Statistical Inference - Project 1
Compare Exponential Distribution with the Central Limit Theorem

Rajesh Thallam

Overview

The aim of this project is to 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 λ is the rate parameter. The mean of the exponential distribution is $\frac{1}{\lambda}$ and the standard deviation is also $\frac{1}{\lambda}$. For this simulation, we set λ = 0.2 and investigate the distribution of averages of 40 samples drawn from the exponential distribution with λ=0.2.

Simulations

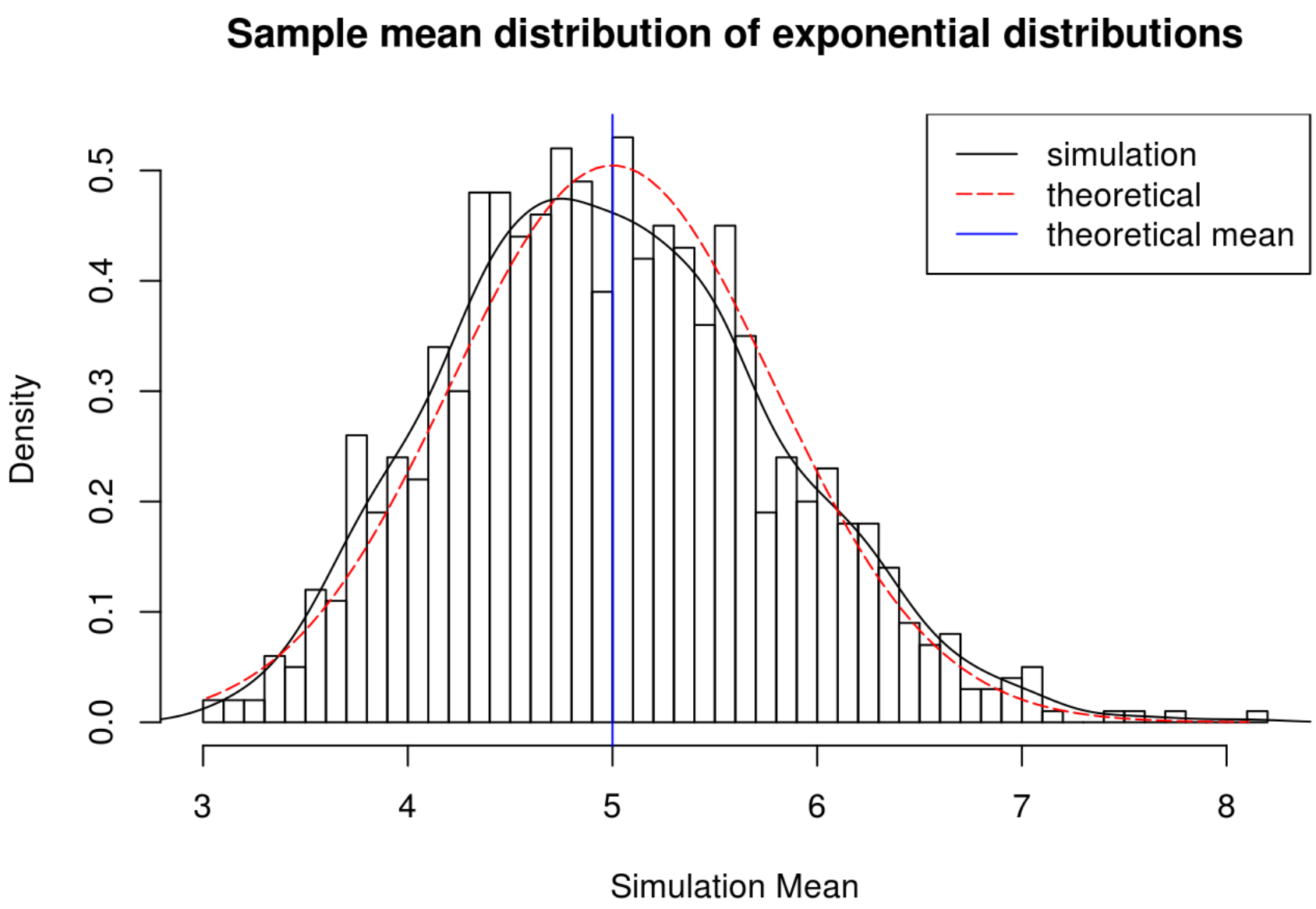
Set the simulation parameters - lambda (0.2), number of simulations (1000), and sample size (40). Before starting the simulations, seed is set to reproduce the same results. `rexp` is used to run the simulations with results of simulations aptured in a 1000-row matrix.Tthe averages for the 40 exponentials are captured in a row matrix.

```
# set values for simulation
# set seed to reproduce the results
set.seed(40)
# set number of simulations, sample size and rate parameter
nosim <- 1000
n <- 40
lambda <- 0.2

# create a 1000-row matrix with each row representing
# 40 samples drawn from the exp distribution
simulated_data <- matrix(rexp(nosim * n, rate = lambda), nosim)

# calculate the mean for each row
row_means <- rowMeans(simulated_data)
```

Distribution of Sample Means

Plotting the distribution of sample means will help answering the questions better.

```
hist(row_means, breaks = 50, prob = TRUE, main = "Sample mean distribution of exponen
tial distributions", xlab = "Simulation Mean")
lines(density(row_means))
abline(v = 1/lambda, col = "blue")
x_fit <- seq(min(row_means), max(row_means), length = 100)
y_fit <- dnorm(x_fit, mean = 1/lambda, sd = (1/lambda/sqrt(n)))
lines(x_fit, y_fit, pch = 50, col = "red", lty = 5)
legend('topright', c("simulation", "theoretical", "theoretical mean"), lty = c(1,5),
col = c("black", "red", "blue"))
```



1. Sample Mean versus Theoretical Mean

Let's compare the theoretical mean $\frac{1}{\lambda}$ with the sample mean `mean(row_means)`.

```
t_mean <- 1/lambda
s_mean <- round(mean(row_means), 3)
```

```
## [1] "Theoretical mean:  5"
```

```
## [1] "Sample mean:  4.989"
```

As we can observe, the distribution of sample means is centered at 4.989 and the theoretical center of the distribution is $\frac{1}{\lambda}$ (5), which is a close estimate of the population mean.

2. Sample Variance versus Theoretical Variance

Comparing the theoretical variance $\frac{1}{\lambda^2}$ with the sample variance `var(row_means)` and the theoretical standard error $\frac{1}{(\lambda*\sqrt{n})}$ with the sample standard error `sd(row_means)`.

```
t_var <- (1/lambda)^2/n;
t_sd <- round(1/(lambda*sqrt(n)), 3);

s_var <- round(var(row_means), 3)
s_sd <- round(sd(row_means), 3)
```

```
## [1] "Theoretical variance:  0.625"
```

```
## [1] "Sample variance:  0.643"
```

```
## [1] "Theoretical standard error:  0.791"
```

```
## [1] "Sample standard error:  0.802"
```

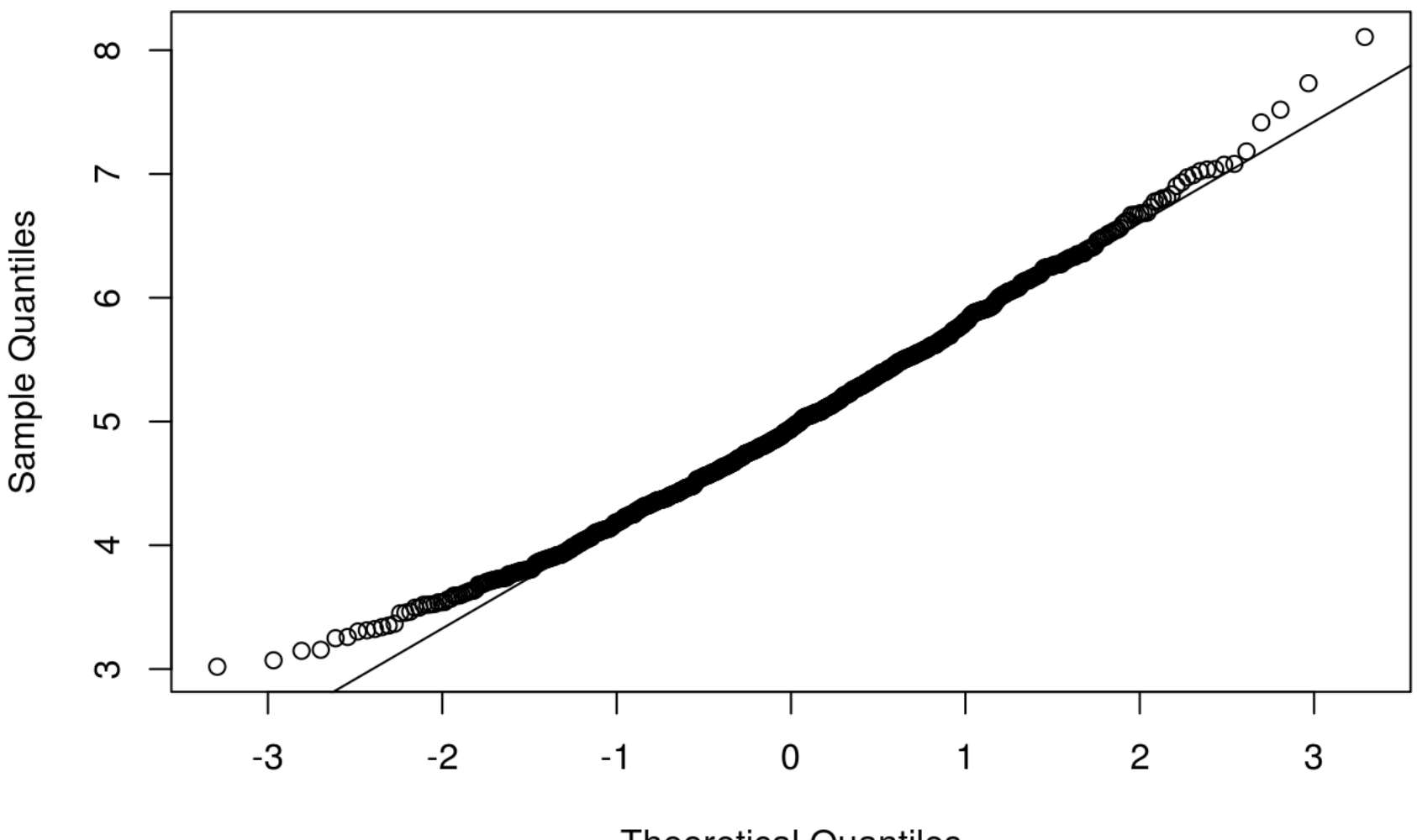
The variance of the sample means is 0.643 and the theoretical variance of the distribution is 0.625 which are pretty close and may be same if we run more simulations. Similarly the standard error values.

3. Distribution is approximately normal

By the central limit theorem, the averages of samples will follow a normal distribution. The plot above also shows the density computed using the histogram and the normal density plotted with theoretical mean and variance values. Both distributions are almost similar suggestinf normality.

Another way to confirm normality is by plotting qq (quantile quantile) showing that simulation quartiles closely match the theoretical normal quartiles.

```
qqnorm(row_means)
qqline(row_means)
```



Confidence Intervals

```
t_ci <- t_mean + c(-1,1)*1.96*sqrt(t_var)/sqrt(n)
s_ci <- round(mean(row_means) + c(-1,1)*1.96*sd(row_means)/sqrt(n),3)
```

```
## [1] "Theoretical Confidence Interval:  4.755"
## [2] "Theoretical Confidence Interval:  5.245"
```

```
## [1] "Simulated Confidence Interval:  4.74"
## [2] "Simulated Confidence Interval:  5.237"
```

Looking at the confidence intervals, both the simulated and theoretical versions are pretty close.

Conclusion

All the tests confirm that distribution of the simulation results and theoretical results are siginificantly similar.