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
```

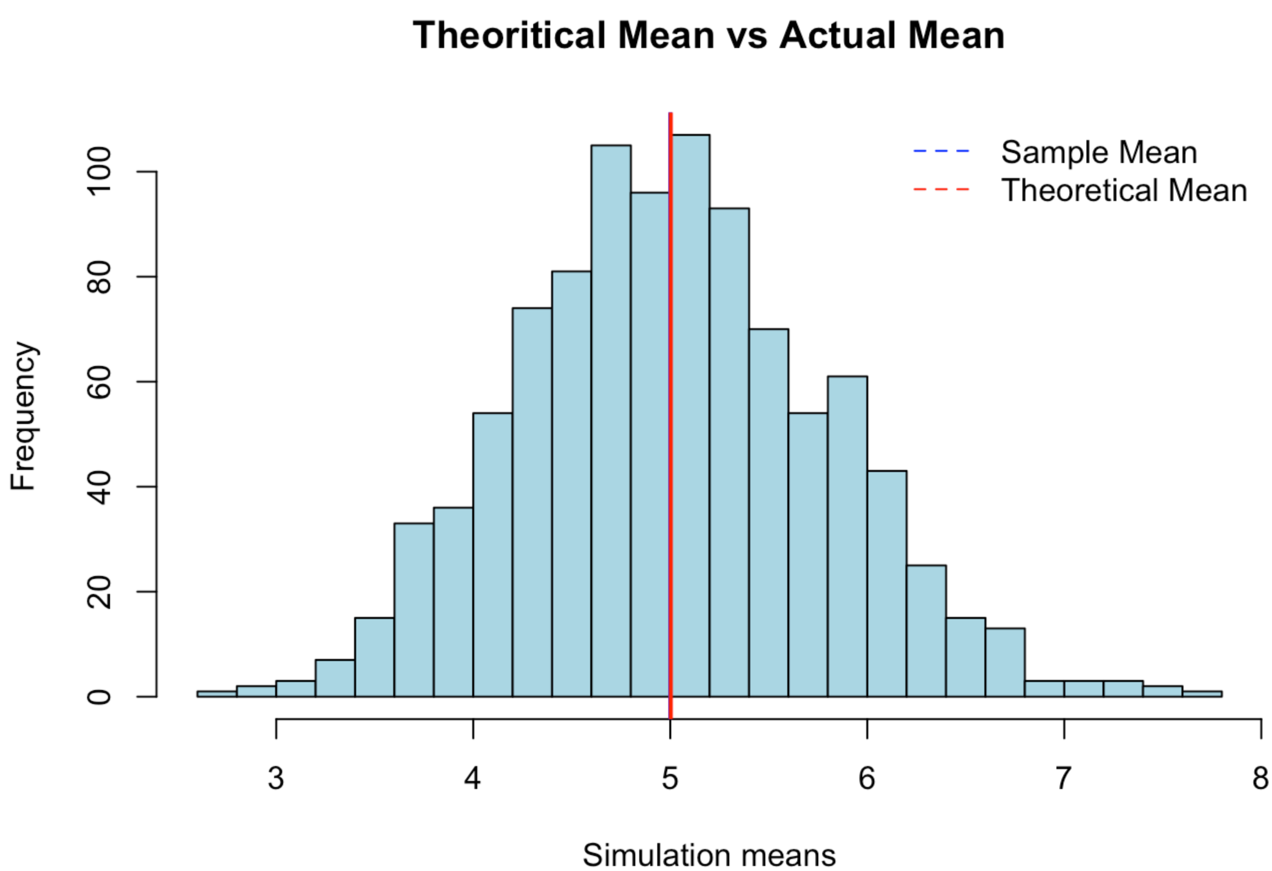
```
## [1] 5.002583
```

```
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"))
```



From the above plot we can see the distribution of the means is centered around the mean. 5.002583 which is very close and overlaps the theoritical mean 1/lambda = 5

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
```

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

```
sample_variance = var(mean_exponential)
sample_variance
```

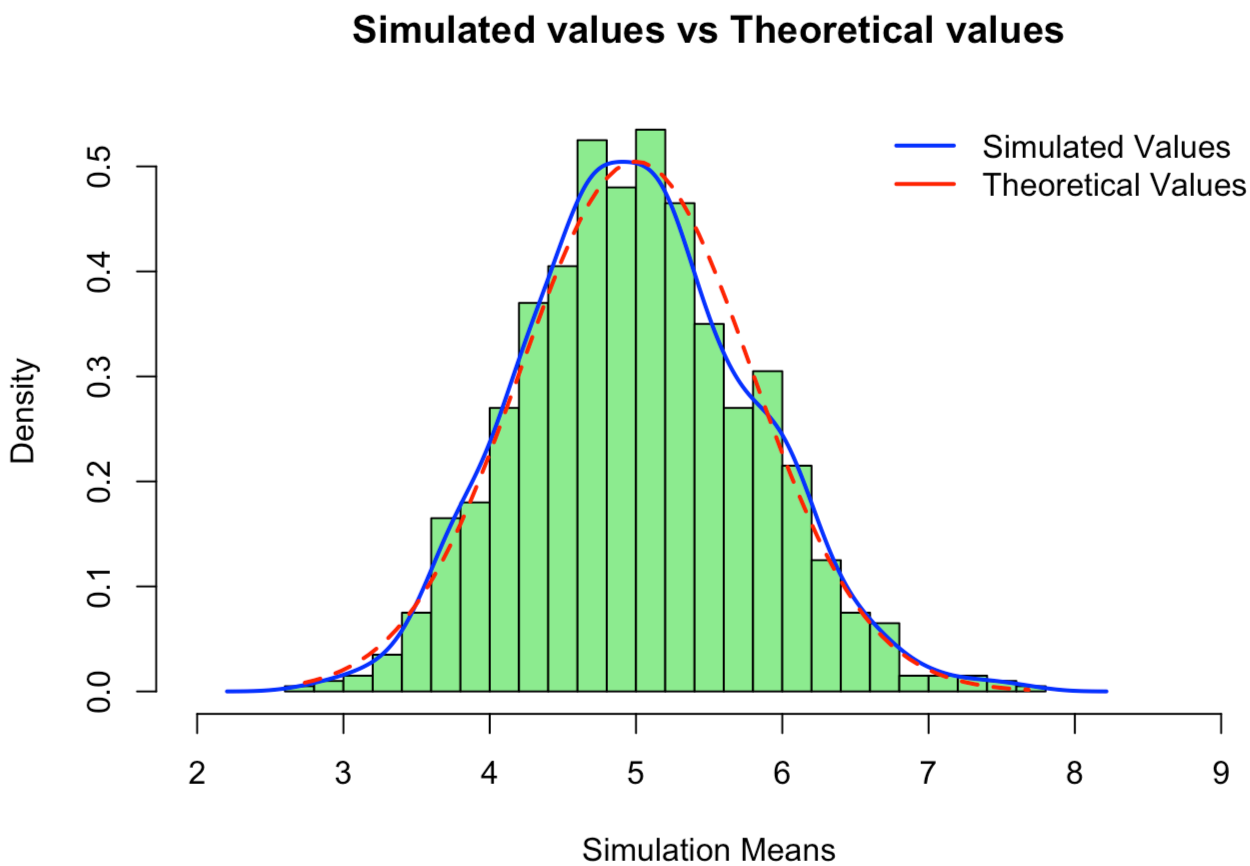
```
## [1] 0.6152212
```

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"))
```



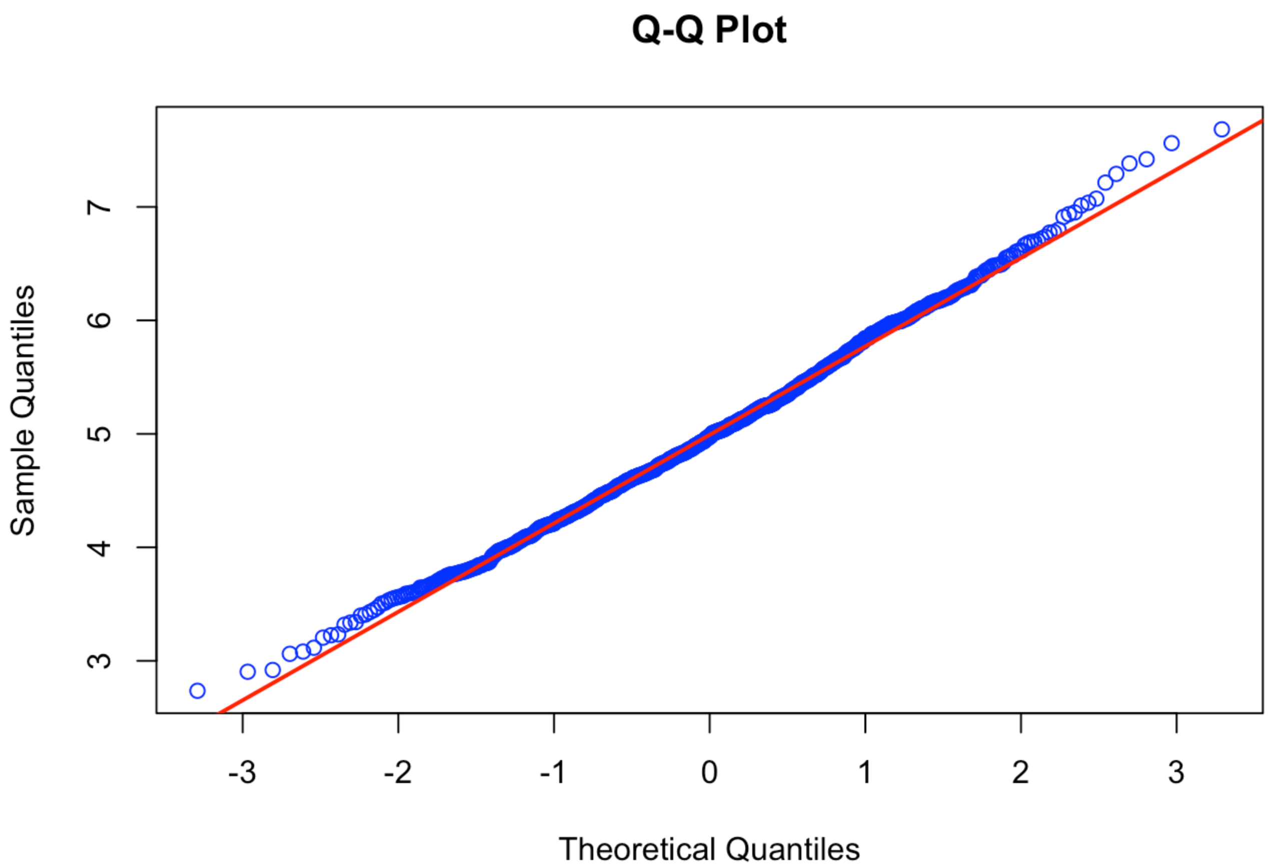
Your output shows a very close match between the theoretical and sample variances:

Theoretical Variance: 0.625 Sample Variance: 0.6152

This small difference is expected due to the inherent variability in the simulation process. The theoretical variance is derived from the properties of the exponential distribution and assumes an infinite number of samples, while the sample variance is based on the 1000 simulations you ran. The result demonstrates that your sample data closely follows the theoretical expectations, which is a good indicator of a well-behaved simulation.

Plot the Graph QQNormal

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



qqnorm(mean_exponential): Creates the QQ plot to compare your sample means with a normal distribution.
qqline(mean_exponential): Adds a reference line that represents the theoretical normal distribution.

If the sample means are normally distributed, the points should fall close to this line. This plot will help you visually assess if the sample means follow a normal distribution, which is expected by the Central Limit Theorem, especially as the sample size increases.