Statistical_Infer_Part1.Rmd

Joe Okelly

Overview

The project consists of two parts:

1. A simulation exercise.
2. Basic inferential data analysis.

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Simulation excercise

```
library(ggplot2)
library(knitr)
```

```
lambda <- .2
num_exponential <- 40
num_of_simulations <- 1000
set.seed(13411)
```

Number of exponentials 40, num of simulations 1000, lambda .2

```
simulation_exp <- replicate(num_of_simulations, rexp(num_exponential, lambda))
mean_exponential <- apply(simulation_exp, 2, mean)

mean_sample <- mean(mean_exponential)
mean_sample
```

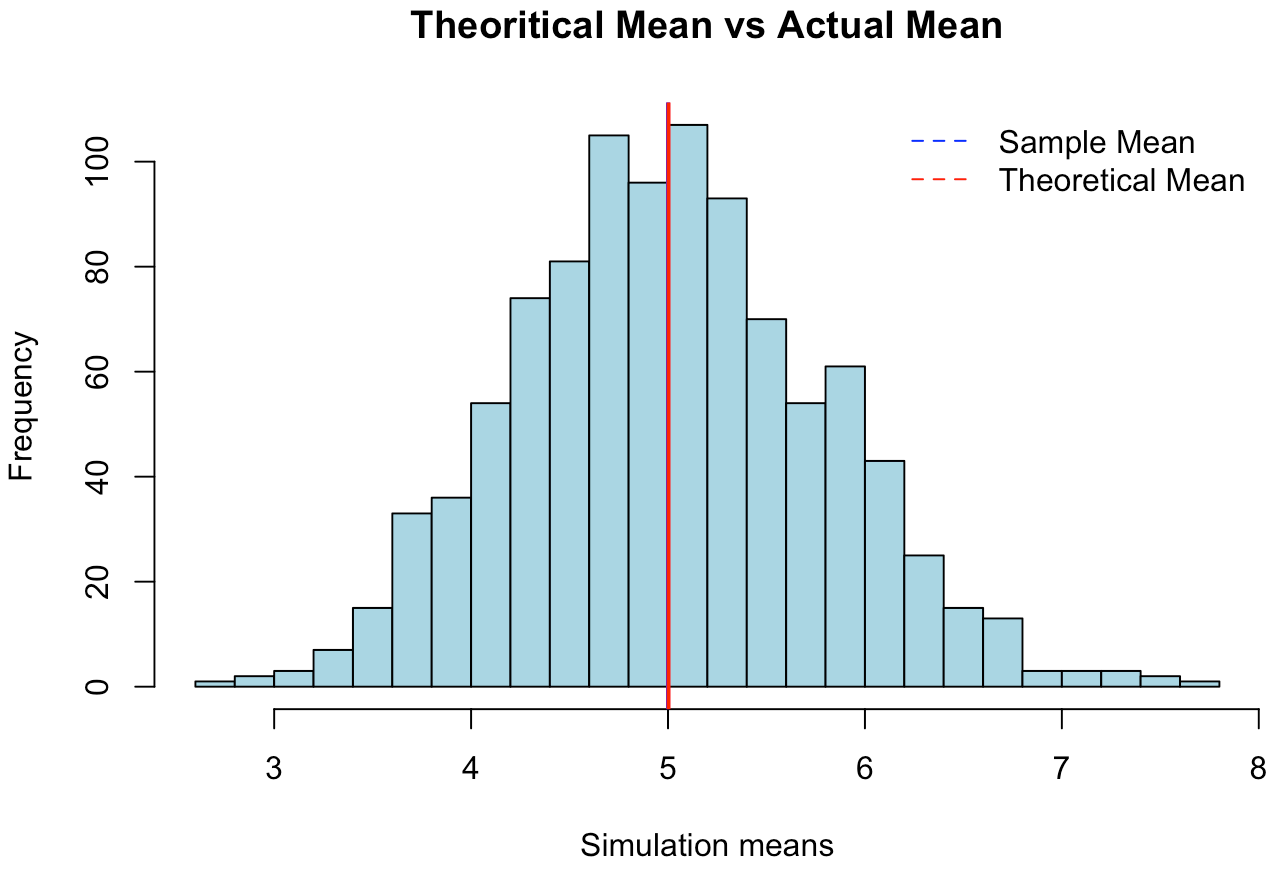
```
## [1] 5.002583
```

```
mean_theory <- 1 / lambda
mean_theory
```

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

Plot the sample mean vs Theoretical mean

```
hist(mean_exponential, col="light blue", main = "Theoritical Mean vs Actual Mean", break
s=30, xlab = "Simulation means")
abline(v=mean_theory, lwd=2, col="blue")
abline(v=mean(mean_sample), lwd=2, col="red")
legend('topright', c("Sample Mean", "Theoretical Mean"),
       bty = "n",
       lty = c(2,2),
       col = c(col = "blue", col = "red"))
```



Compare sample Standard Deviation vs Theoretical Standard Deviation

```
theoretical_deviation <- round((1/lambda) / sqrt(num_exponential), 4)
theoretical_deviation
```

```
## [1] 0.7906
```

```
sample_standard_deviation <- round(sd(mean_exponential), 4)
sample_standard_deviation
```

```
## [1] 0.7844
```

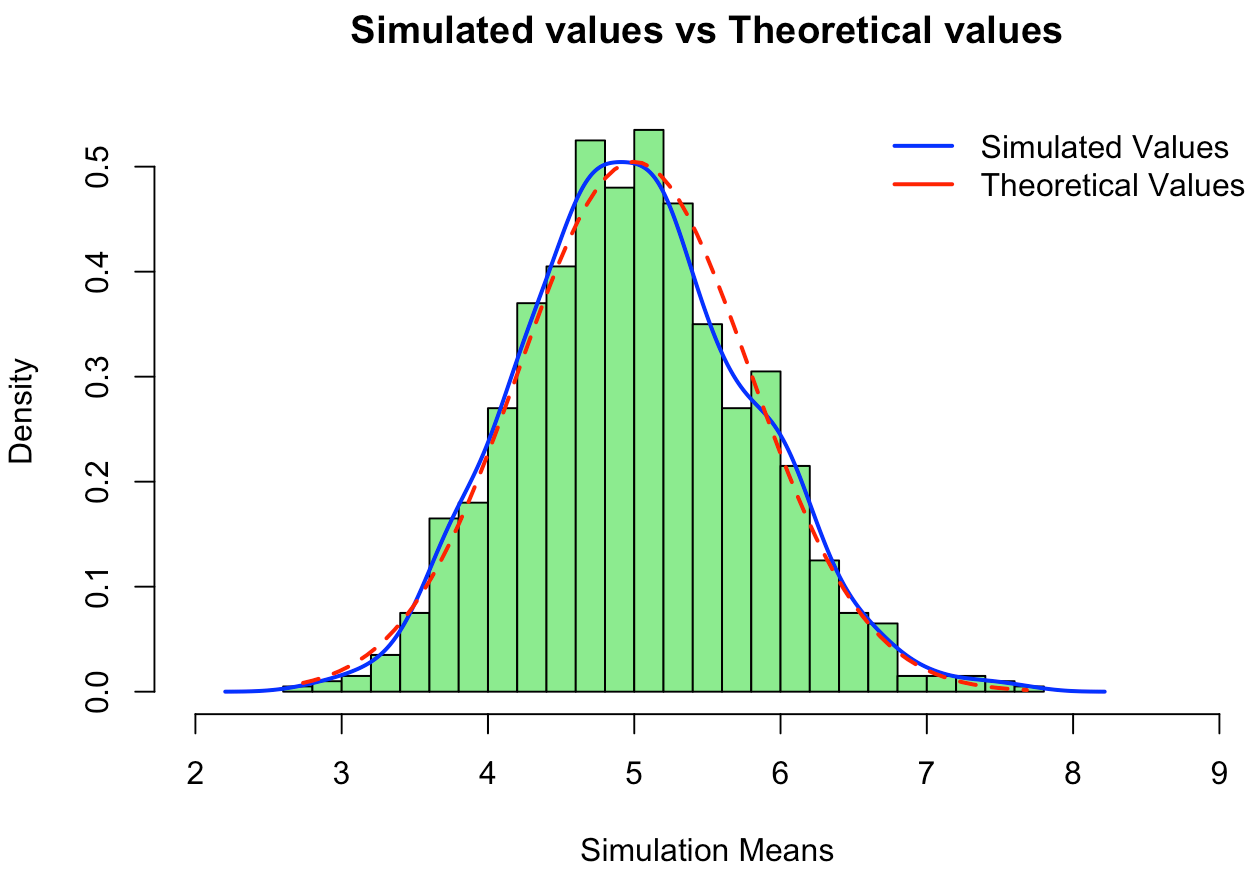
Compare sample Variance vs Theoretical variance

theoretical_variance = (1/lambda)^2/num_exponential theoretical_variance

sample_variance = var(mean_exponential) sample_variance

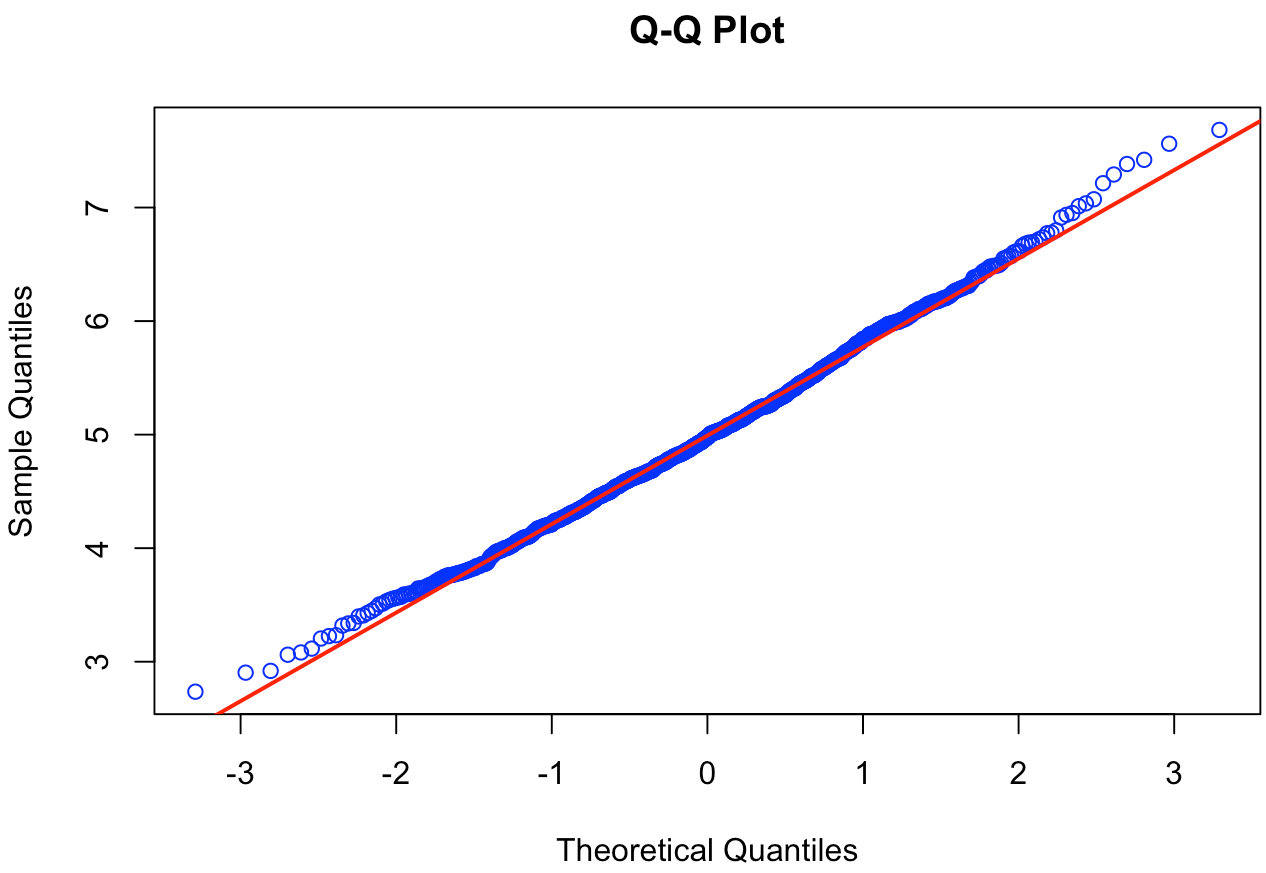
Plot the Graph

```
hist(mean_exponential, prob=TRUE, col="light green", main="Simulated values vs Theoretic
al values", breaks=30, xlim=c(2,9), xlab = "Simulation Means")
lines(density(mean_exponential), lwd=2, col="blue")

x <- seq(min(mean_exponential), max(mean_exponential), length=2*num_exponential)
y <- dnorm(x, mean=1/lambda, sd=sqrt(((1/lambda)/sqrt(num_exponential))^2))
lines(x, y, col="red", lwd=2, lty = 2)

legend('topright', c("Simulated Values", "Theoretical Values"),
      bty = "n", lwd = c(2,2), col = c("blue", "red"))
```



Plot the Graph QQNormal

```
qqnorm(mean_exponential,main ="Q-Q Plot", col = "blue")
qqline(mean_exponential, col = "red", lwd = 2)
```


Q-Q Plot
Sample Quantiles
3
4
5
6
7
-3
-2
-1
0
1
2
3
Theoretical Quantiles