SI_PA1_template_part1

abhibhaku

April 27, 2018

TITLE - Statistical Inference - Part 1

OVERVIEW - In this project, aim is to investigate the exponential distribution in R and compare it with the Central Limit Theorem. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda

Info given -

a. lambda = 0.2
b. Investigate the distribution of averages of 40 exponentials
c. Do thousand simulations

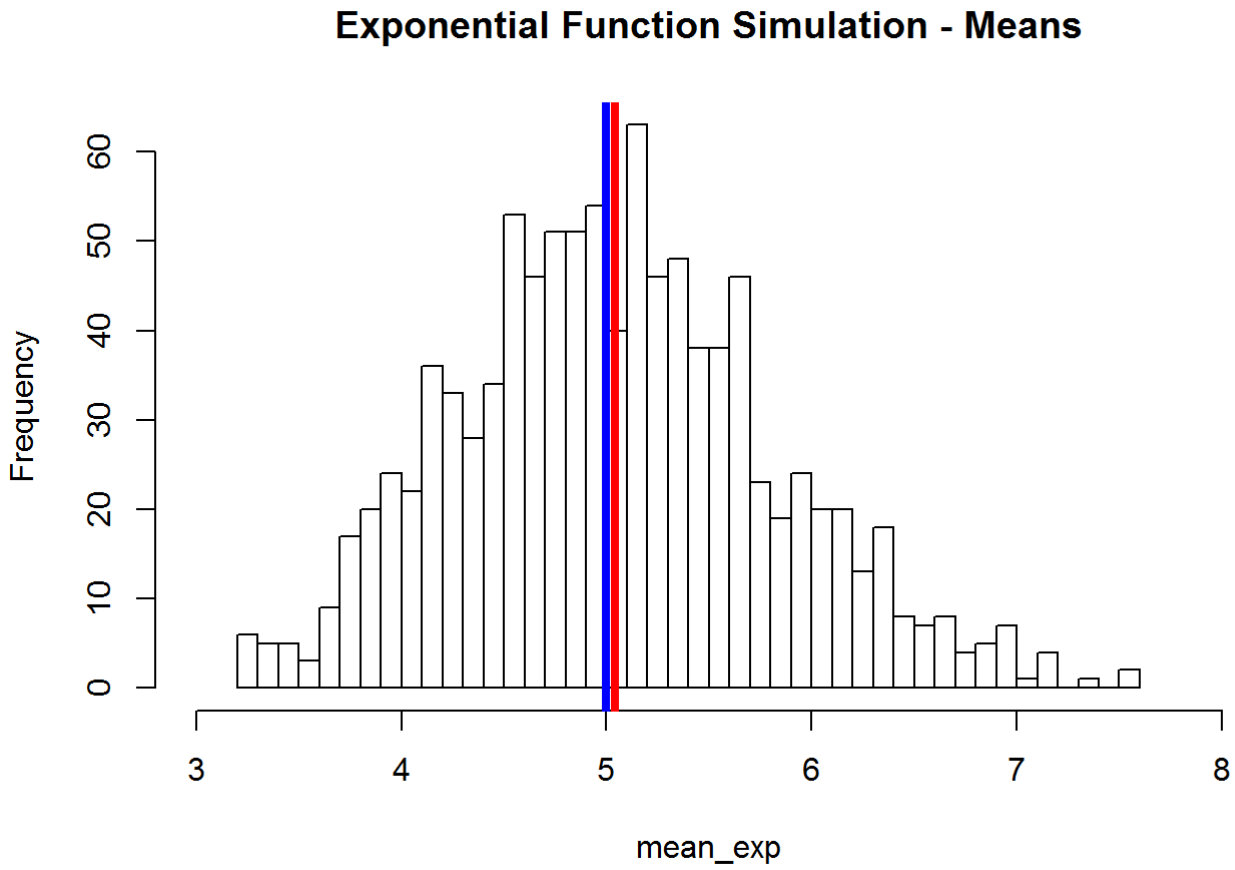
```
knitr::opts_chunk$set(echo = TRUE)

set.seed(5) # setting seed value
lambda <- 0.2 # lambda value specified
nos <- 40 # number of samples
sim <- 1000 # no. of simulations specified

# Q1 - Show the sample mean and compare it to the theoretical mean of the distribution

sim_exp <- replicate(sim,rexp(nos,lambda)) # running simulations

mean_exp <- apply(sim_exp,2,mean) # calculating mean for each sample

s_mean <- mean(mean_exp) # calculated mean from the sample
s_mean
```

```
## [1] 5.043053
```

```
t_mean <- 1/lambda # calculated theoretical mean
t_mean
```

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

```
hist(mean_exp, col="white", main = "Exponential Function Simulation - Means",breaks=40, xlim
 = c(3,8)) # plotting histogram for means
abline(v=s_mean, lwd = "4", col = "red")
abline(v=t_mean, lwd = "4", col = "blue")
```



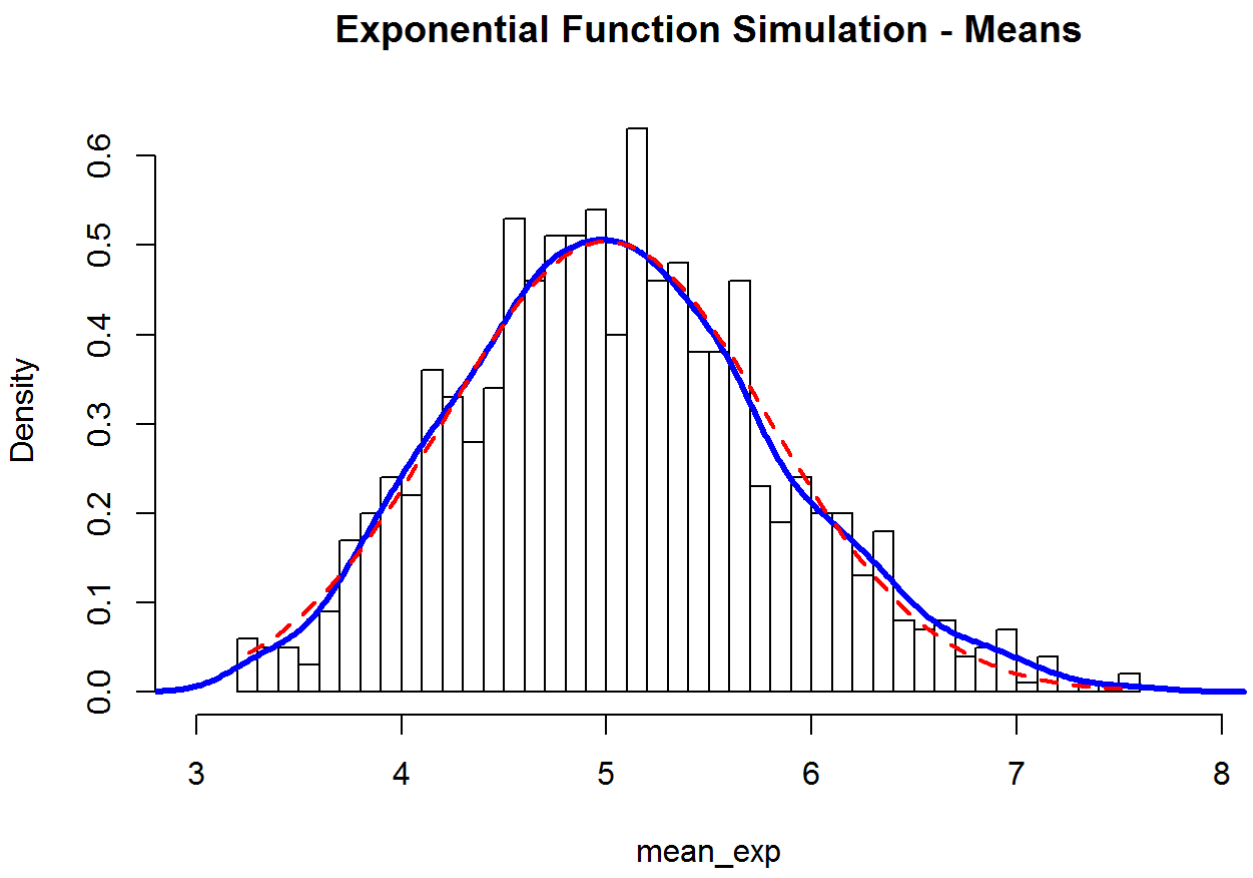

```
# Result for Q1 - We can clearly see from above results & plot, that sample mean and theoreti
cal mean are very close to each other.

# Q2 - Show how variable the sample is (via variance) and compare it to the theoretical varia
nce of the distribution

s_var <- var(mean_exp) # calculating sample variance
s_var
```

```
## [1] 0.6026047
```

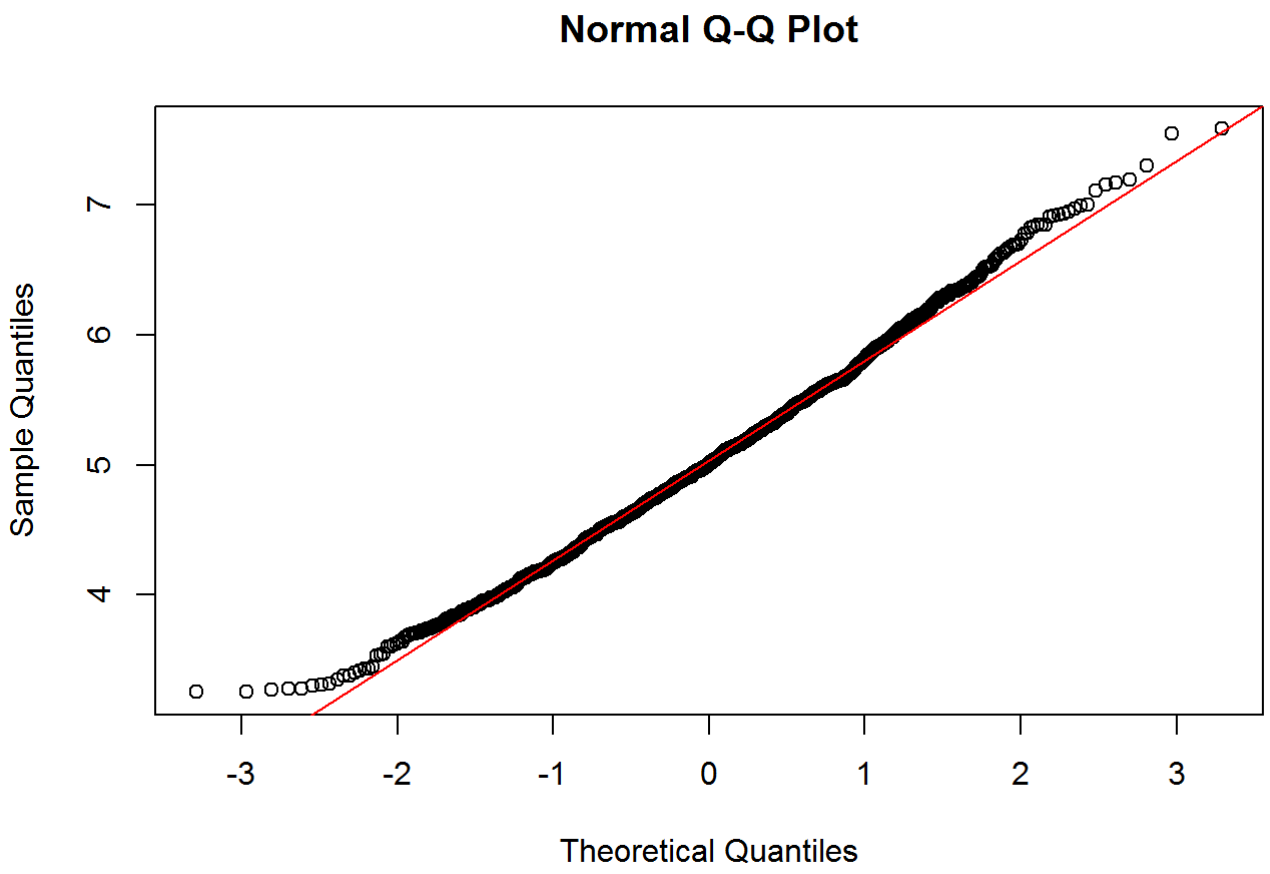
```
t_var <- ((1/lambda)/sqrt(nos))^2 # calculating theoretical variance
t_var
```

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

```
# Result for Q2 - We can clearly see from above results, that sample variance and theoretical
 variance are very close to each other.

# Q3 - Show that the distribution is approximately normal

hist(mean_exp, prob=TRUE,col="white", main = "Exponential Function Simulation - Means",breaks
=40, xlim = c(3,8)) # plotting histogram for means
lines(density(mean_exp),lwd=3,col="blue") # plotting the distribution curve
# plotting the normal distribution curve
a <- seq(min(mean_exp), max(mean_exp), length=2*nos)
b <- dnorm(a, mean=1/lambda, sd=sqrt(((1/lambda)/sqrt(nos))^2))
lines(a, b, pch=22, col="red", lwd=2, lty = 2)
```




```
# comparing the quantile plots

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




```
# Result for Q3 - From the above two plots, we can safely conclude that the distribution of m
ean of 40 exponentials is very close to normal distribution.
```