Assignment

Jose V. Die

February 26, 2016

Sample Project Report

Overview

In this project we will investigate the exponential distribution and compare it with the Central Limit Theorem. We will estimate a statistic (mean) from our observations and get an idea about its distribution. We will investigate the distribution of averages of 40 exponentials. For this purpose, we will run 1,000 simulations.

- Question 1. Show the sample mean and compare it to the theoretical mean of the distribution.

1.1 We will create a simulated data set by sampling n observations with replacement from the observed data

```
# Set variables
set.seed(2340)

n <-  40
B <- 1000
lamb <- 0.2

# Run simulation
sim.data = NULL
for(i in 1:1000) {
    sim.data = c(sim.data,mean(rexp(n,lamb)))
}
```

1.2 We will take the sample mean

```
resampledMeans <- mean(sim.data)
```

1.3 We will compare the estimated distribution of means with the theoretical mean.

```
theor <- 1/lamb
empirical <- resampledMeans
diff <- round(abs(empirical - theor),3)
```

The difference between the estimated- and the theoretical mean of the distribution is **0.031**. So, we see how close they are.

- Question 2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

The variance of the estimated mean from our simulation is:

```r
var(sim.data)
```

```
## [1] 0.6062894
```

Again, our estimate is pretty close to the the theoretical variance of the distribution:

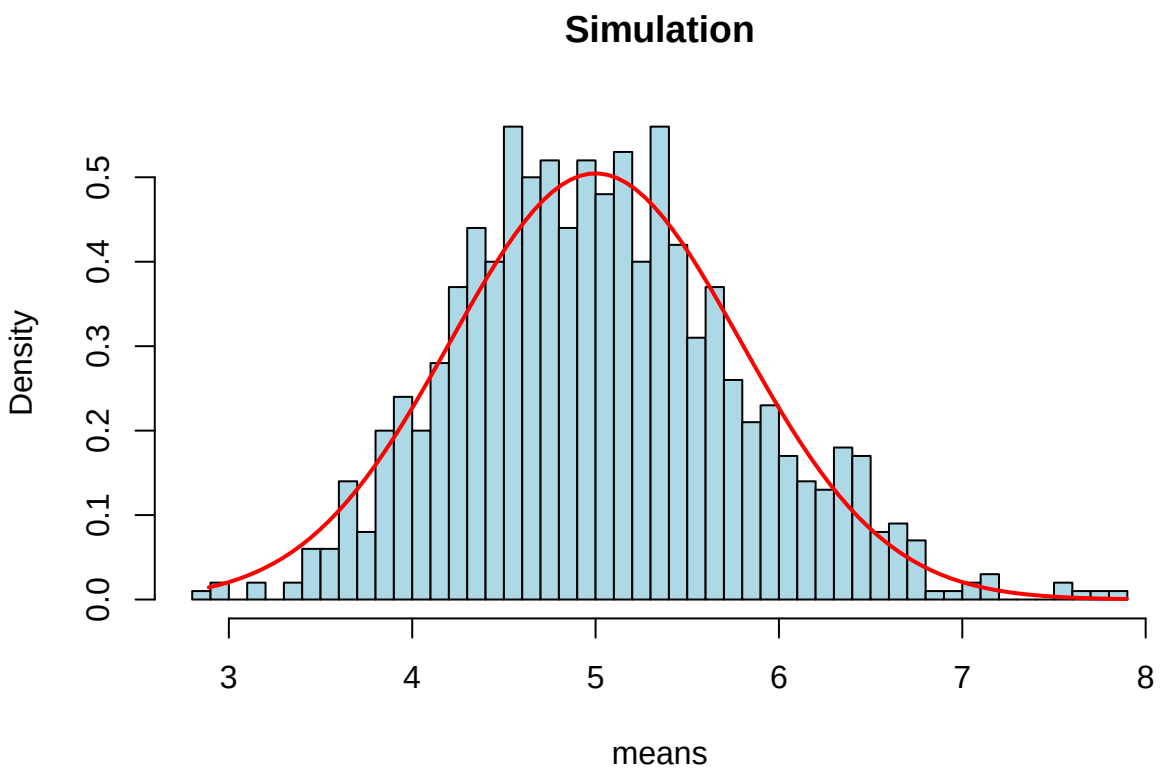
```r
theor.var <- (1/(lamb*sqrt(n)))**2
theor.var
```

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

- Question 3. Show that the distribution is approximately normal
  We can compare the distribution of our simulation data (large collection of averages of 40 exponentials) with the theoretical normal distribution.

```r
#Simulated data
hist(sim.data, breaks=n, probability = TRUE, col="lightblue",
    xlab="means", main="Simulation")

#Normal distribution
x = seq(min(sim.data), max(sim.data), length=100)
lines(x, dnorm(x, mean=1/lamb, sd=(1/(lamb*sqrt(n)))), col=2, lwd=2)
```



Also, we could compare the quantiles of the simulated data with the normal distribution. We can see how the distribution is approximately normal

```r
qqnorm(sim.data)
qqline(sim.data, col=2, lwd=2)
```