Final Project- Statistical Inference (Part 1)

06/27/2019

Part 1: Simulation Exercise

Overview:

In this project we investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can 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. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

The following discussions will be addressed in this report:

1. Compare the sample mean 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

Simulation and Comparisons:

First we show the sample mean and compare it with theoritical mean of exponential distribution. ### Loading libraries and Simulating 1000 times exponential distribution with 40 samples

```
library("data.table")
library("ggplot2")
```

```
# set seed for reproducability
set.seed(30)

# set lambda to 0.2 as is noted in problem description
lambda <- 0.2

# n is sample size = 40
n <- 40

# set the number of simulations
simulations <- 1000

# simulate
simulated_exp <- replicate(simulations, rexp(n, lambda))

# calculate mean of exponentials (the columns of the above matrix are corresponding to each simulation
means_exponentials <- apply(simulated_exp, 2, mean)
```

Question 1: Show the sample mean and compare it to the theoretical mean of the distribution.

```r
# sample mean
sample_mean <- mean(means_exponentials)
sample_mean
```

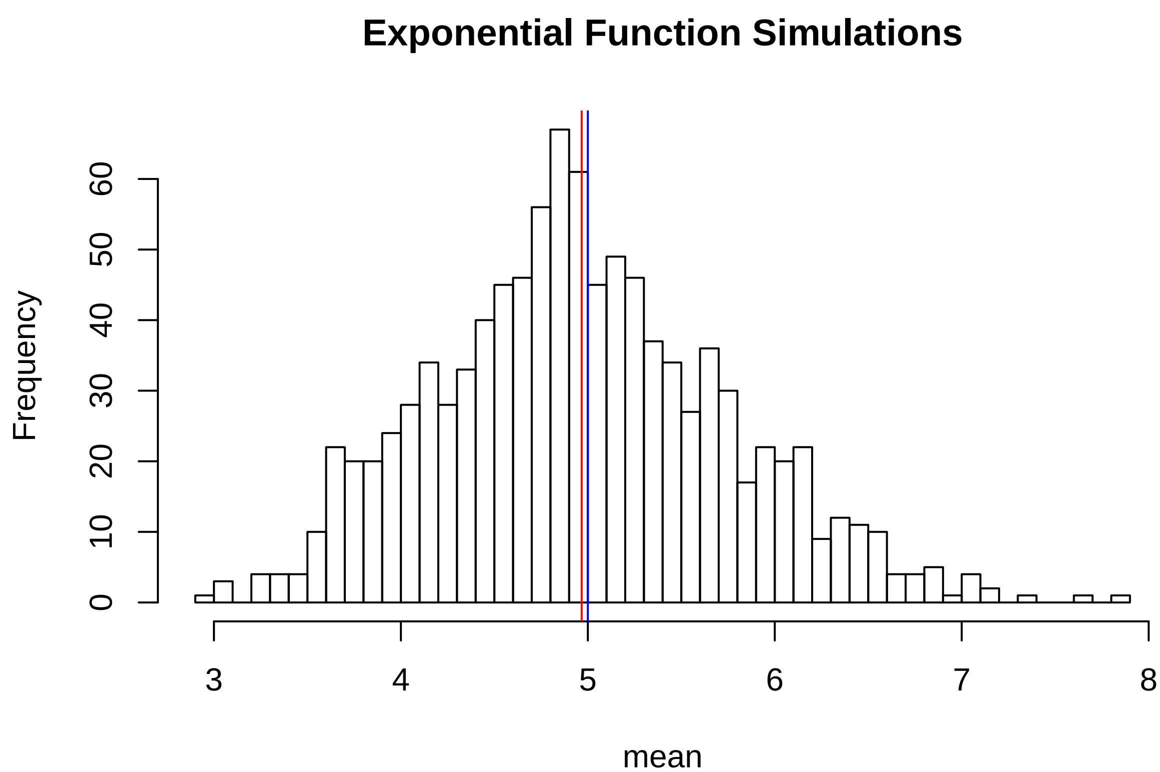
```
## [1] 4.967156
```

```r
# Theoretical mean
theoretical_mean <- 1/lambda
theoretical_mean
```

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

Histogram of sample mean for simulating 1000 times exponential distribution with 40 sample size (n=40), and lambda=0.2

```r
# visualization
hist(means_exponentials, breaks=n, xlab = "mean", main = "Exponential Function Simulations")
abline(v = sample_mean, col = "red")
abline(v = theoretical_mean, col = "blue")
```



As the above histogram plot shows, the sample mean is 4.993867 the theoretical mean 5. The center of distribution of means of 40 exponentials (center of sample means) is very close to the theoretical center of the distribution.

Question 2: Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
# calculating standard deviation of distribution
standard_deviation_dist <- sd(means_exponentials)
standard_deviation_dist
```

```
## [1] 0.784799
```

```
# calculating variance of distribution
variance_dist <- standard_deviation_dist^2
variance_dist
```

```
## [1] 0.6159094
```

```
# calculating standard deviation from analytical expression for exponential distribution
standard_deviation_theory <- (1/lambda)/sqrt(n)
standard_deviation_theory
```

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

```
# variance from analytical expression for exponential distribution
variance_theory <- ((1/lambda)/sqrt(n))^2
variance_theory
```

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

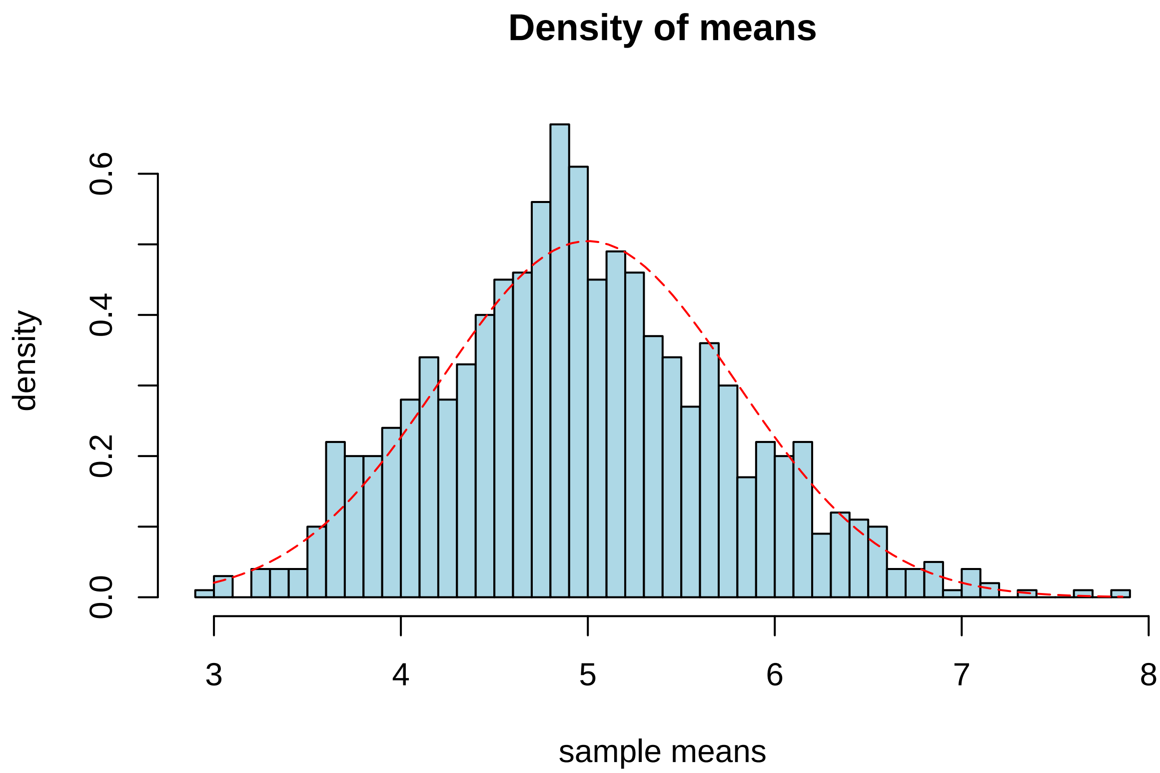
As the above calculation shows, the standard deviation for sample mean distribution and the analytical distribution follows as 0.785 and 0.791, respectively. Which are almost the same.

The theoretical variance for exponential distribution as ((1/lambda)*(1/sqrt(n)))^2 = 0.62. and the variance for sample mean distribution is 0.615 which again are both almost equal.

Question 3: Show that the distribution is approximately normal.

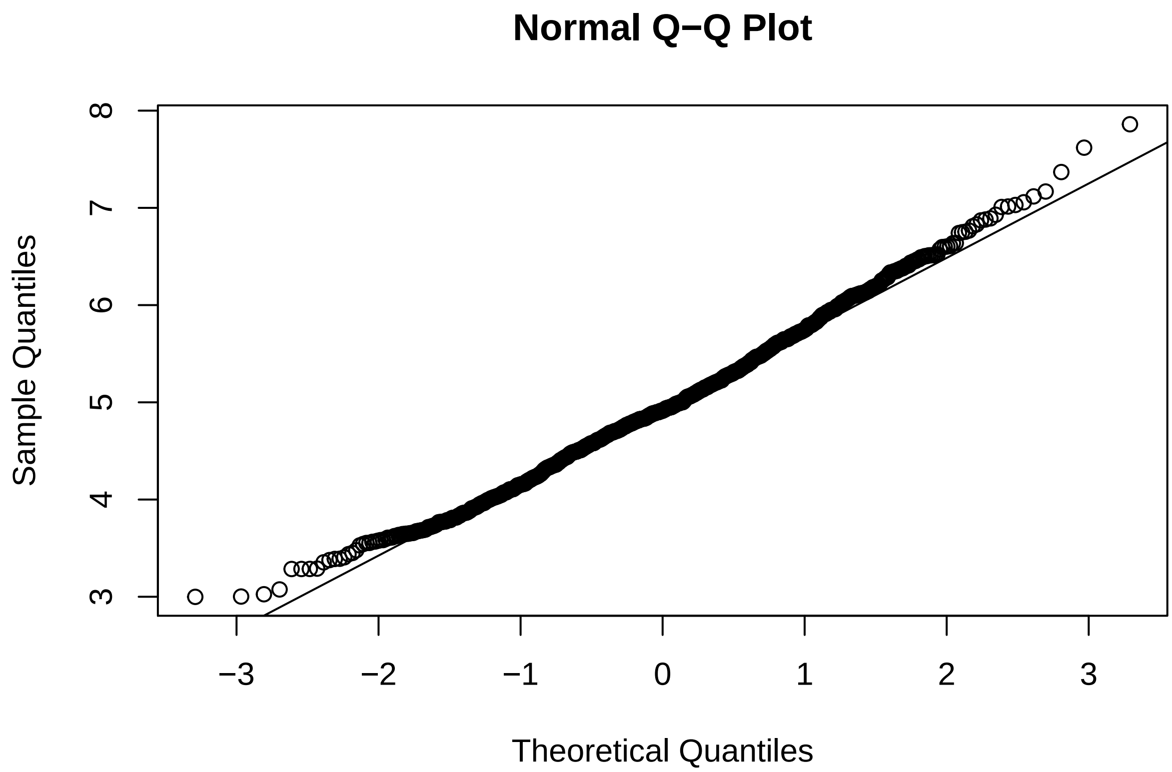
```
#draw a histogram of density for sample means
hist(means_exponentials,breaks=n, prob=T,col="lightblue",xlab = "sample means", ylab="density",main="Den

# fit a density plot to data
x <- seq(min(means_exponentials), max(means_exponentials), length=100)
y <- dnorm(x, mean=1/lambda, sd=(1/lambda/sqrt(n)))
lines(x, y, pch=16, col="red", lty=5)
```



Below we show a quantile-qualtile (Q-Q) plot to demonstrate the fact that due to Central Limit Theorem (CLT) the distribution of sample mean for n=40 exponentials converges to normal distribution.

```
# compare the distribution of averages of 40 exponentials to a normal distribution
qqnorm(means_exponentials)
qqline(means_exponentials)
```


Normal Q−Q Plot
Sample Quantiles
3
4
5
6
7
8
−3
−2
−1
0
1
2
3
Theoretical Quantiles