Statistical Inference Course Project -part 1

Nidhi Shrivastava

Wednesday, March 18, 2015

In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials.

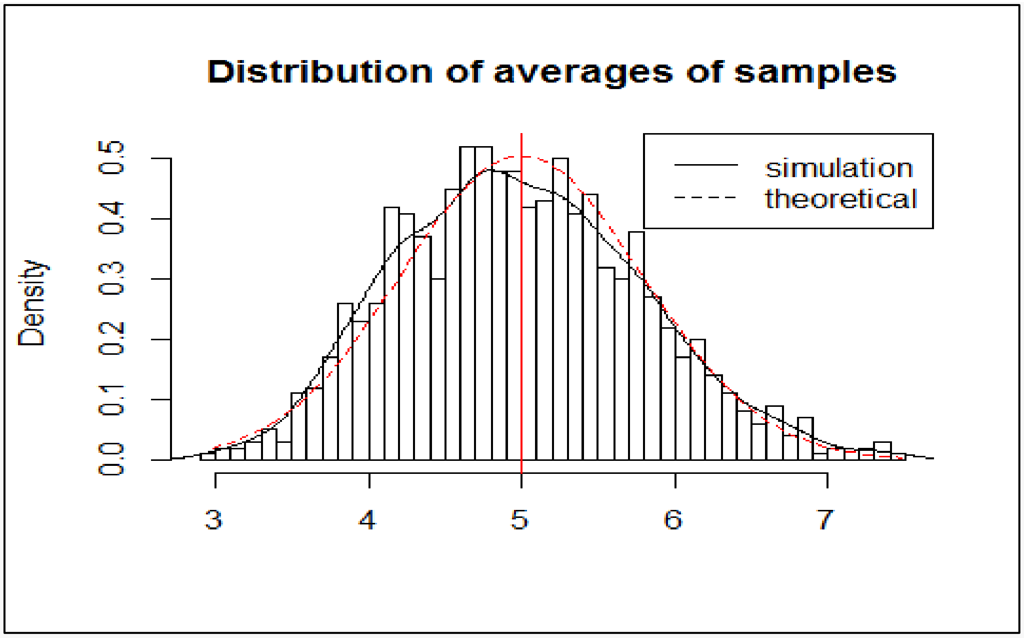
```
## set seed for reporducibility
set.seed(3)

## set lambda to 0.2
lambda<-0.2

## 1000 simulations
nsims<-1000

## 40 exponential samples
n<-40

## see the number of simulation in a matrix form
sim<-matrix(rexp(nsims*n, rate=lambda), nsims, n)

## calculate means
row_means<-rowMeans(sim)
```

Question 1

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

```
#distribution mean
theoretical_mean<-mean(row_means)
theoretical_mean
## [1] 4.987876

#sample mean
 sample_mean<-1/lambda
 sample_mean
## [1] 5

#visualization - draw a histogram
hist(row_means, breaks=50, prob=TRUE, main="Distribution of averages of samples, drawn
from exponential distribution with lambda=0.2", xlab="")

## density of the averages sample
```

```
lines(density(row_means))

##theoretical center of distribution
abline(v=1/lambda, col='red')
```



Answer 1

The analytics mean is 4.993867 the theoretical mean 5. The center of distribution of averages of 40 exponentials is very close to the theoretical center of the distribution.

Question 2

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

```
##standard deviation of distribution
s<-sd(row_means)
s
## [1] 0.7834318

##standard deviation from the sample

s1<- (1/lambda)/sqrt(n)
 s1
##[1] 0.7905694

##variance of distribution

var_dist<-s^2
 var_dist
```

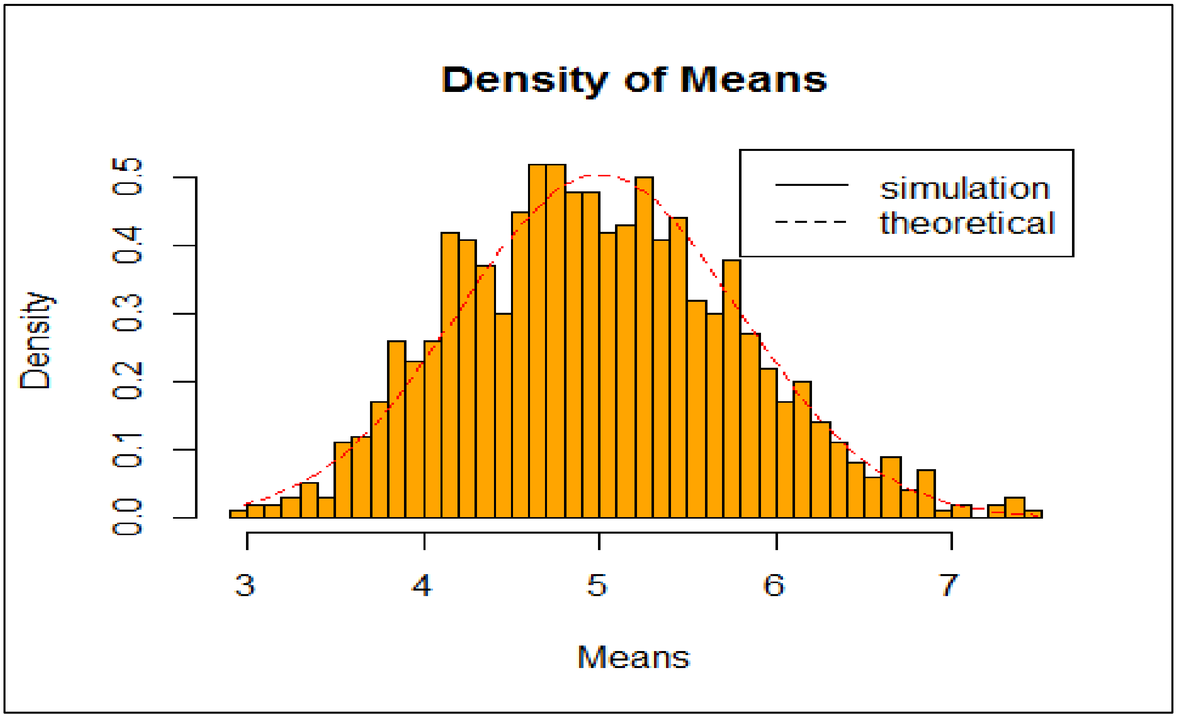
```
## [1] 0.6137653

##variance of theoretical mean
var_theory<-((1/lambda)*(1/sqrt(n)))^2
var_theory
## [1] 0.625
```

Answer 2 Standard Deviation of the distribution is 0.7834318 with the theoretical SD calculated as 0.7905694. The Theoretical variance is calculated as $(\frac{1}{\text{lambda}} * \frac{1}{\text{sqrt(n)}})^2$ = 0.625. The actual variance of the distribution is 0.6137653.

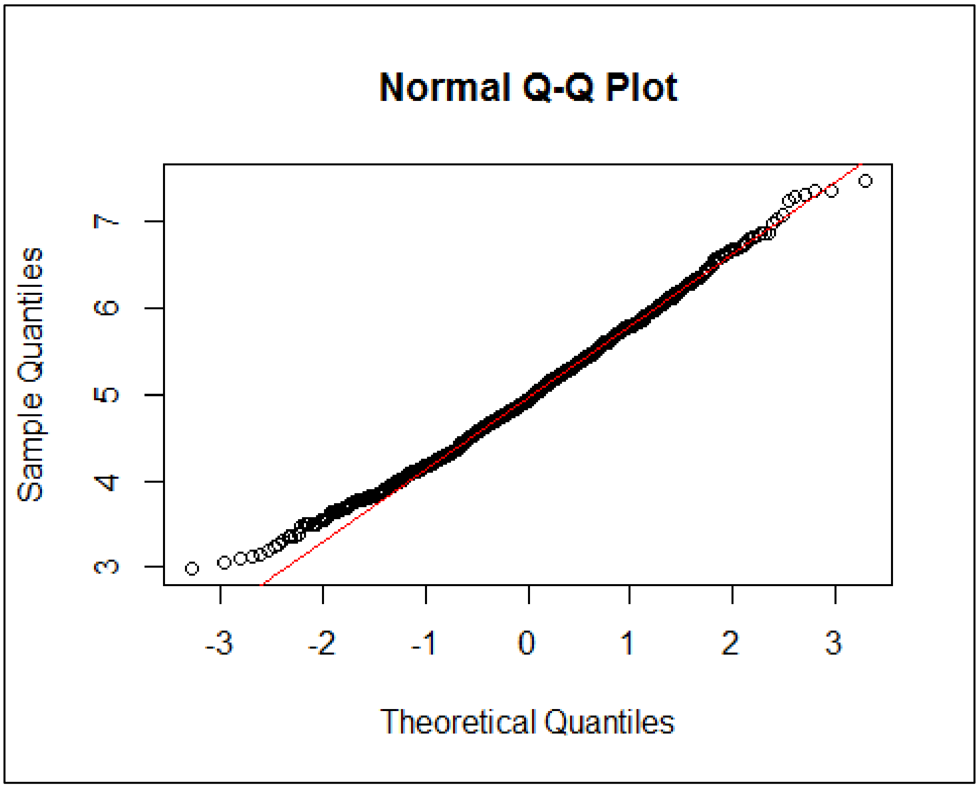
Question 3

Show that the distribution is approximately normal.

```
xfit<-seq(min(row_means), max(row_means), length=100)
yfit<-dnorm(xfit, mean=1/lambda, sd=((1/lambda)/sqrt(n)))
hist(row_means, breaks=n, prob=TRUE, col="orange", xlab="Means", main="Density of
Means", ylab="Density")
lines(xfit, yfit, pch=22, col="red", lty=2)
## add legend
legend('topright', legend=c("simulation","theoretical"), lty=1:2, col("black", "red"))
```



```
##compare the distribution of averages of 40 exponentials to a normal distribution

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



Answer 3:
Due to the central limit theorem (CLT), the distribution of averages of 40 exponentials is very close to a normal distribution.