Part1: Statistical Inference

Minki Jo

November, 12, 2017

Overview

This project is to investigate the exponential distribution in R and compare it with the Central Limit Theorem. This creates a report to answer three questions. The report is submitted as a pdf type document

Simulation

- Investigate the distribution of averages of 40 exponentials with 1,000 simulations.
- For all simulation, the parameter, lambda is to set 0.2. The mean and the standard deviation of exponential distribution are known to be 1/lambda (population)
- In R, the exponential distribution can be simulated with rexp(n, lambda).

Problems

This project is to answer three problems as follows:

1. Show the sample mean and compare it to the theoretical mean of the distribution
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

Sample Mean versus Theoretical Mean:

Here is simulation for 40 samples mean with 1,000 times simulation

```
lambda <-0.2
n <-40
rep <-1000

set.seed(1234)
s_mean_dist <-c()
for (i in 1:rep){
s_mean_dist <-c(s_mean_dist,mean(rexp(n,lambda)))
}
```

Let's save mean of sample distribution and theoretical mean, 1/lamdba

```
s_mean <-mean(s_mean_dist)
t_mean <-1/lambda
```

Here is `sample mean` for 40 Samples with 1,000 simulatios

```
s_mean
```

```
## [1] 4.974239
```

Here is `theoretical mean`, 1/lambda

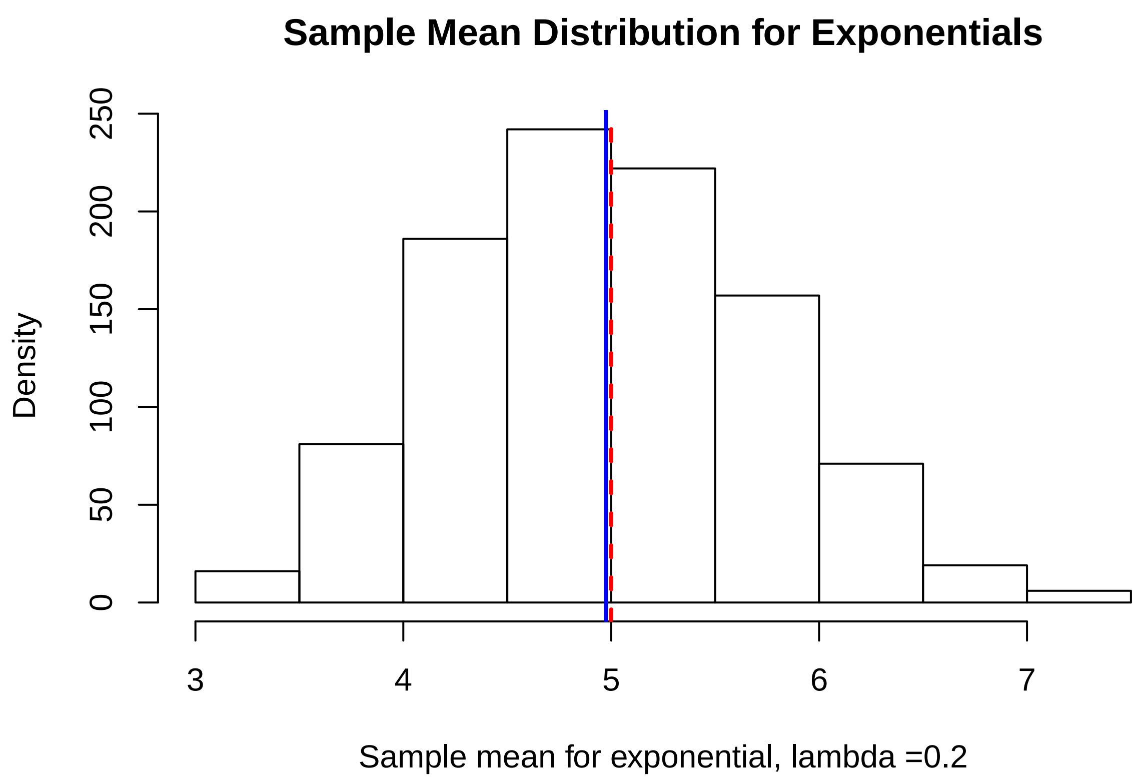
```
t_mean
```

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

we can see both are pretty close to each other

Now, compare sample mean and theoretical mean with sample mean distribution for exponential

```
hist(s_mean_dist, xlab = "Sample mean for exponential, lambda =0.2", ylab ="Density",main = "Sample Mea
abline(v = s_mean, col = "blue", lwd=2)
abline(v = t_mean, col = "red", lwd=2, lty =2)
```



Sample Variance versus Theoretical Variance:

Let's compare sample variance with theoretical variance. The theoretical standard deviation is eqaul to standard error of mean and populationss standard deviation is already known as 1/lambda. Here are `theoretical standard deviation and varinace`

```
t_std <-1/lambda
t_se <-t_std/sqrt(n)
t_var <-t_se^2
t_se
```

```
## [1] 0.7905694
```

```
t_var
```

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

Here are `sample's variacne and standard deviation`

```
s_std <- sd(s_mean_dist)
s_var <- var(s_mean_dist)
s_std
```

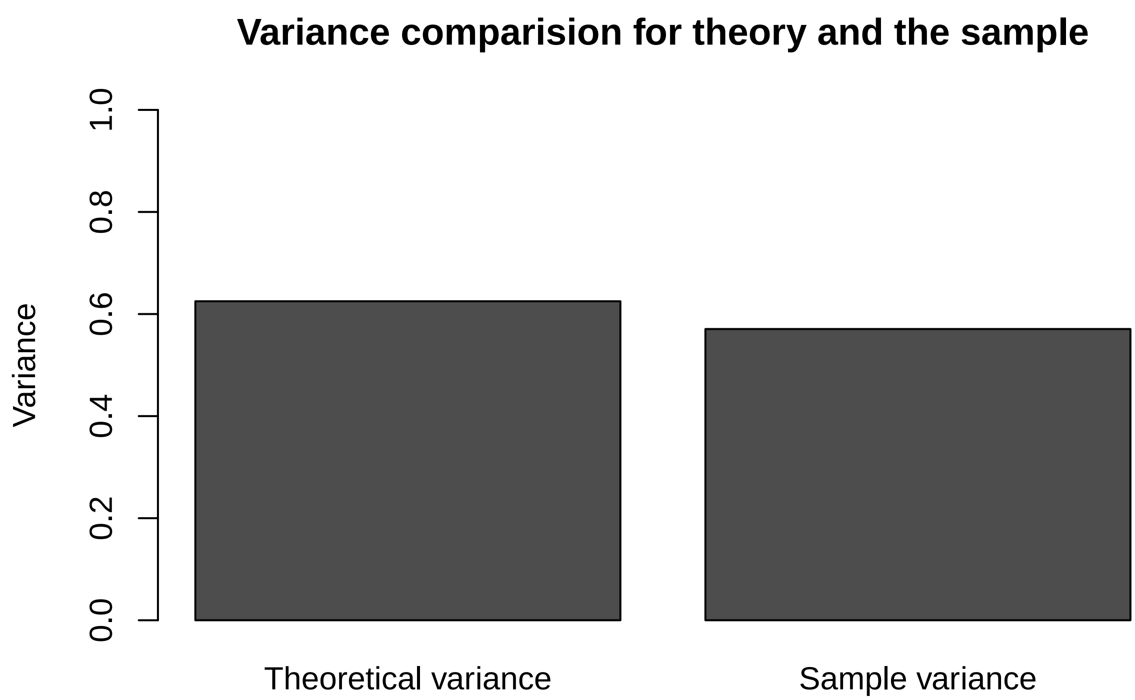
```
## [1] 0.7554171
```

```
s_var
```

```
## [1] 0.5706551
```

Let's compare theoretical variacne with sample variance with barplot

```
var <-matrix(c(t_var,s_var),ncol=2)
colnames(var) <-c("Theoretical variance", "Sample variance")
barplot(var, main ="Variance comparision for theory and the sample", ylab ="Variance", ylim=c(0,1))
```



In sum, mean, variance, and standard deviation for theory and simulation are shwon in the below table

```
average <- c(t_mean,s_mean)
variance <-c(t_var,s_var)
standard_deviation <- c(t_se,s_std)
comparison <-rbind(average,variance,standard_deviation)
colnames(comparison) <-c("Theoretical Value", "Simulation Value")
library(knitr)
kable(round(comparison,3), caption ="Comparison Table")
```

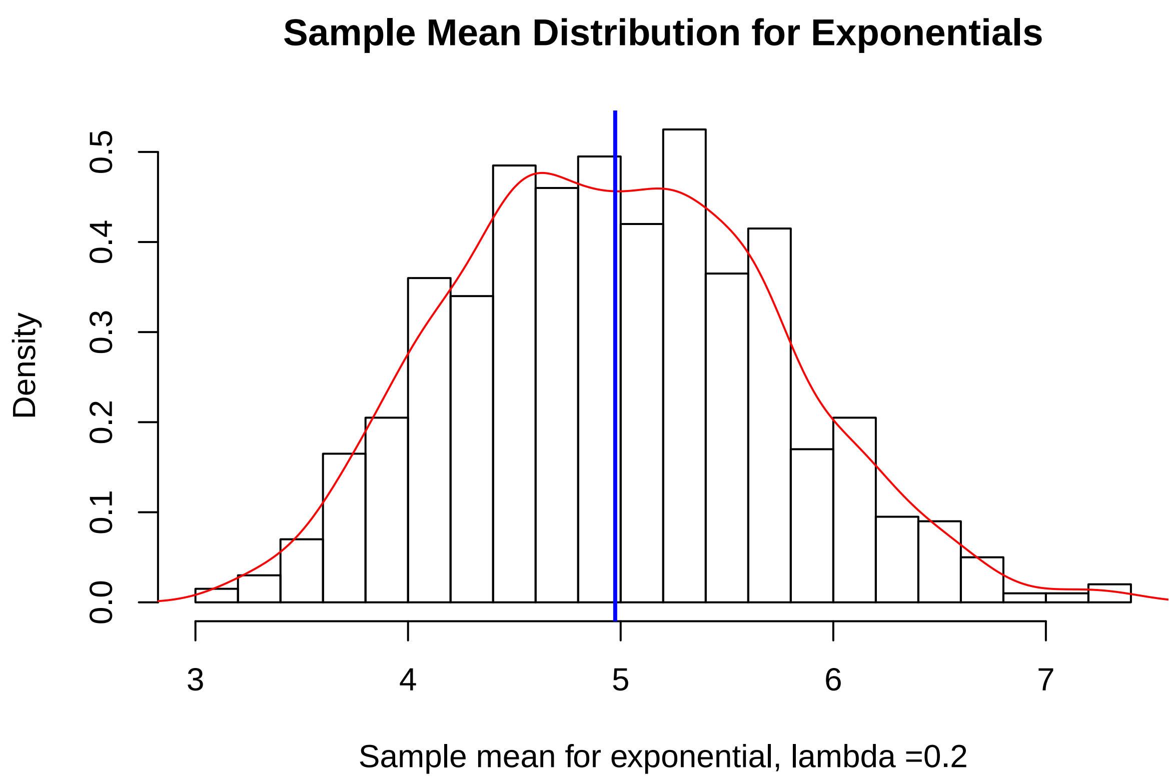
Table 1: Comparison Table

	Theoretical Value	Simulation Value
average	5.000	4.974
variance	0.625	0.571
standard_deviation	0.791	0.755

Distribution

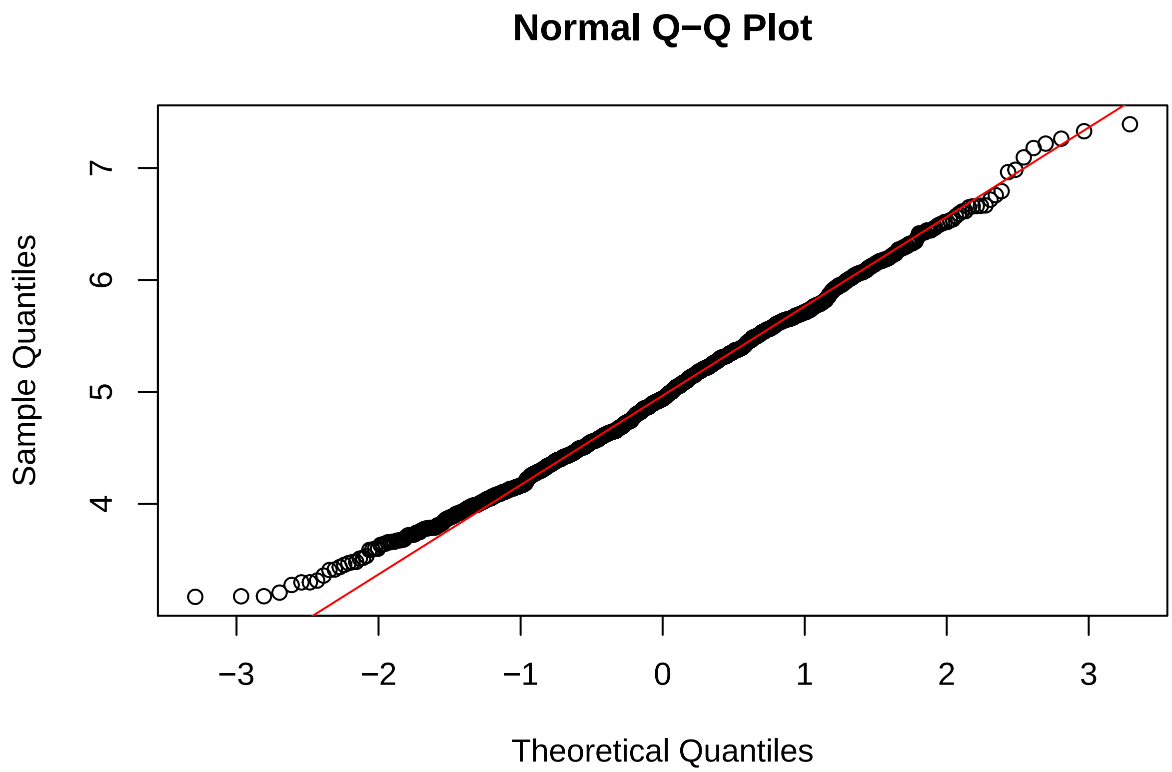
Let's draw contidnuous density distribution with discrete distribution

```
hist(s_mean_dist, prob = TRUE, xlab = "Sample mean for exponential, lambda =0.2", main = "Sample Mean Di
lines(density(s_mean_dist), col = "red")
abline(v = s_mean, col = "blue", lwd=2)
```



The continuous distribution seems a lot like normal distribution. The qqplot is a technical tool to see if the distribution has normality or not

```
qqnorm(s_mean_dist)
qqline(s_mean_dist, col = 2)
```



From the density plot and the normal qq-plot, it is proved that the average exponential has approximately normality.