Statistical Inference Project Part 1

Rajaram

01/08/2020

Overview

The objective of this project is to investigate the exponential distribution in R and compared to central limit theorem.

1. Show the sample mean and compare it 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.

Simulations

In R, We simulate exponential distribution through rexp function by passing number of samples and lambda. As given in the problem statement, lambda is set to 0.2 and size of the sample is 40.

```
set.seed(12)
lambda <- 0.2
samples <- 40
simulations <- 1000

sim_exponentials <- replicate(simulations,rexp(samples,lambda))
```

Sample Mean versus Theoretical Mean

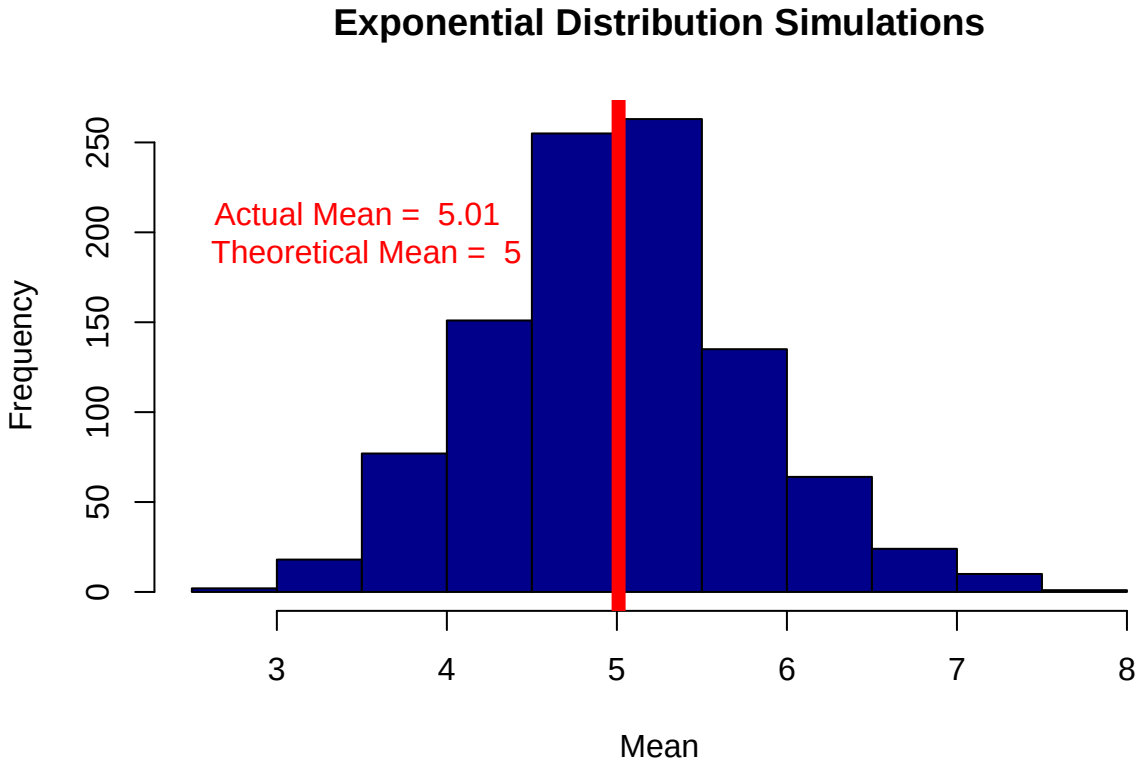
```
mean_exponentials <- apply(sim_exponentials,2,mean)

sample_mean <- mean(mean_exponentials)
sample_mean
```

```
## [1] 5.010015
```

```
theoretical_mean <- 1/lambda
theoretical_mean
```

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

```
hist(mean_exponentials, xlab = "Mean",
     main = "Exponential Distribution Simulations",col="darkblue")
abline(v = sample_mean, col = "red",lwd="7")
text(3.5,200,paste("Actual Mean = ",round(sample_mean,2),
                   "\n Theoretical Mean = ",
                   theoretical_mean), col="red")
```



Conclusion:

```
The sample mean rounded to 2 decimal places is 5.01 and the theoretical mean is 5.
The center of distribution is very close to the theoretical center of the distribution.
```

Sample Variance versus Theoretical Variance

```
sd_dist <- sd(mean_exponentials)
sd_dist
```

```
## [1] 0.7740594
```

```
sd_theory <- (1/lambda)/sqrt(samples)
sd_theory
```

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

```
variance_dist <- sd_dist^2
variance_dist
```

```
## [1] 0.599168
```

```
variance_theory <- ((1/lambda)*(1/sqrt(samples)))^2
variance_theory
```

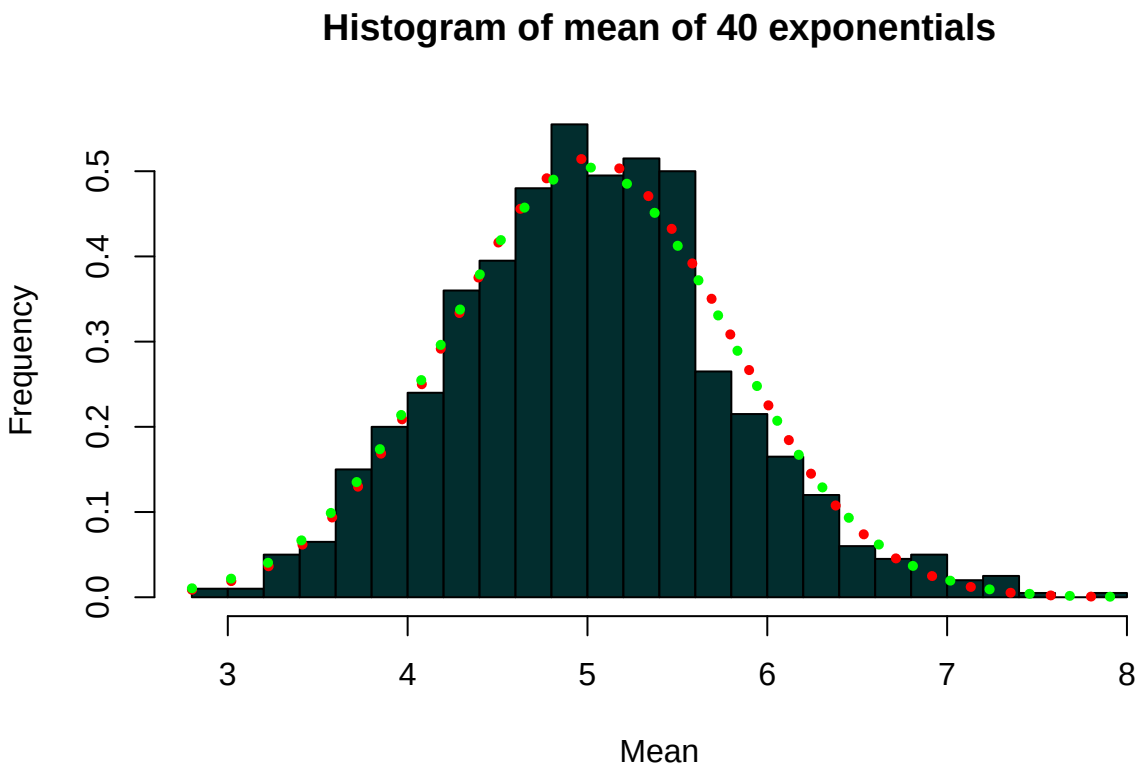
```
## [1] 0.625
```

Conclusion:

```
 Variance of the distribution is 0.599168 and the theoretical variance is calculated as 0.625.
 Variances are in similar range.
```

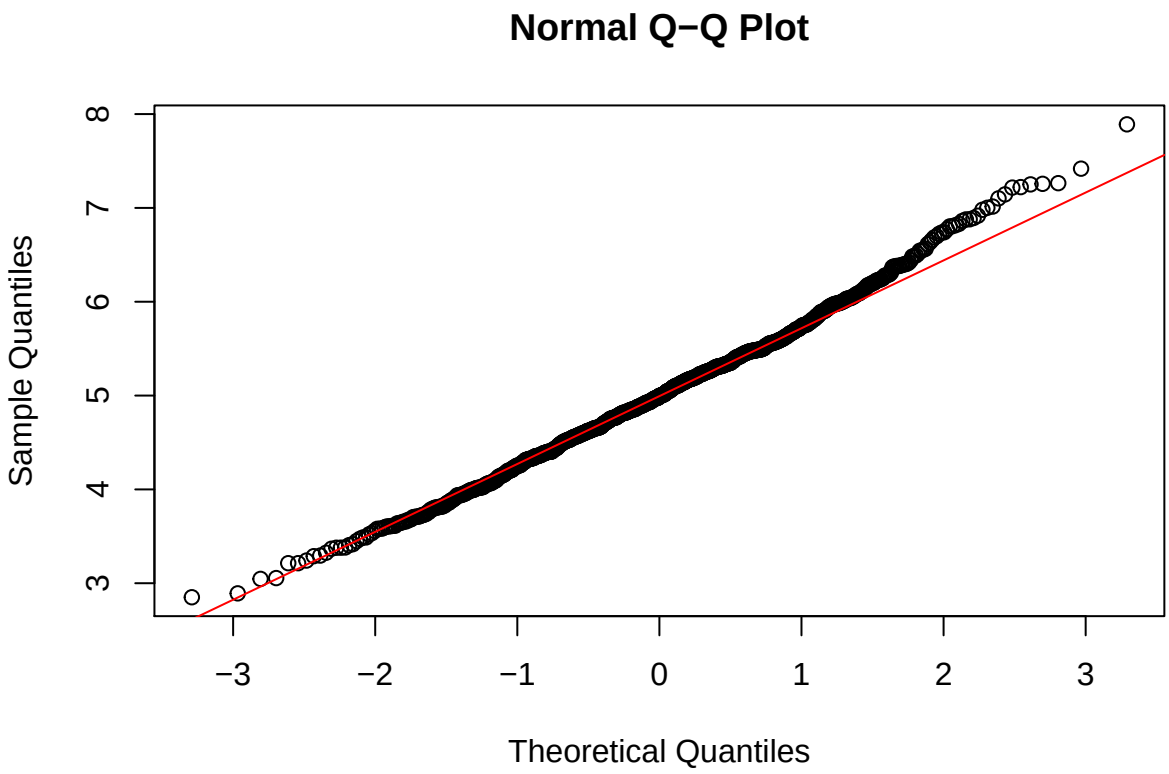
Distribution

```
hist(mean_exponentials,breaks=20,prob=TRUE,
     xlab="Mean", ylab="Frequency",
     main="Histogram of mean of 40 exponentials",
     col="#022d2e")
curve(dnorm(x, mean=sample_mean, sd=sd_dist),
      col="red", lwd="5",
      lty = "dotted",
      add=TRUE)
curve(dnorm(x, mean=theoretical_mean, sd=sd_theory),
      col="green", lwd="5",
      lty = "dotted",
      add=TRUE)
```



This plot shows that histogram is close to normal distribution.

```
qqnorm(mean_exponentials, main="Normal Q-Q Plot",
       xlab="Theoretical Quantiles",
       ylab="Sample Quantiles")
qqline(mean_exponentials,
       col="red")
```



Conclusion:

Q-Q plot shows that Sample Quantiles matches with theoretical Quantiles. We can conclude that distribution is normal.