Statistical Inference Course Project Part 1

Overview: In this Report we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution will 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. We will investigate the distribution of averages of 40 exponentials. We will need to do a thousand simulations.

Simulation

n=40 lambda = 0.2 theoretical mean = 5 standard deviation = 5

1. Create a exponential distribution with 40 exponentials.

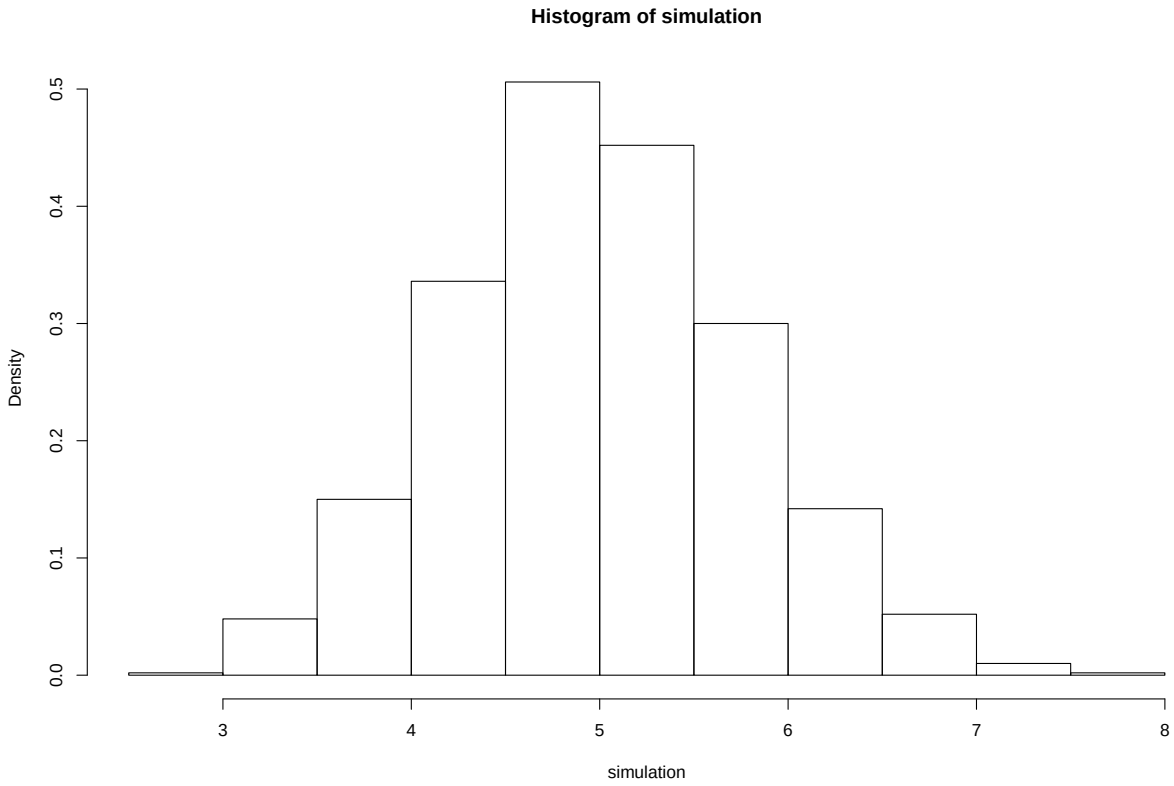
```
set.seed(4000)
n <- 40
lam <- 0.2
sample <- rexp(n, lam)
```

2. Run a thousand simulations

```
simulation = NULL
for (i in 1 : 1000) simulation = c(simulation, mean(rexp(n, lam)))
```

3. Histogram of the simulations

```
hist(simulation,prob=TRUE)
```



simulation mean vs theoretical mean

```
sim_mean <- mean(simulation)
```

The simulation mean is 4.988818 is very close to the theoretical mean of 5.

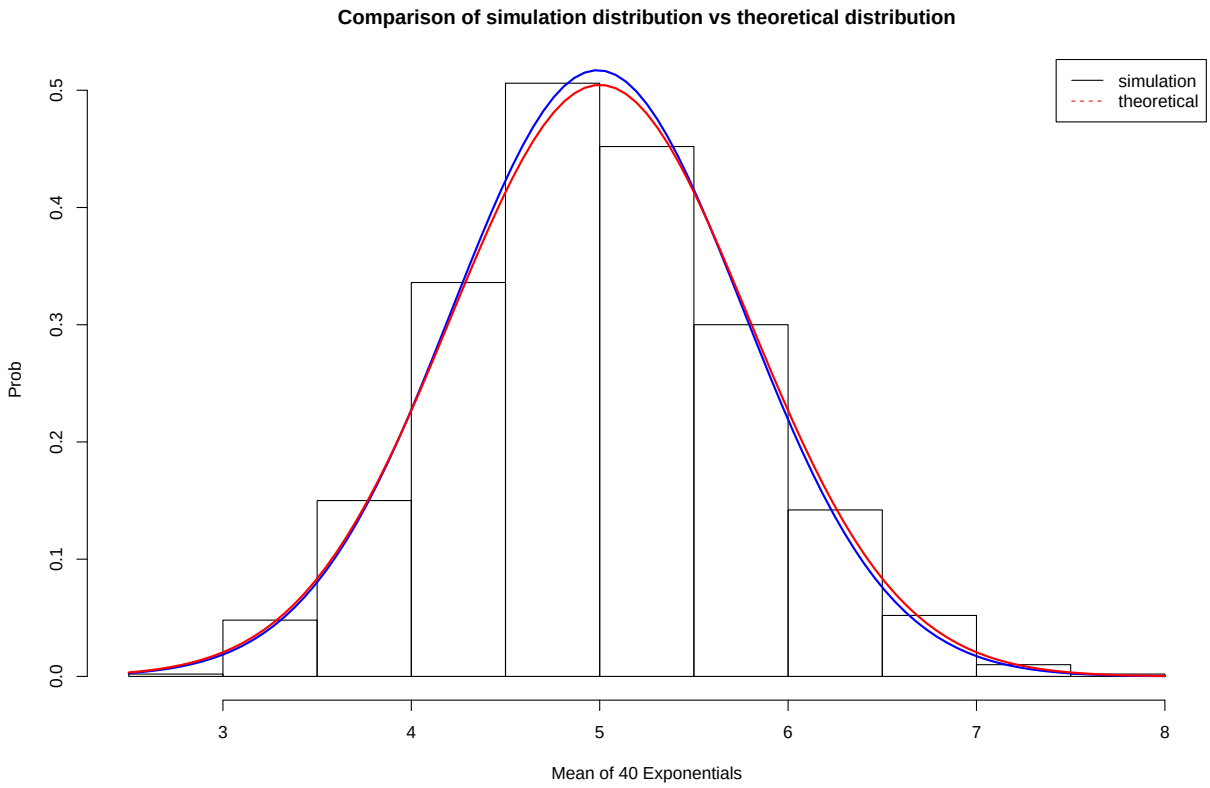
simulation variance vs theoretical variance

```
sim_var <- var(simulation)
theo_var <- (1/lam)^2 /n
```

The simulation variance is 0.5951333 is very close to the theoretical variance of 0.625.

Distribution of simulation vs theoretical distribution

```
hist(simulation,prob=TRUE, xlab = "Mean of 40 Exponentials", ylab = "Prob"
     , main = "Comparison of simulation distribution vs theoretical distribution")
curve(dnorm(x, mean=mean(simulation), sd=sd(simulation)), add=TRUE, col="Blue", lwd = 2)
curve(dnorm(x, mean=5, sd=sqrt(theo_var)), add=TRUE, col="Red", lwd = 2)
legend('topright', c("simulation", "theoretical"), lty=c(1,2), col=c("black", "red"))
```



As shown in the graph above, the simulation will be approximated with the normal distribution

Confidence intervals

The simulation confidence intervals is:

```
sim_mean + c(-1,1) * qt(0.975, 39) * sqrt(sim_var) * sqrt(1/40)
```

```
## [1] 4.742097 5.235539
```

The theoretical confidence intervals is:

```
5 + c(-1,1) * qt(0.975, 39) * sqrt(theo_var) * sqrt(1/40)
```

```
## [1] 4.747164 5.252836
```

Conclusion

This exercise shown that, based on the central limit theorem, when the sample size is big enough, the sample distribution will be approximated with the normal distribution.