Exercice n°4

Sample distribution and Central Limit Theorem

Grégoire de Lambertye

2022-10-27

I Task n°1 :

We consider the 12 sample data points: 4.94 5.06 4.53 5.07 4.99 5.16 4.38 4.43 4.93 4.72 4.92 4.96

I.1 How many possible bootstrap samples are there, if each bootstrap sample has the same size as the original?

Two bootstrap sample are considered equals if only the order of the point change from one bootstrap sample to another. With 12 point we can have up to XX different bootstrap samples: $b = n - 1 * n - 1$

I.2 Compute the mean and the median of the original sample.

```
data <- c(4.94, 5.06, 4.53, 5.07, 4.99, 5.16, 4.38, 4.43, 4.93, 4.72, 4.92, 4.96)

original_mean <- mean(data)
original_median <- median(data)
```

The mean of the original sample is 4.8408333 and the median is 4.935

I.3 Create 2000 bootstrap samples and compute their means.

We create 2000 samples with the same length than the original sample.

```
set.seed(12202211)
m <- 2000
samples <- replicate(m, sample(data, size=12, replace=TRUE))
means <- numeric(m)
for(i in 0:m){
  means[i] <- mean(samples[,i])
}
```

Compute the mean on the first 20.0.0 bootstrap means.

```r
mean_of_mean <- function(n){
  sum <- 0
  for(i in 1:n){
    sum <- sum + mean(samples[,i])
  }
  return(sum/n)
}
mean_20b <- mean_of_mean(20)
mean_200b <- mean_of_mean(200)
mean_2000b <- mean_of_mean(2000)
means_first_sample <- c(mean_20b, mean_200b, mean_2000b)
```

We have the following result :

number of sample	20	200	2000
mean	4.8549167	4.8411458	4.8378087

Visualise the distribution all the different bootstrap means to the sample mean. Does the Central Limit Theorem kick in?

Based on the three different bootstrap sample lengths in 3. compute the corresponding 0.025 and 0.975 quantiles. Compare the three resulting intervals against each other and the "true" confidence interval of the mean under the assumption of normality. (Use for example the function t.test to obtain the 95% percent CI based on asympotic considerations for the mean.)

I.4 Create 2000 bootstrap samples and compute their medians.

Compute the mean on the first 20 bootstrap medians. Compute the mean of the first 200 bootstrap medians. Compute the mean based on all 2000 bootstrap medians. Visualise the distribution all the different bootstrap medians to the sample median. Based on the three different bootstrap sample lengths in 3. compute the corresponding 0.025 and 0.975 quantiles. Compare the three resulting intervals against each other.

II Task n°2 :

III Task n°3 :

IV Task n°4 :