Statistical Inference Project Part 1

Paulo Cardoso

January 28, 2016

Overview

This first part of Statistical Inference class project that aims to simulate an exponential distribution and perform comparisons with the central limit theorem and is composed of four stages: - Simulations - Sample Mean versus Theoretical Mean - Sample Variance versus Theoretical Variance - Distribution The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of the exponential distribution is 1/lambda and the standard deviation is also 1/lambda. For this simulation, we set lambda=0.2. And we investigate the distribution of averages of 40 samples drawn from the exponential distribution with lambda=0.2.

Simulations

Given the requirements of the exercises were held 1000 simulations using a distribution of averages of 40 and a lambda value of 0.2, as presented by de code below.

```
# Set seed
set.seed(456)
# Set lambda
lambda <- 0.2
# 1000 simulations with 40 samples
sampleSize <- 40
nSim <- 1000

# Performing the 1000 simulations
simExp <- matrix(rexp(nSim*sampleSize, rate=lambda), nSim, sampleSize)
# Averages of 40 exponentials
rMean <- rowMeans(simExp)
```

The vector rMean contains the mean value of 40 samples. And below you can observe its summary.

```
# Presenting Summary
summary(rMean)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.054   4.459   4.935   4.984   5.506   8.244
```

Sample Mean versus Theoretical Mean

Presentation of the distribtion of the sample mean and the theoretical mean.

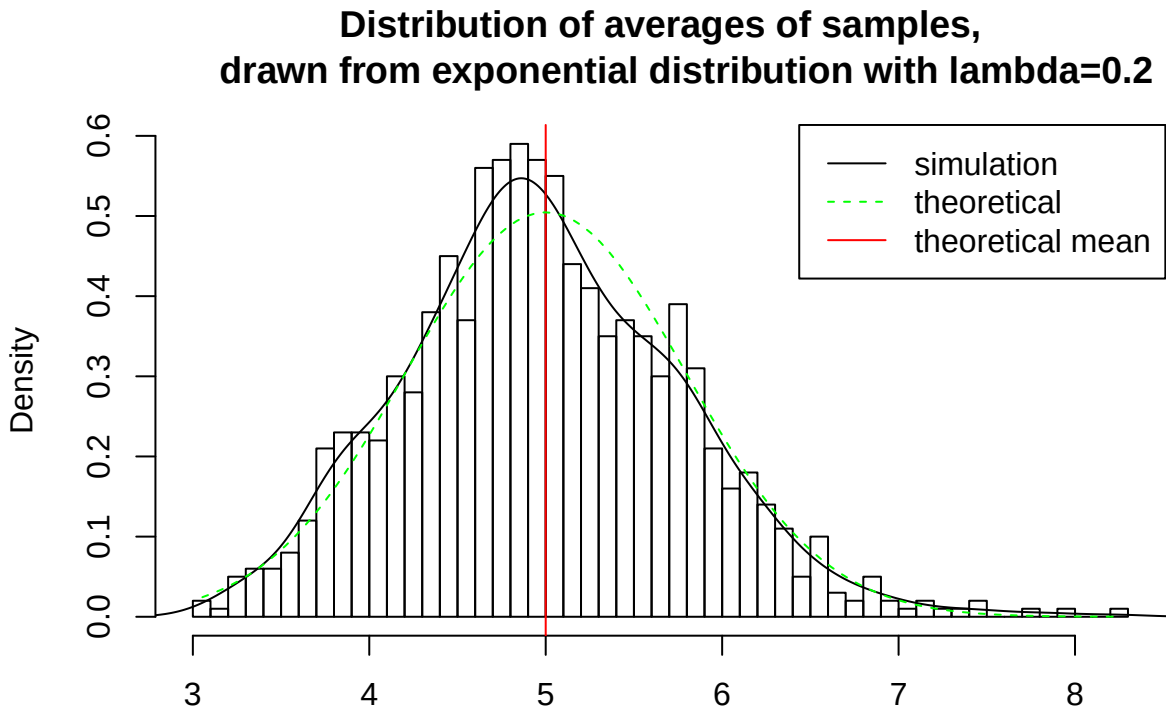
```
# Sample mean
sampleMean <- mean(rMean)
sampleMean
```

```
## [1] 4.984081
```

```
# Theoretical mean
tMean <- 1/lambda
tMean
```

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

```
# Plot histogram of averages
hist(rMean, breaks=50, prob=TRUE,
     main="Distribution of averages of samples,
     drawn from exponential distribution with lambda=0.2",
     xlab="")
# Density of averages of the samples
lines(density(rMean))
# Theoretical center of distribution
abline(v=1/lambda, col="red")
# Theoretical density of the averages of samples
xfit <- seq(min(rMean), max(rMean), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(sampleSize)))
lines(xfit, yfit, pch=22, col="green", lty=2)
# Adding legend
legend('topright', c("simulation", "theoretical", "theoretical mean"), lty=c(1,2,1), col=c("black", "gr
```



Sample Variance versus Theoretical Variance

This is the comparison of theoretical variance with the sample variance and the theoretical standard error with the sample standard error.

```
# Theoretical variance
tVar <- (1/lambda)^2/sampleSize
round(tVar,3)
```

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

```
# Sample variance
round(var(rMean), 3)
```

```
## [1] 0.609
```

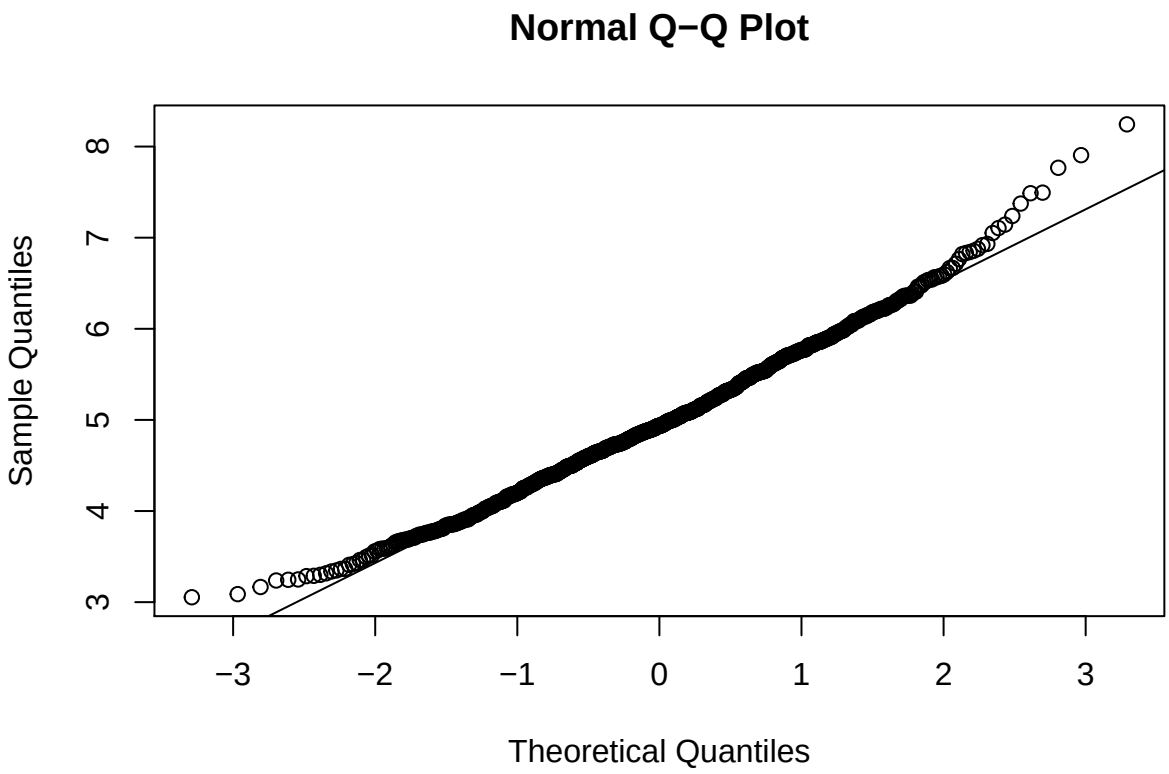
```
# Theoretical standard error
tse <- 1/(lambda*sqrt(sampleSize))
round(tse, 3)
```

```
## [1] 0.791
```

```
# Sample standard error
round(sd(rMean), 3)
```

```
## [1] 0.781
```

```
# qqplot visualization
qqnorm(rMean); qqline(rMean)
```



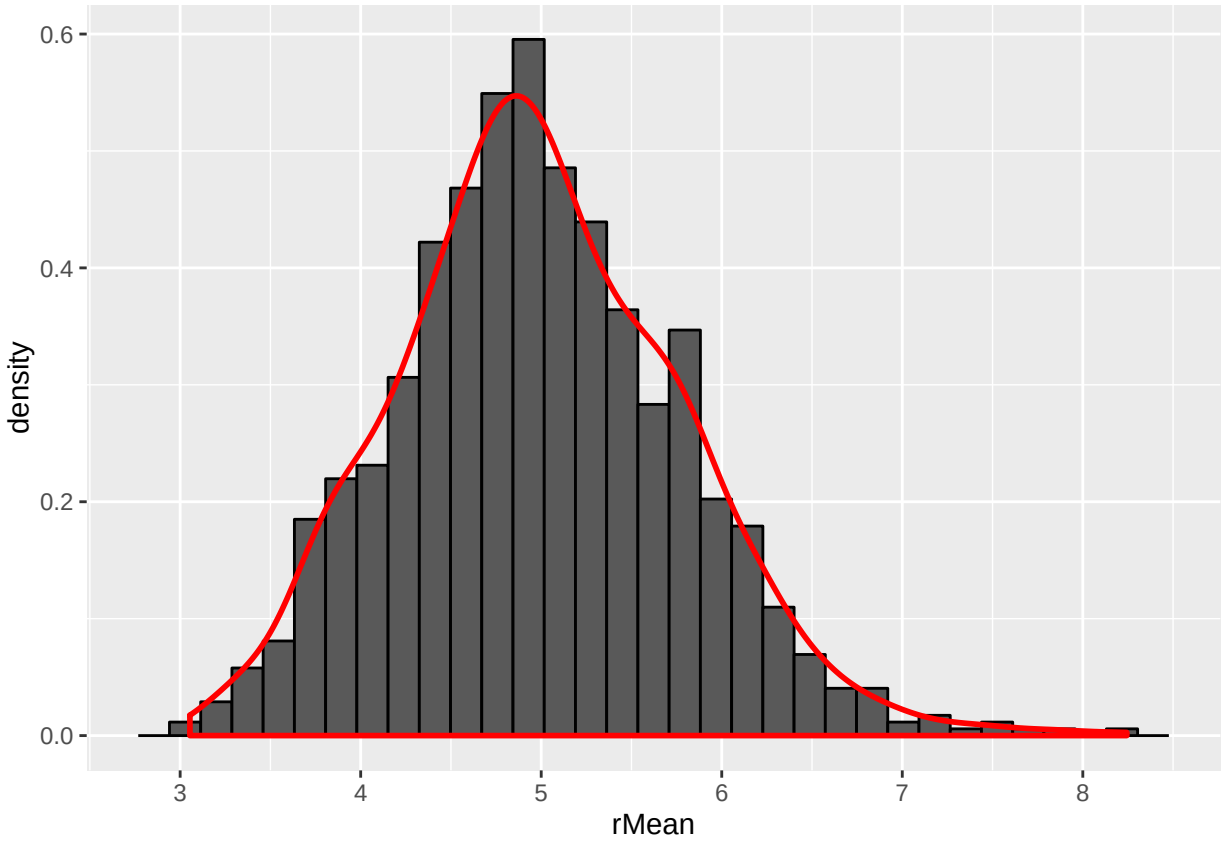
Distribution

The plot below shows that the distribution is approximately normal.

```
library(ggplot2)

vis <- data.frame(rMean)
a <- ggplot(vis, aes(x = rMean))
a <- a + geom_histogram(aes(y=..density..), colour="black")
a + geom_density(colour="red", size=1)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```


0.6
0.4
0.2
0.0
density
3
4
5
6
7
8
rMean