Data Science Course with R:Statistical Inference:Course Project

Question 1:Simulation Exercise

Sagnik Das

6/24/2020

Generating a exponential distribution:

We first set seed and the define the parameters of the distribution aacording to the information provided in the question. Then generate the simulated data named sim_data.

```r
set.seed(100)

lambda<- 0.2

### number of samples

n<- 40

### number of simulations

sim<- 1000

### simulated data

sim_data<- replicate(sim, rexp(n,lambda))
```

Next, we generate the distribution of the mean of the simulated sample:

```r
mean_simp<- apply(sim_data,2,mean)
```

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

```r
### Smaple Mean

sample_mean<- mean(mean_simp)

### Population Mean

pop_mean<- 1/lambda

print(c(paste("Sample Mean:", sample_mean),paste("Population (Theoretical Mean):", pop_mean)))
```

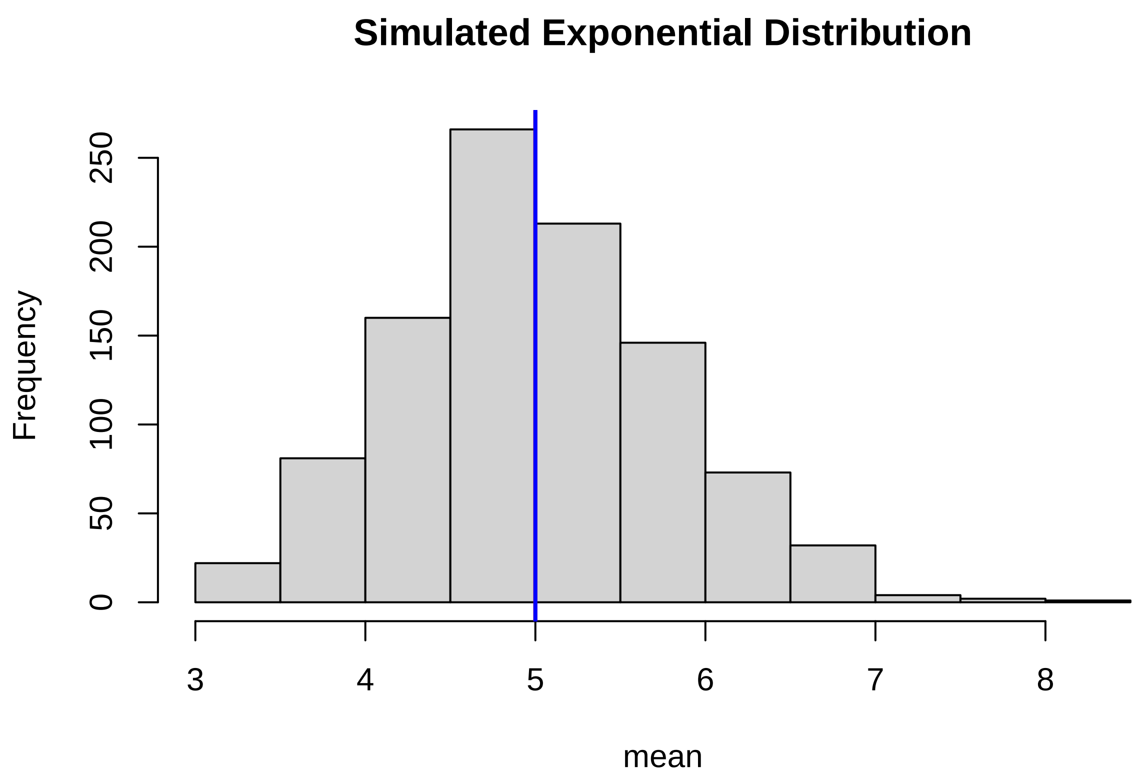
Answer:

```
## [1] "Sample Mean: 4.9997019268744"     "Population (Theoretical Mean): 5"
```

As can be seen from the means, the mean of the distribution of the sample mean is very close to the actual theoretical mean of the distribution.

```
hist(mean_simp, xlab = "mean", main = "Simulated Exponential Distribution")
abline(v=4.999702, col="red", lwd=2)
abline(v=5, col="blue",lwd=2)
```

Visualizing the sampling and the population distributions



From the figure, it can be seen that the mean of the distribution of the sample mean and the theoretical mean is so close to each other, that they are overlayed over one another.

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

```
### standard deviation of the means of the simulated sample distribution

std_sim_dist<- sd(mean_simp)

### Standard deviation of the population

std_pop_dist<- (1/lambda)/sqrt(n)

print(c(paste("Standard deviation of sample:", std_sim_dist),paste("Standard deviation of population:",
```

Answer:

```
## [1] "Standard deviation of sample: 0.802025077464672"
## [2] "Standard deviation of population: 0.790569415042095"
```

```
### Variance of the means of simulated sample distribution

var_sim_dist<- (std_sim_dist)^2

### Variance of the population distribution

var_pop_dist<- ((1/lambda)*(1/sqrt(n)))^2

print(c(paste("Variance of sample:", var_sim_dist),paste("Variance of population:", var_pop_dist)))
```

```
## [1] "Variance of sample: 0.643244224882213"
## [2] "Variance of population: 0.625"
```

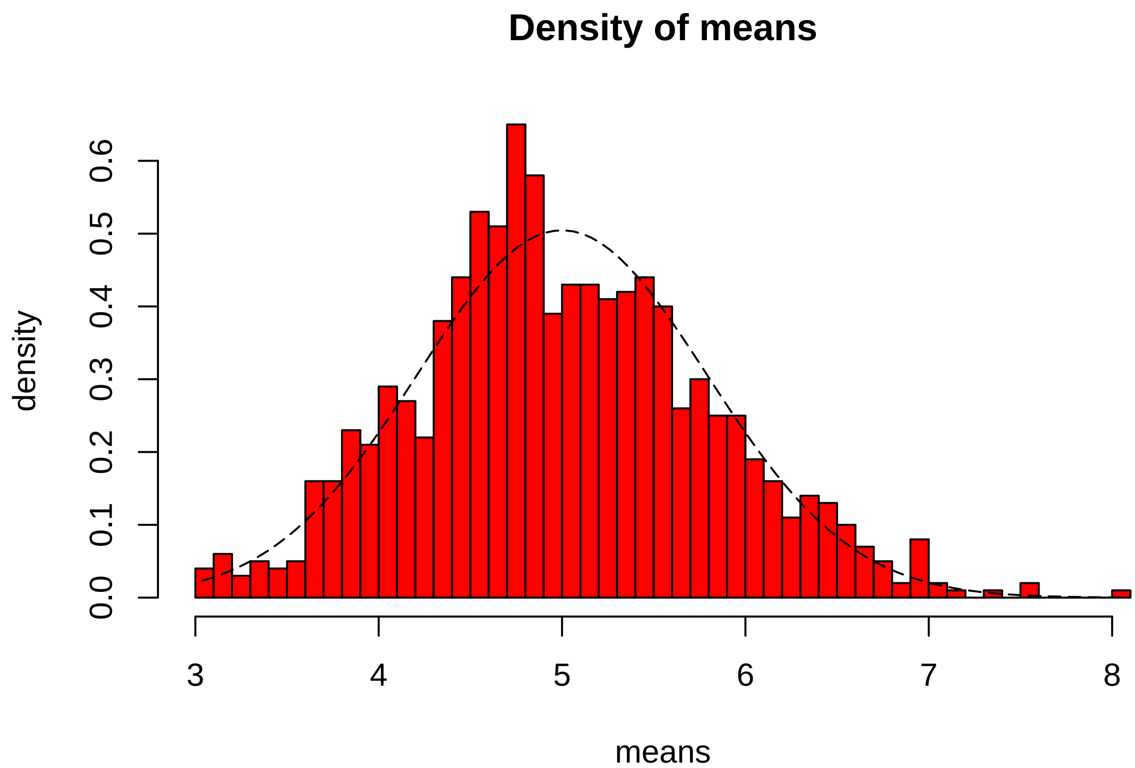
Just like the means, the standard ddeviation and the variance of the sampling distribution and the population distribution show how close they are to each other.

Show that the distribution is approximately normal.

```
xfit <- seq(min(mean_simp), max(mean_simp), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(n)))

hist(mean_simp,breaks=n,prob=T,col="red",xlab = "means",main="Density of means",ylab="density")
lines(xfit, yfit, pch=20, col="black", lty=5)
```

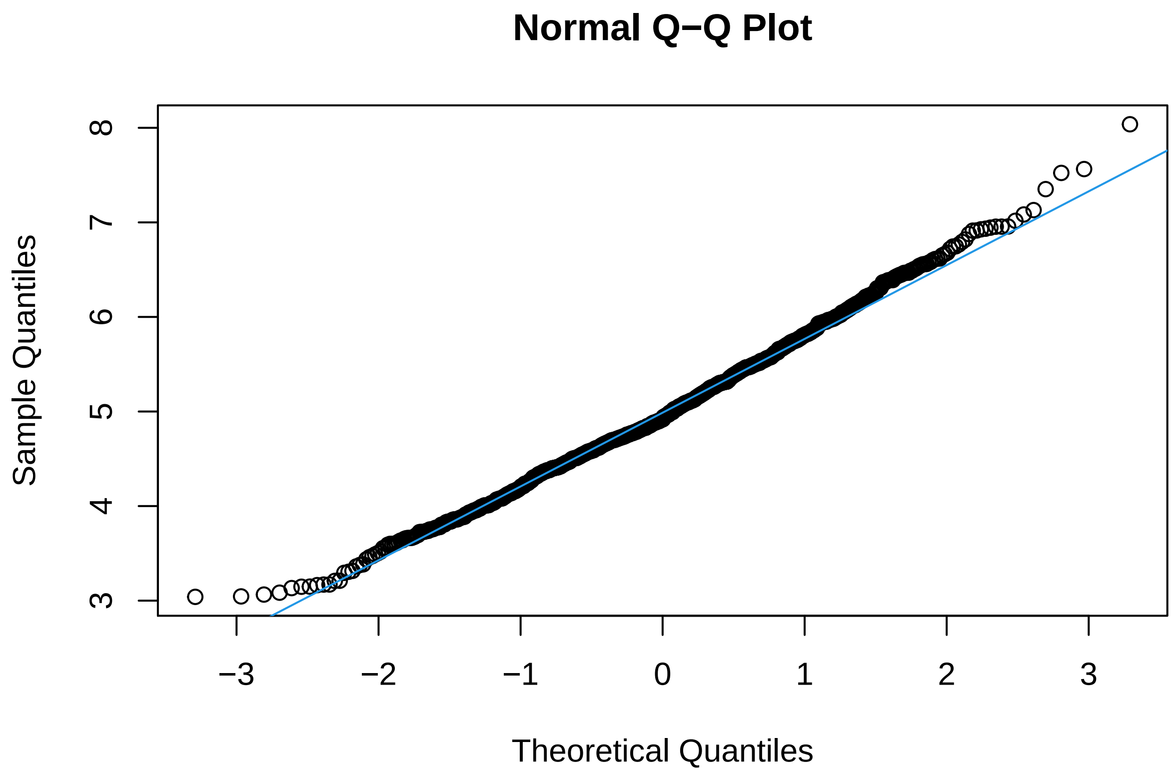
Answer:



As can be seen clearly from the figure, the distribution of the means of the sample is approximately normal.

```
### Comparing the average of the 40 exponentials to a normal distribution

qqnorm(mean_simp)
qqline(mean_simp,col=4)
```



This figure shows the quantile wise similarity between the theoretical and the sampling distribution.