EXPONENTIAL DISTRIBUTION SIMULATION USING R

By: Piyush Neupane

(PLEASE NOTE: 3 Pages report and 3 additional pages of Appendix (total of 6 pages) is allowed (according to the Course Project instruction in Coursera!))

OVERVIEW

The purpose of this project is to investigate the exponential distribution in R and compare it with the Central Limit Theorem(using rexp(n, lambda) in R) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda.

RUN SIMULATIONS:

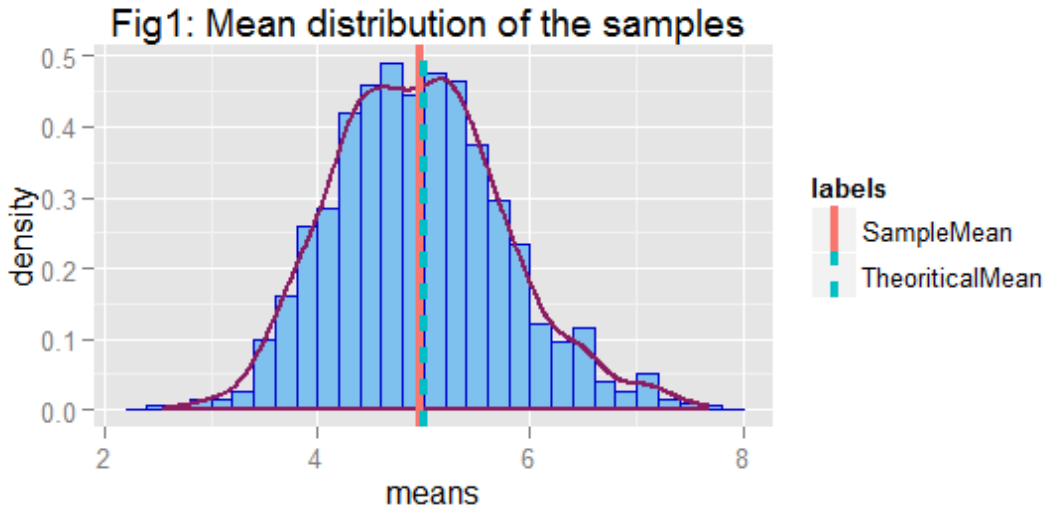
Given value for lambda is 0.2 for all of the simulations and we have to create a sample of 40 exponentials. The simulation have to be run 1000 times.

```
# set libraries
library(ggplot2)

set.seed(56248) # set seed
  lambda = 0.2    # given
  nsim = 1:1000  # Index of simulations
  n = 40 # Number of samples from Exponential distribution

# Pull 40 samples of an exponential distribution, and repeat it 1000 time
samples = sapply(nsim, function(x) rexp(n,lambda))
```

RESULTS

Question1: Sample Mean versus Theoretical Mean

The distribution of sample means is centered around 5, which is very close to theoritical mean 4.96.




```
##            labels   values
## 1      SampleMean 4.958495
## 2 TheoriticalMean 5.000000
```

Question2: Sample Variance versus Theoretical Variance

The theoritical variance and Standard Deviation are very close to Actual variance and SD. So, the the mean of exponential distribution is spread out very similar to the way it is expected.

```
# Theoritical Variance
TheoriticalVar = ( (1/lambda)/sqrt(n) )^2
TheoriticalVar
```

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

```
# Sample Var
SampleVar = var(means)
SampleVar
```

```
## [1] 0.6506186
```

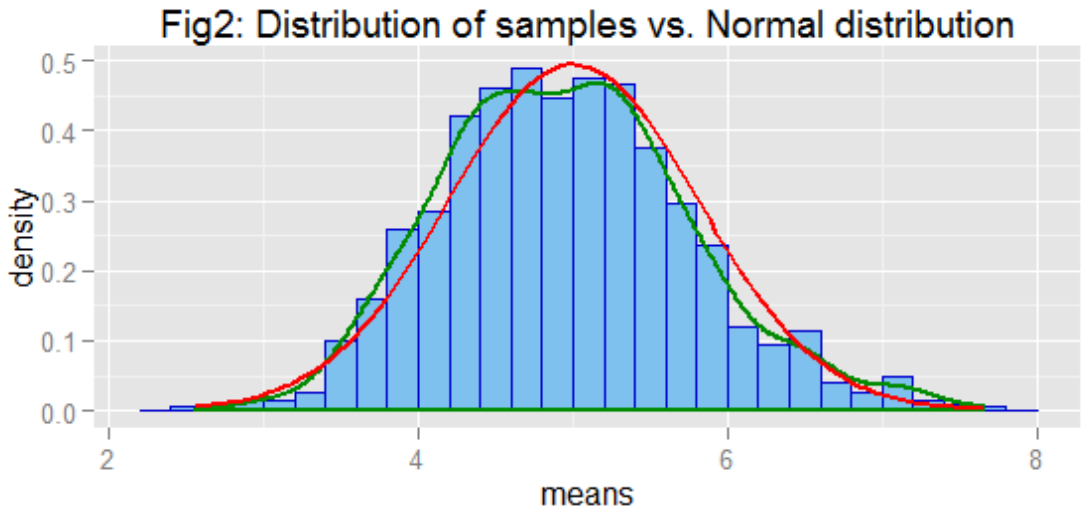
```
# Theoritical SD
sqrt(TheoriticalVar)
```

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

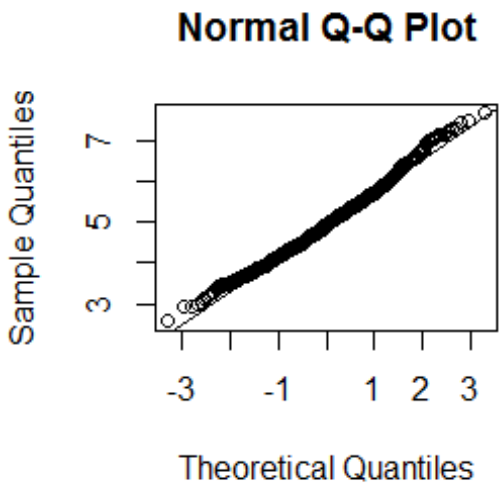
```r
# Sample SD
sqrt(SampleVar)

## [1] 0.8066093
```

Question3: Distribution: Explain how one can tell the distribution is approximately normal.

The figure shows the Normal distribution in RED curve and the actual distribution of means in GREEN curve. The actual distribution is very similar to the normal distribution.



QQ normal plots also allows us to visually check how closely the quantiles in actual distribution matches those in theoritical distribution. From this plot, the actual distribution of mean seem to match normal distribution very closely.



APPENDIX

Code to generate Fig1:

```
# Mean of each sample
means = apply(samples,2,mean)

# Mean of all sample means
mean_line = data.frame(labels = c("SampleMean", "TheoriticalMean") , values =
c( mean(means),  1/lambda))


# Plot the histogram showing the distribution of the sample means. Also
display the theoritical vs. Sample mean
ggplot(NULL, aes(x=means)) +   geom_histogram(binwidth=.2,aes(y =
..density..), fill='skyblue2', color="mediumblue" ) +
    geom_density(color="maroon4", size= 1 ) + ggtitle("Fig1: Mean
distribution of the samples ") +
    geom_vline(data=mean_line, aes(xintercept=values, linetype = labels,
color = labels ), size=1.5, show_guide=T)

# theoritical vs. Sample mean
mean_line
```

Code to generate Fig2:

```
# Plot the histogram showing the distribution of the sample means compared
with the normal distribution
ggplot(NULL, aes(x=means)) +   geom_histogram(binwidth=.2, aes(y =
..density..), fill='skyblue2', color="mediumblue" ) +
  geom_density(color="green4", size= 1 ) +
  stat_function(color="red", size =1, fun = dnorm, arg = list(mean =
1/lambda, sd = sqrt(SampleVar))) +
  scale_color_manual("",
                     breaks = c("Normal Distribution", "Actual
Distribution"),
                     values = c("red", "green4")) + ggtitle("Fig2:
Distribution of samples vs. Normal distribution")
```

Code to generate QQ

```
# In addition, create a qq plot for sample quantiles and qq line for
theoritical quantiles ;
qqnorm(means)
qqline(means)
```