Statistical Inference Course Project 1- Simulations

Joel DSilva

Tuesday, May 19, 2015

Overview The objective 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. This will compare the sample and theoretical means, std deviation and variance. It will also show the distribution to be Normal.

```
## Warning: package 'ggplot2' was built under R version 3.1.3
```

Set the variables. Setting a seed value so the output can be regenerated.

Structure of the generated random sample set is as below

```
lambda <- 0.2
n <- 1000
samples <- 40
set.seed(26)
generate_random <- replicate(n, rexp(samples, rate=lambda))

# Print structure of the simulation
str(generate_random)
```

```
##  num [1:40, 1:1000] 17.64 4.26 1.08 6.01 5.28 ...
```

Now we calculate the means, standard deviance and variance

```
#calculate  means of the exponential samples
exp_means<- apply(generate_random,2,mean)

#calculate standard deviation of the exponential samples
exp_sd<- sd(exp_means)

#calculate variance of the exponential samples
exp_var<- exp_sd ^ 2
```

1. Show the sample mean and compare it to the theoretical mean of the distribution. ***Sample mean and Theoretical Mean is printed below respectively***

```
mean(exp_means)
```

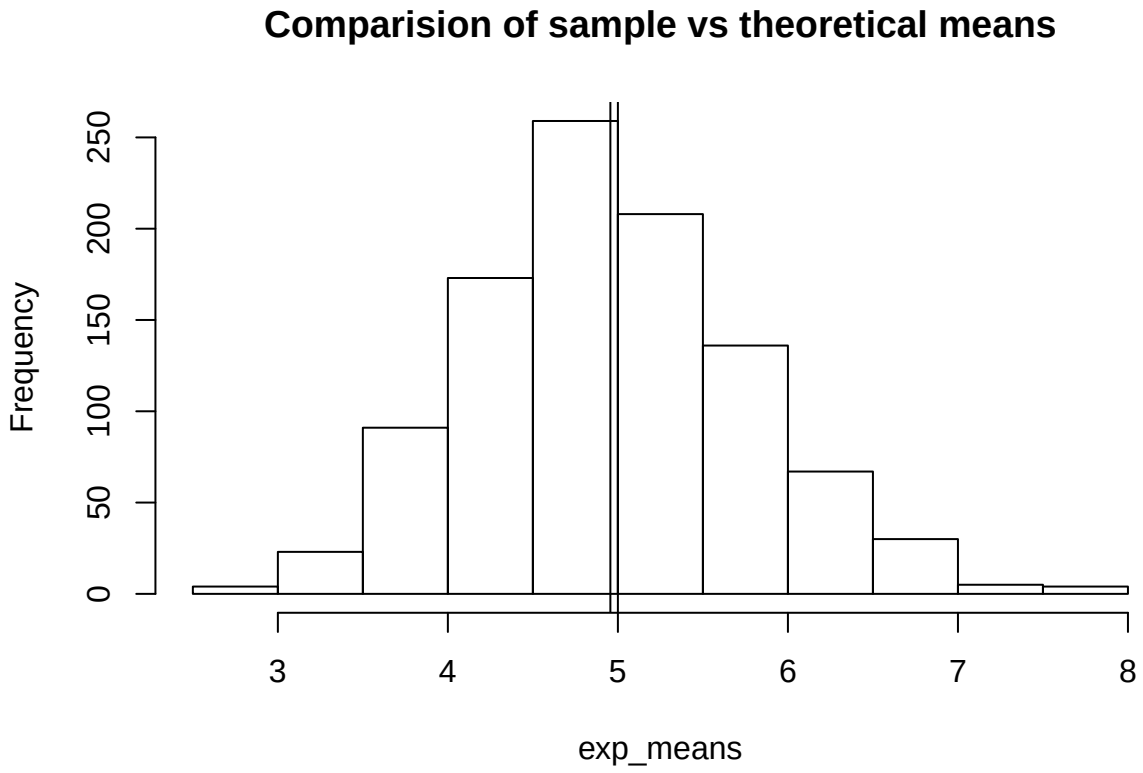
```
## [1] 4.955705
```

```
# Theoretical mean
1/lambda
```

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

The sample mean comes to around 4.9 and the theoretical is at 5. These values indicate that the sample and theoretical mean are very close to each other. For infinite observations the sample and theory should match. ***In figures the sample mean vs theoretical looks likebelow. The left vertical is the sample mean and the right vertical line is the theoretical mean.***

```
hist(exp_means,main="Comparision of sample vs theoretical means")
abline(v=mean(exp_means))
abline(v=1/lambda)
```



2.Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution. ***Sample deviation and Theoretical std deviation are printed below respectively***

```
exp_sd
```

```
## [1] 0.8159913
```

```
(1/lambda)/(sqrt(samples))
```

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

Sample variance and Theoretical variance are printed below respectively

```
exp_var
```

```
## [1] 0.6658418
```

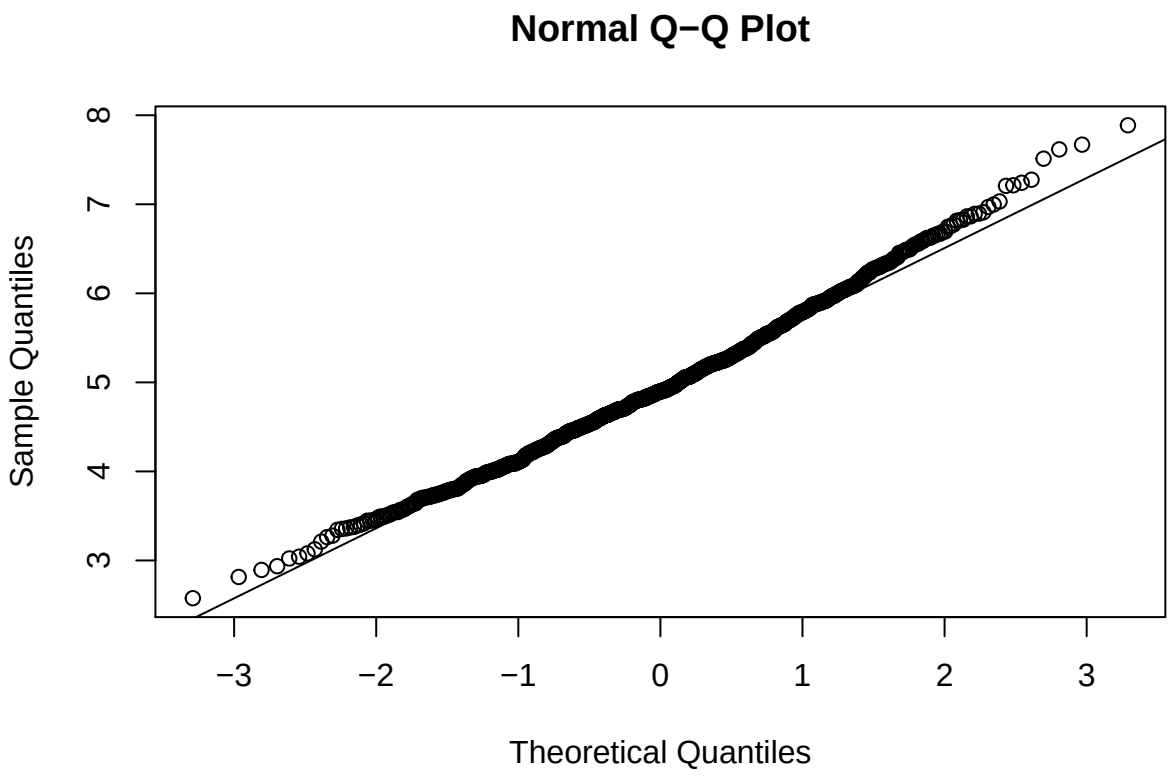
```
((1/lambda)/(sqrt(samples)))^2
```

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

The standard deviation of the samples turns out to be 0.81. THe theoretical standard deviation has turned out to be 0.79.

The sample variance has turned out to be 0.66 and theoretical variance is 0.62. ***This shows that the SD and Var are very close indicating a sufficient sample would provide the same values.***

3.Show that the distribution is approximately normal.



As per this the sample and theoretical points are aligned closedly. So a sufficiently large number of samples would indicate that the data would overlap and meet the premise of the CLT -> that the distribution would become a normal distribtion.