Statistical Inference - Course Project 1

Mark C

10 Nov 2014

Part 1: Simulation Exercise

Simulations The exponential distribution can be simulated in R with `rexp(n, lambda)` where `lambda`(λ) is the rate parameter. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. For this simulation, we set λ=0.2. In this simulation, we investigate the distribution of averages of 40 exponential(0.2)s.

Simulation Step : Generating a thousand simulated averages of 40 exponentials.

```
library(ggplot2)

# Set a seed
set.seed(10)
lambda <- 0.2
# Generating 1000 simulations with 40 samples
sample_size <- 40
simulations <- 1000

# Storing the 1000 simulations output
simulated_exponentials <- matrix(rexp(simulations*sample_size, rate=lambda), simulations, sample_size)
# Getting the average of 40 exponentials from the simulations
row_means <- rowMeans(simulated_exponentials)
```

Simulation Output

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.

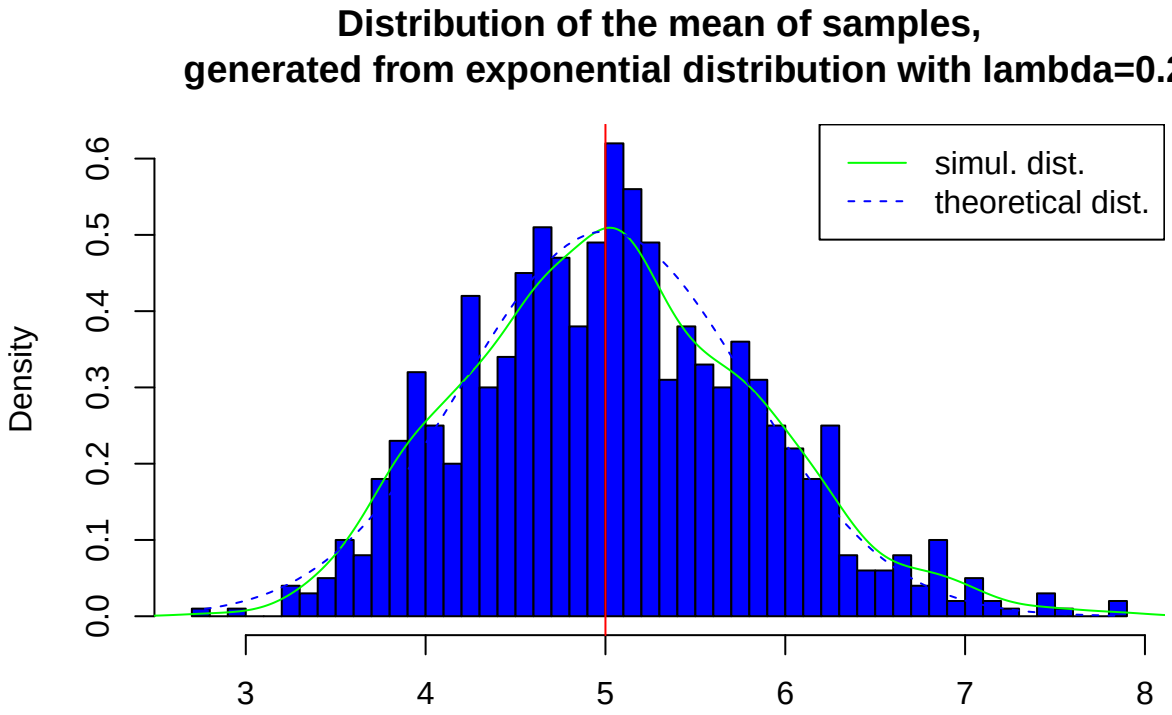
```
# output the mean of distribution of averages of 40 exponentials
mean(row_means)
```

```
## [1] 5.04506
```

```
# output the mean from analytical expression
1/lambda
```

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

The distribution of sample means is shown below:



The distribution of mean of 40 exponentials is centered at 5.0450596 and it is shown to be close to the theoretical center of the distribution,$(1/\lambda^{-1} = 1/\lambda)$.

2. Show how variable it is and compare it to the theoretical variance of the distribution.

```
# Calculate the standard deviation of mean of 40 exponentials
sd(row_means)
```

```
## [1] 0.80881
```

```
# Calculate the standard deviation from theoretical distribution expression
(1/lambda)/sqrt(sample_size)
```

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

```
# Calculate the Variance of the sample mean
var(row_means)
```

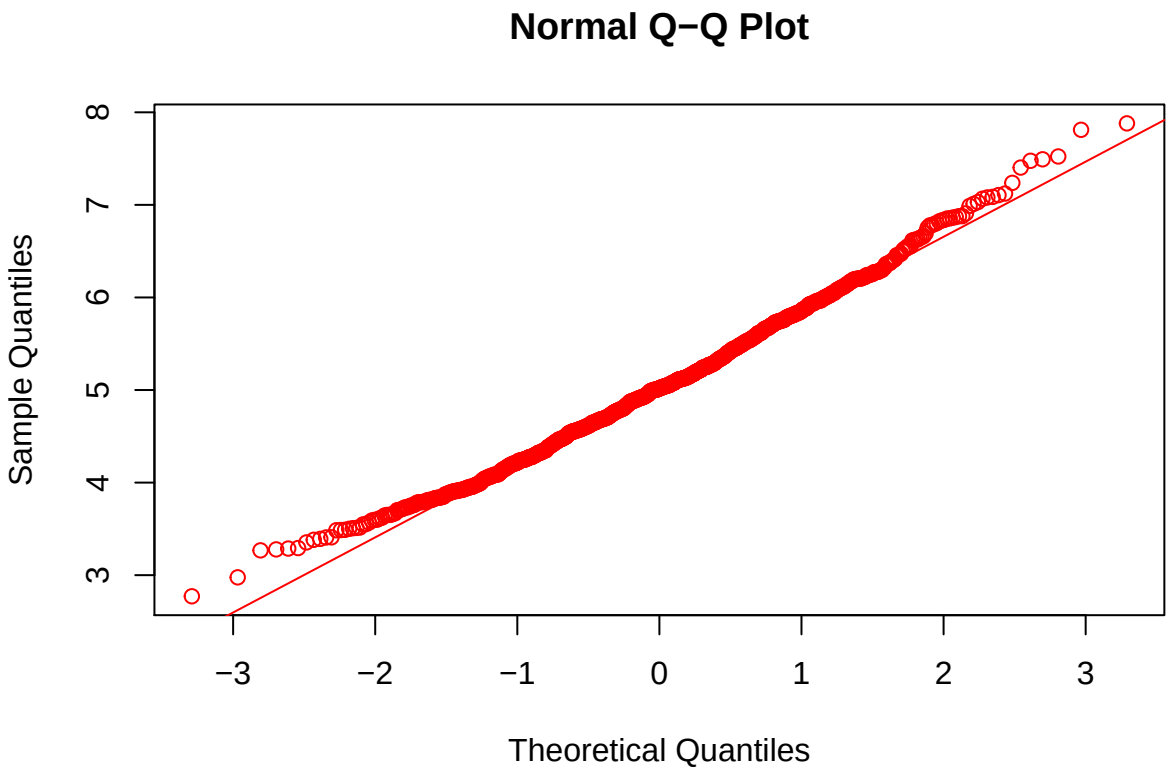
```
## [1] 0.6541736
```

```
# Calculate the theoretical variance of the distribution
1/((0.2*0.2) * 40)
```

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

From the calculation of the variance, it is shown that the simulated variance of the 40 exponential distribution is close to the theoretical variance of the distribution.

3. Show that the distribution is approximately normal.



Following the Central Limit Theorem (CLT), the average of the generated samples follows the normal distribution. The figure above also shows the density computed using the histogram and the normal density plotted with theoretical mean and variance values. The plot above also can be seen that the distribution of averages of 40 exponentials is approximated to close to a normal distribution.