III. RESULTS AND DISCUSSION:

- Results:





- The results that what we expected. In the first picture, approximately 95% of sample means are fallen within the two red curves. Approximately 99% of the sample means are fallen within the two green curves at 99% confidence level.

IV. CONCLUSION:

- We can see that the means of samples are close to the mean $\mu = 100$ when n is larger in two figures above.
- We used the information that professor gave, and had no problem when solving this question.

V. APPENDIX:

```
import numpy as np
import matplotlib.pyplot as plt
import random

N = 1000000; # Bearings
mean = 100;
sigma = 12;
M = 10000;
nSample = 200;

# Generating population
bearings = np.random.normal(mean, sigma, N);

# Z value for confidence interval
z_95 = 1.96;
z_99 = 2.58;

result = [];
for i in range(0, nSample):
```

```
        n = i + 1;
        sample = np.random.choice(bearings, i + 1);
        result.append(sum(sample) / n);

    x = np.linspace(1, nSample);

    # Plotting Figure a
    fig1 = plt.figure(1);
    plt.title("Sample Means and 95% confidence Intervals");
    plt.xlabel("Sample Size");
    plt.ylabel("x_bar");
    plt.axhline(y = mean, color = "black");
    plt.plot(x, mean + z_95 * sigma / (x**(1/2)), 'r--');
    plt.plot(x, mean - z_95 * sigma / (x**(1/2)), 'r--');
    plt.scatter(range(nSample), result, marker = 'x', color = "blue");
    fig1.show();

    #Plotting Figure b
    fig2 = plt.figure(2);
    plt.title("Sample Means and 99% confidence Intervals");
    plt.xlabel("Sample Size");
    plt.ylabel("x_bar");
    plt.axhline(y = mean, color = "black");
    plt.plot(x, mean + z_99 * sigma / (x**(1/2)), 'g--');
    plt.plot(x, mean - z_99 * sigma / (x**(1/2)), 'g--');
    plt.scatter(range(nSample), result, marker = 'x', color = "blue");
    fig2.show();
```

Question 2: The difference between Normal Distribution and Student's t distribution for different sample sizes

I. INTRODUCTION

From a barrel of a million ball bearings in the problem above, choosing random samples of size n= 5, 40, 120. Create the 95% and 99% confidence intervals using normal distribution and t-distribution. Checking if the confidence intervals include the actual means of the population, then set this experiment is success. Calculating the percentage of successful outcomes to see the difference between normal distribution and t distribution on the samples of different sizes.

II. PROCEDURE:

- For each n, run the for loop from 1 to 10,000. For each experiment:
  - Step 1: Choose a random sample of size n (n = 5, 40, 120) bearings from the N bearings was created in the previous problem.
  - Step 2: Calculate the sample mean and the sample standard deviation of the sample by using these formulas:

$$\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j \text{ and } \hat{S} = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \bar{X}\right)^2}$$

  - Step 3: Create the 95% and 99% confidence interval using the normal distribution: $z_c = 1.96\ for\ 95\%\ level, and\ z_c = 2.58\ for\ 99\%\ level$)

$$\left[\mu_{lower}, \mu_{upper}\right] = \left[\bar{X} - z_c\frac{\hat{S}}{\sqrt{n}}, \bar{X} + z_c\frac{\hat{S}}{\sqrt{n}}\right]$$

  - Step 4: Check if the confidence interval includes the actual mean μ of the population of N bearings, then decide it as a success, and increase the counter by 1.
  - Step 5: Create the 95% and 99% confidence interval using the t-distribution: $At\ 95\%\ level\ t_c = 2.78, 2.02, 1.98, and\ at\ 99\%\ level\ t_c = 4.6, 2.7, 2.62\ for\ n = 5, 40, 120, respectively$

$$[\mu_{lower},\mu_{upper}] = \left[\bar{X} - t_c\frac{\hat{S}}{\sqrt{n}}, \bar{X} + t_c\frac{\hat{S}}{\sqrt{n}}\right]$$

- Step 6: Check if the confidence interval includes the actual mean μ of the population, then decide it as a success, and increase the counter by 1.

- Calculate the percentage of successful outcomes.

III. RESULTS AND DISCUSSION:

- Results:

```
Normal Distribution with n =  5

At 95% confidence level:  87.96000000000001

At 99% confidence level:  93.94

Student-T's Distribution with n =  5

At 95% confidence level:  95.06

At 99% confidence level:  99.03

Normal Distribution with n =  40

At 95% confidence level:  94.39999999999999

At 99% confidence level:  98.77

Student-T's Distribution with n =  40

At 95% confidence level:  95.00999999999999

At 99% confidence level:  99.08

Normal Distribution with n =  120

At 95% confidence level:  94.66

At 99% confidence level:  98.83999999999999

Student-T's Distribution with n =  120

At 95% confidence level:  94.98

At 99% confidence level:  98.96000000000001
```

Sample size (n)	95% Level (Normal Distribution)	99% Level (Normal Distribution)	95% Level (Student's t distribution)	99% Level (Student's t distribution)
5	87.96	93.94	95.06	99.03
40	94.4	98.77	95.01	99.08
120	94.66	98.84	94.98	98.96

IV. CONCLUSION:

- As the results we received in this question, for a large sample size (n > 30) the Student's T distribution is very close to the normal distribution.
- The instructions that professor gave us are easy to understand and very helpful. We just plugged in the different value for n to find the of t_c

V. APPENDIX:

```
import numpy as np

N = 1000000; # Bearings
mean = 100;
sigma = 12;
M = 10000;

# Generating population
bearings = np.random.normal(mean, sigma, N);

def ConfidenceInterval(bearings, mean, sigma, M, n):
    successes_n95 = 0;
    successes_n99 = 0;
    successes_t95 = 0;
    successes_t99 = 0;

    # Giving z value
    z_95 = 1.96;
    z_99 = 2.58;

    # Giving t value
    if (n == 5):
        t_95 = 2.78;
        t_99 = 4.6;
    if (n == 40):
        t_95 = 2.02;
        t_99 = 2.7;
    if (n == 120):
        t_95 = 1.98;
        t_99 = 2.62;

    for i in range(M):
        # Generating sample of size n from population
        sample = np.random.choice(bearings, n);
        mean_s = sum(sample) / n;
```

```
            sigma_s = 0;
            for j in range(len(sample)):
                sigma_s = sigma_s + ((sample[j] - mean_s)**2);
            sigma_s = (sigma_s / (n - 1))**(1/2);

            # Finding upper and lower limits for Normal Distribution
            lowerLimit_n95 = mean_s - z_95 * sigma_s / (n**(1/2));
            upperLimit_n95 = mean_s + z_95 * sigma_s / (n**(1/2));

            lowerLimit_n99 = mean_s - z_99 * sigma_s / (n**(1/2));
            upperLimit_n99 = mean_s + z_99 * sigma_s / (n**(1/2));

            if (lowerLimit_n95 <= mean and mean <= upperLimit_n95):
                successes_n95 += 1;
            if (lowerLimit_n99 <= mean and mean <= upperLimit_n99):
                successes_n99 += 1;

            # Finding upper and lower limits for T-SDistribution
            lowerLimit_t95 = mean_s - t_95 * sigma_s / (n**(1/2));
            upperLimit_t95 = mean_s + t_95 * sigma_s / (n**(1/2));

            lowerLimit_t99 = mean_s - t_99 * sigma_s / (n**(1/2));
            upperLimit_t99 = mean_s + t_99 * sigma_s / (n**(1/2));

            if (lowerLimit_t95 <= mean and mean <= upperLimit_t95):
                successes_t95 += 1;
            if (lowerLimit_t99 <= mean and mean <= upperLimit_t99):
                successes_t99 += 1;

        # Calculating the success for 95 and 99 confidence level
        successes_n95 = successes_n95 / M * 100;
        successes_n99 = successes_n99 / M * 100;
        successes_t95 = successes_t95 / M * 100;
        successes_t99 = successes_t99 / M * 100;

        # Print results
        print("Normal Distribution with n = ", n, "\n");
        print("At 95% confidence level: ", successes_n95, "\n");
        print("At 99% confidence level: ", successes_n99, "\n");

        print("Student-T's Distribution with n = ", n, "\n");
        print("At 95% confidence level: ", successes_t95, "\n");
        print("At 99% confidence level: ", successes_t99, "\n");

    ConfidenceInterval(bearings, mean, sigma, M, 5);
    ConfidenceInterval(bearings, mean, sigma, M, 40);
    ConfidenceInterval(bearings, mean, sigma, M, 120);
```