Statistical Inference Part 1

Overview

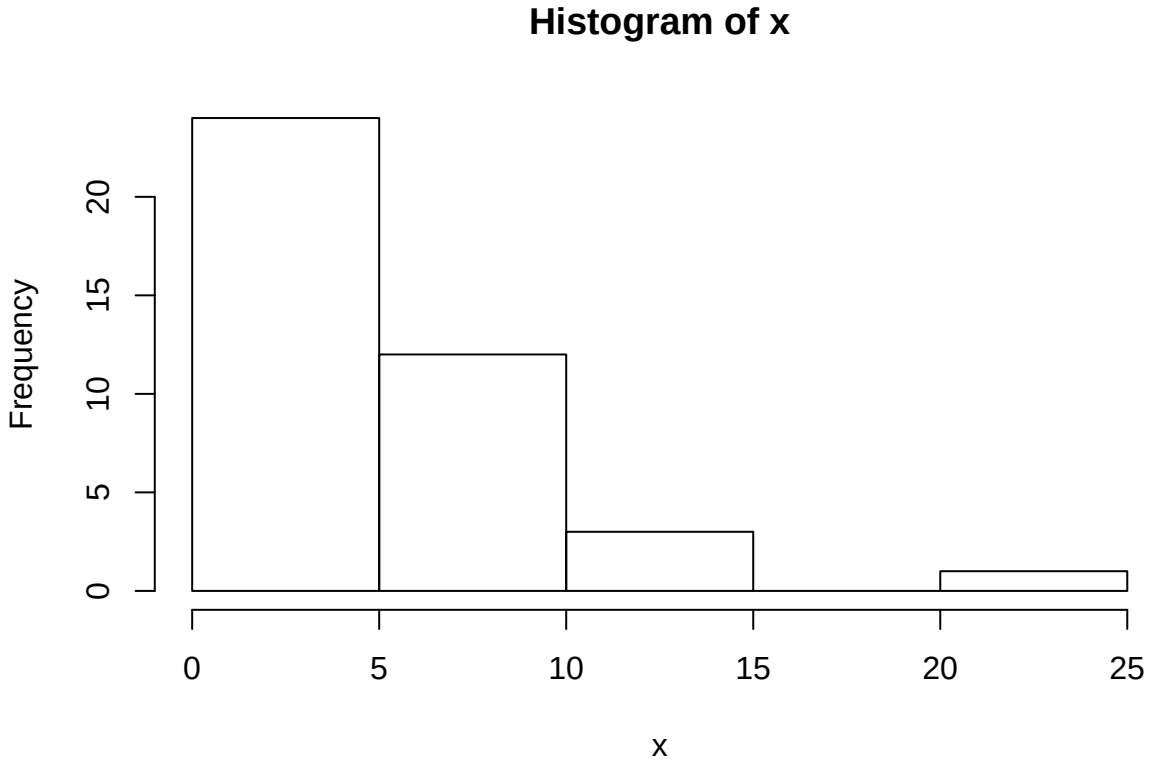
In this project we are doing a simulation exercise.

We will be investigating the exponential distribution in R and comparing it with the Central Limit Theorem by analysing the distribution of averages of 40 exponentials, and doing 1,000 simulations.

Simulations

The code below simulates 40 exponentials x happening at rate lambda = 0.2:

```
set.seed(123)
x <- rexp(40, 0.2)
hist(x)
```



To carry out the bootstrap principle, we will sample 40 (n) obvervations with replacement from the observed data 1,000 times (B), and take the mean of each data set.

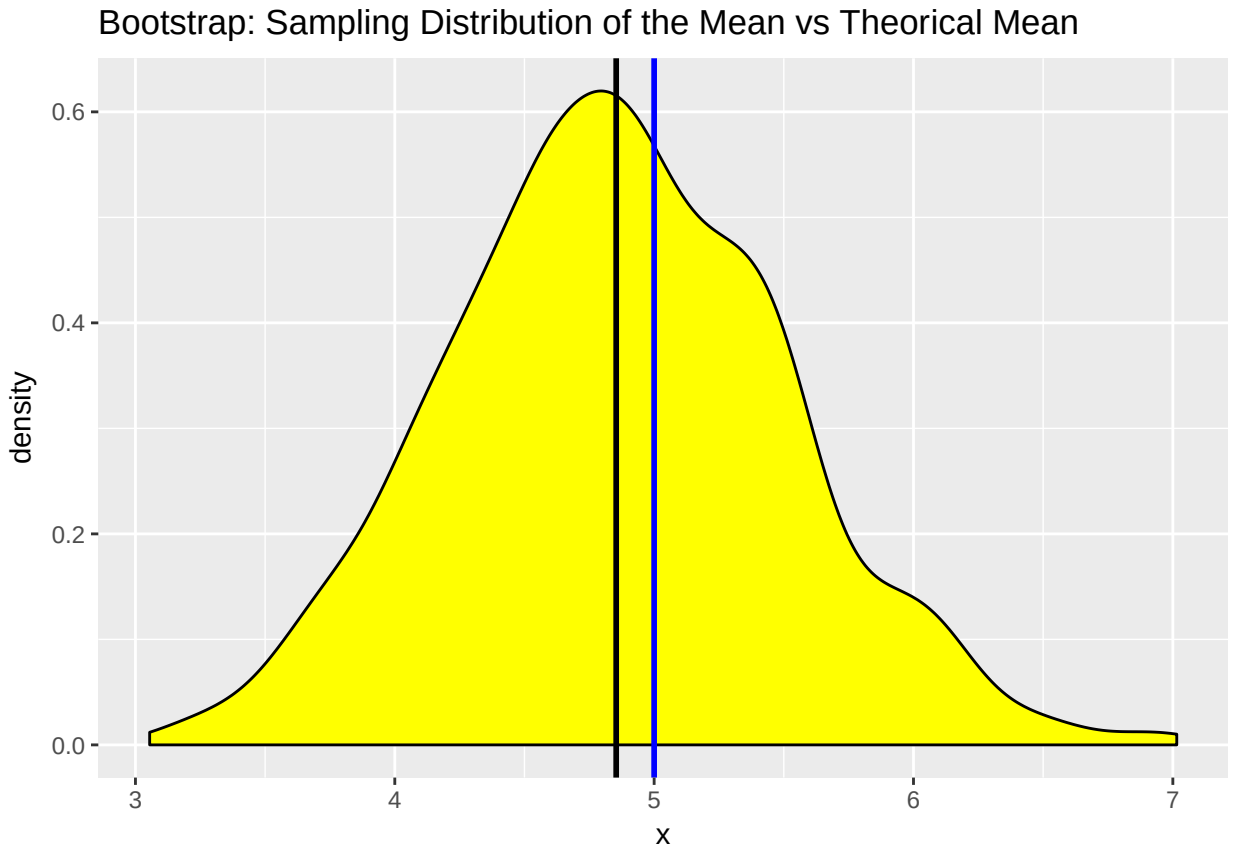
```
B <- 1000
n <- length(x)
resamples <- matrix(sample(x, n*B, replace=TRUE), B, n)
means <- apply(resamples, 1, mean)
```

This is approximately like drawing from the sampling distribution of that statistic.

Sample Mean versus Theoretical Mean

This is the sampling distribution of the mean using the bootstrap (black line), compared to the theorerical mean of an exponential distribution 1/lambda (blue line).

```
library(ggplot2)
g <- ggplot(data.frame(x = means), aes(x = x))
g <- g + geom_density( fill = "yellow")
g <- g + geom_vline(xintercept = 1/0.2, size = 1, colour = "blue")
g <- g + geom_vline(xintercept = mean(means), size = 1, colour = "black")
g <- g + ggtitle("Bootstrap: Sampling Distribution of the Mean vs Theorical Mean")
g
```



Sample mean:

```
sample_mean <- mean(means)
sample_mean
```

```
## [1] 4.853696
```

Theorical mean:

```
theoretical_mean <- 1/0.2
theoretical_mean
```

```
## [1] 5
```

Sample Variance versus Theoretical Variance

Our sample mean has variance 0.42, compared to our theorical mean (1 / lambda)^2 / (n) = 0.62.

```
sample_var<- var(means)
sample_var
```

```
## [1] 0.4234653
```

```
theoretical_var  <- (1 / 0.2)^2 /40
theoretical_var
```

```
## [1] 0.625
```

Our sample mean has a standard deviation of 0.65, compared to our theorical standard devatino of 0.79:

```
sample_sd <- sd(means)
sample_sd
```

```
## [1] 0.6507421
```

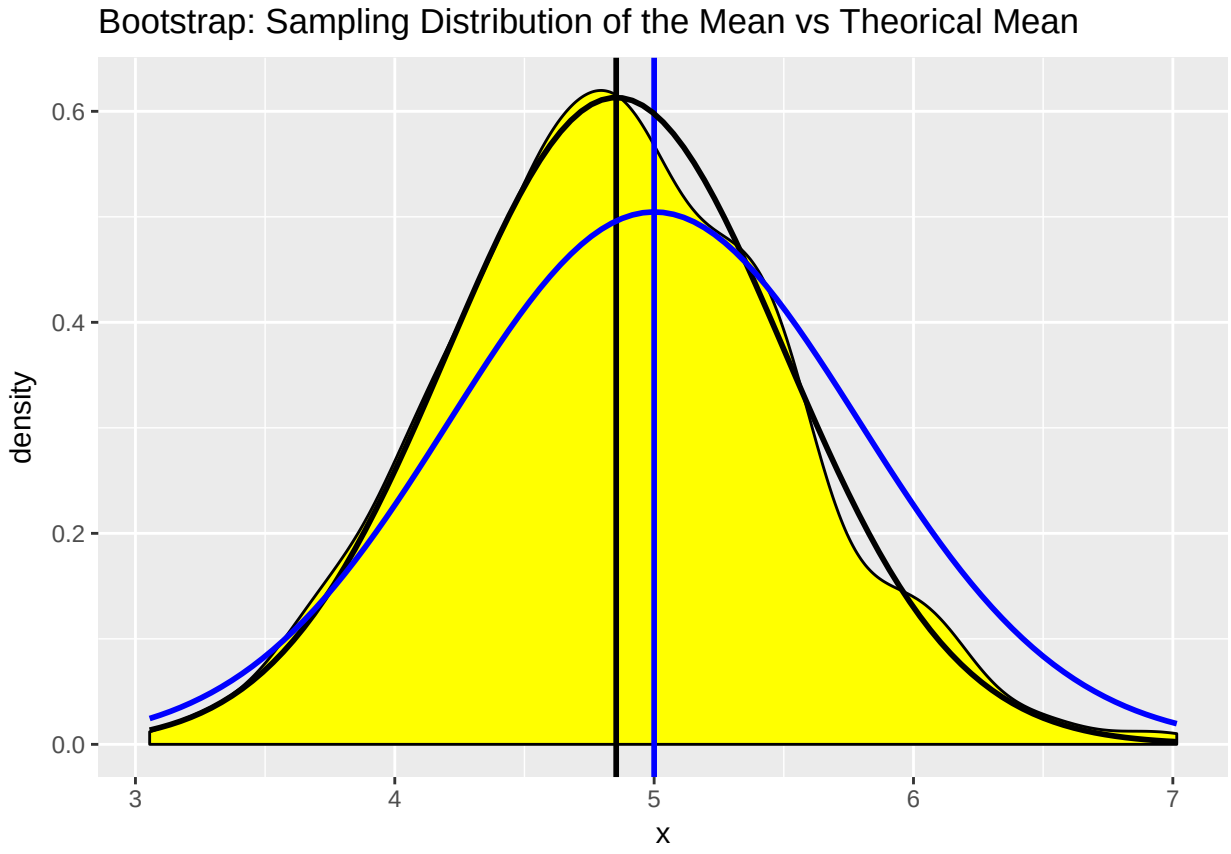
```
theoretical_sd  <- 1 / (0.2 * sqrt(40))
theoretical_sd
```

```
## [1] 0.7905694
```

Distribution

The distribution of sample means (black) is close to the distribution of the theorical mean (blue) given the value of lambda = 0.2.

```
m <- g + stat_function(fun = dnorm, args = list(mean = mean(means), sd = sd(means)), color = "black", si
m <- m + stat_function(fun = dnorm, args = list(mean = 1/0.2, sd = 1 / (0.2 * sqrt(40))), colour = "blue
m
```


Bootstrap: Sampling Distribution of the Mean vs Theorical Mean
density
0.0
0.2
0.4
0.6
3
4
5
6
7
x