

Sample distribution and Central Limit Theorem

Kirill Medovshchikov

2022-11-14

Contents

1	Setup	3
2	Task 1	3
2.1	2.	3
2.2	3.	3
2.3	4.	6
3	Task 2	8
3.1	1.	8
3.2	2.	9

1 Setup

```
set.seed(12144024)
```

2 Task 1

Here the provided data points have been added into the list.

```
dp <- 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)
```

2.1 2.

Here the mean and median of the previously defined sample are calculated and then displayed.

```
mean_dp <- mean(dp)
median_dp <- median(dp)
mean_dp
```

```
## [1] 4.840833
```

```
median_dp
```

```
## [1] 4.935
```

2.2 3.

After that the 2000 bootstrap samples are generated.

```
samples_amount <- 2000
samples <- replicate(samples_amount, sample(dp, replace = TRUE))
```

Then the means are calculated for every sample element.

```
sample_means = c()
count = 1
for (i in samples){
  sample_means[count] = mean(i)
  count = count + 1
}
```

2.2.1 1.

First the means of the 20 bootstrap means are found.

```
mean_dp_20 <- mean(sample_means[1:20])  
mean_dp_20
```

```
## [1] 4.8965
```

2.2.2 2.

Then the means of 200 bootstrap means.

```
mean_dp_200 <- mean(sample_means[1:200])  
mean_dp_200
```

```
## [1] 4.8492
```

2.2.3 3.

Then the means of 2000 bootstrap means.

```
mean_dp_2000 <- mean(sample_means[1:2000])  
mean_dp_2000
```

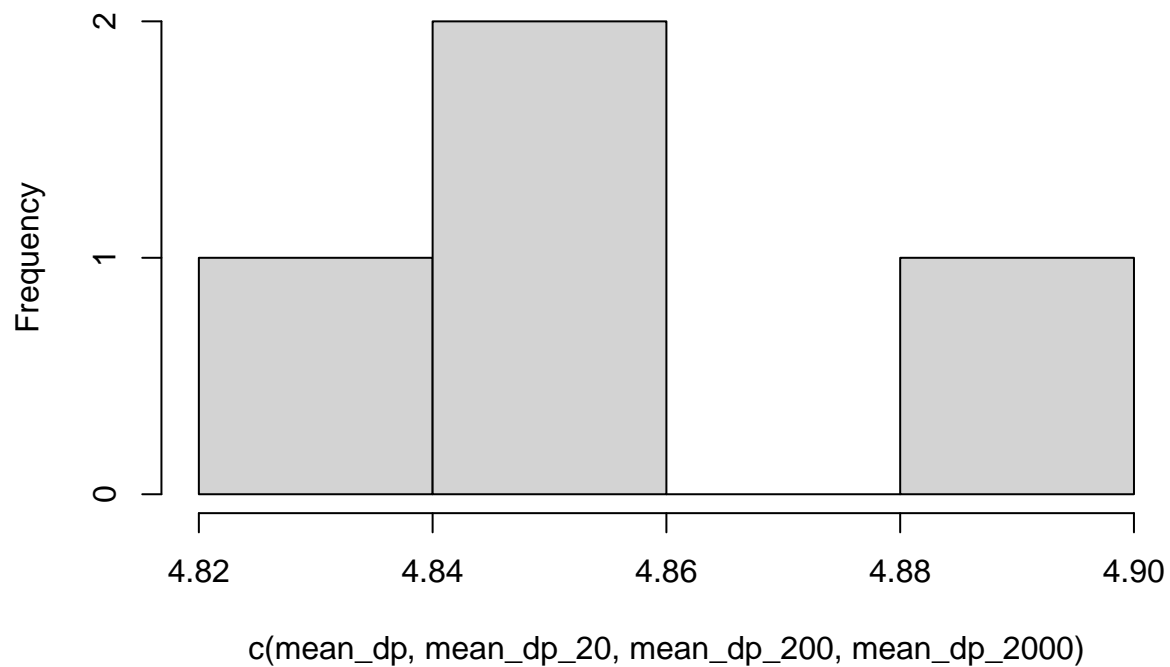
```
## [1] 4.83917
```

2.2.4 4.

Here all of the means are visualized against each other.

```
hist(c(mean_dp, mean_dp_20, mean_dp_200, mean_dp_2000))
```

Histogram of c(mean_dp, mean_dp_20, mean_dp_200, mean_dp_2000)



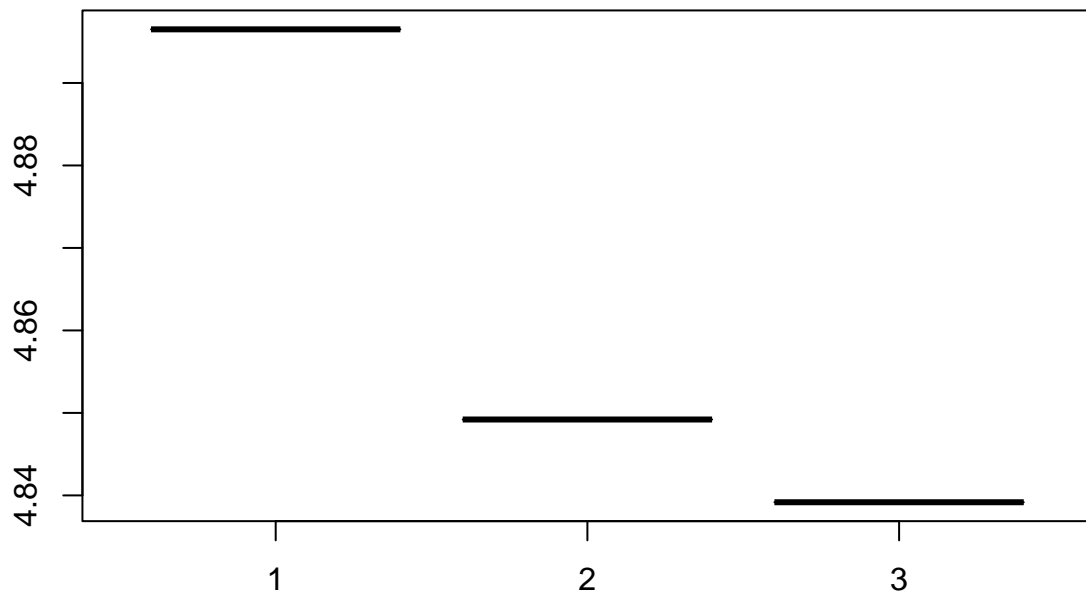
2.2.5 5.

Now the quantiles of every mean subset is calculated.

```
q1 <- quantile(mean_dp_20)
q2 <- quantile(mean_dp_200)
q3 <- quantile(mean_dp_2000)
```

And then are displayed through the use of a boxplot.

```
boxplot(q1, q2, q3)
```



2.3 4.

Now the same process as in the previous subsection is performed once again for the previously generated samples. However this time the median is calculated.

```
sample_medians = c()
count = 1
for (i in samples){
  sample_medians[count] = median(i)
  count = count + 1
}
```

2.3.1 1.

Here the median of 20 bootstrap medians are calculated.

```
median_dp_20 <- median(sample_medians[1:20])
median_dp_20
```

```
## [1] 4.945
```

2.3.2 2.

Then the median of 200 bootstrap medians are calculated.

```
median_dp_200 <- median(sample_medians[1:200])
median_dp_200
```

```
## [1] 4.94
```

2.3.3 3.

Here the median of 2000 bootstrap medians are calculated.

```
median_dp_2000 <- median(sample_medians[1:2000])
median_dp_2000
```

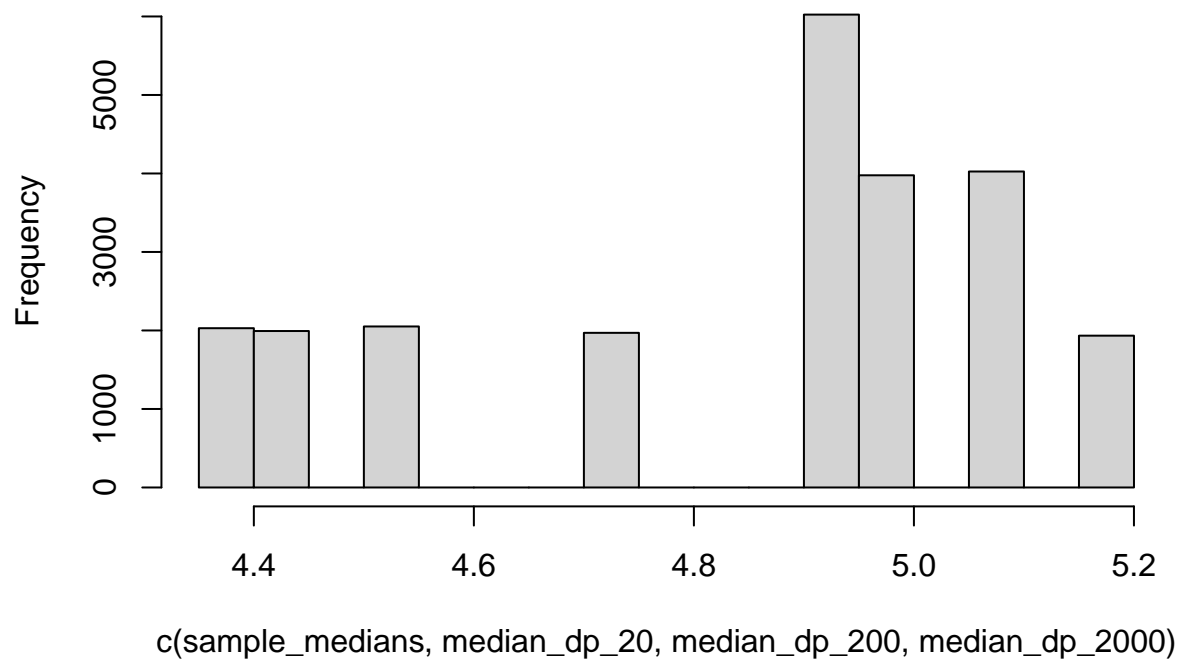
```
## [1] 4.94
```

2.3.4 4.

Here once again the different medians are compared against each other.

```
hist(c(sample_medians, median_dp_20, median_dp_200, median_dp_2000))
```

rogram of c(sample_medians, median_dp_20, median_dp_200, median_dp_2000)



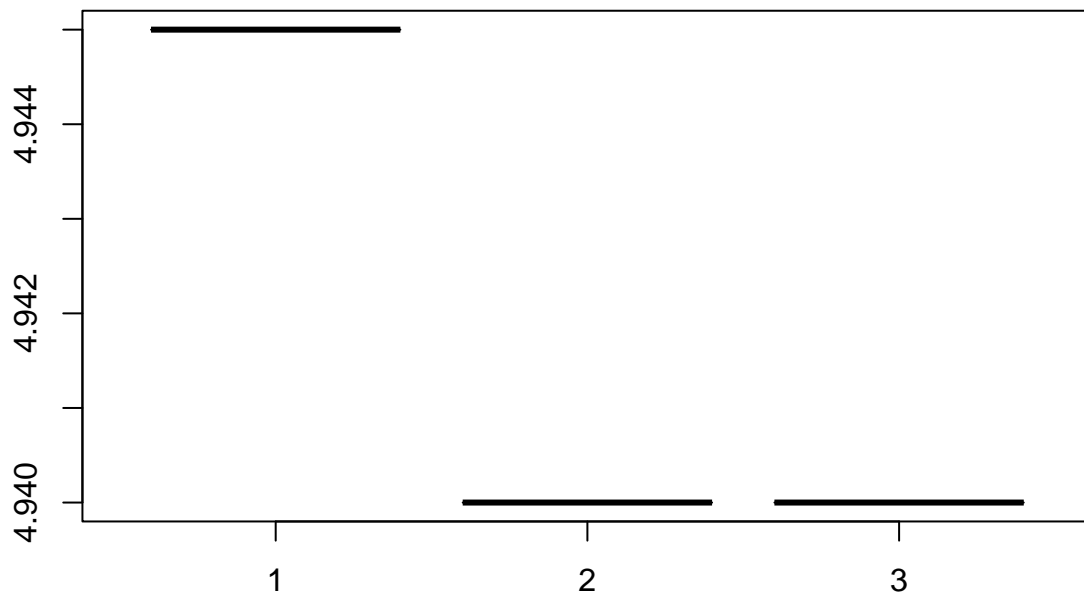
2.3.5 5.

Now the quantiles are calculated for the medians.

```
qm1 <- quantile(median_dp_20)
qm2 <- quantile(median_dp_200)
qm3 <- quantile(median_dp_2000)
```

And the boxplot is built.

```
boxplot(qm1, qm2, qm3)
```



3 Task 2

3.1 1.

Here the seed is set to 1234 and the standard normal distribution is calculated for 1960 points. Then the vector with 4 0observation is created from a uniform random variables. Finally x variable is created.

```
set.seed(1234)
x.clean <- rnorm(1960)
x.cont <- runif(40,4,5)
x <- c(x.clean, x.cont)
set.seed(12144024)
```


3.2 2.

Here the mean, median and mean with alpha are calculated for x.

```
x_mean <- mean(x)
x_median <- median(x)
x_mean_trim <- mean(x, trim = 0.05)
```

And the same operations are done for x.clean variable.

```
x_mean <- mean(x.clean)
x_median <- median(x.clean)
x_mean_trim <- mean(x.clean, trim = 0.05)
```