Investigate the exponential distribution in R and compare it with the Central Limit Theorem

Kishibe Rohan

July 12,2020

Overview

The purpose of this data analysis is to investigate the exponential distribution in R and compare it to the Central Limit Theorem. For this analysis,lambda will be set to 0.2 and the distribution of averages of 40 exponentials over 1000 simulations will be investigated.

Simulations

Set the parameters for simulations,namely lamba,exponentials and seed.

```
#set seed for reproducibility
set.seed(1331)
lambda <- 0.2

#number of samples & simulations
n <- 40
simulations <- 1000

#Simulate
sim_exp <- replicate(simulations,rexp(n,lambda))

#Calculate mean
means_exp <- apply(sim_exp,2,mean)
```

Question 1

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

```
sim_mean <- mean(means_exp)
sim_mean
```

```
## [1] 4.982624
```

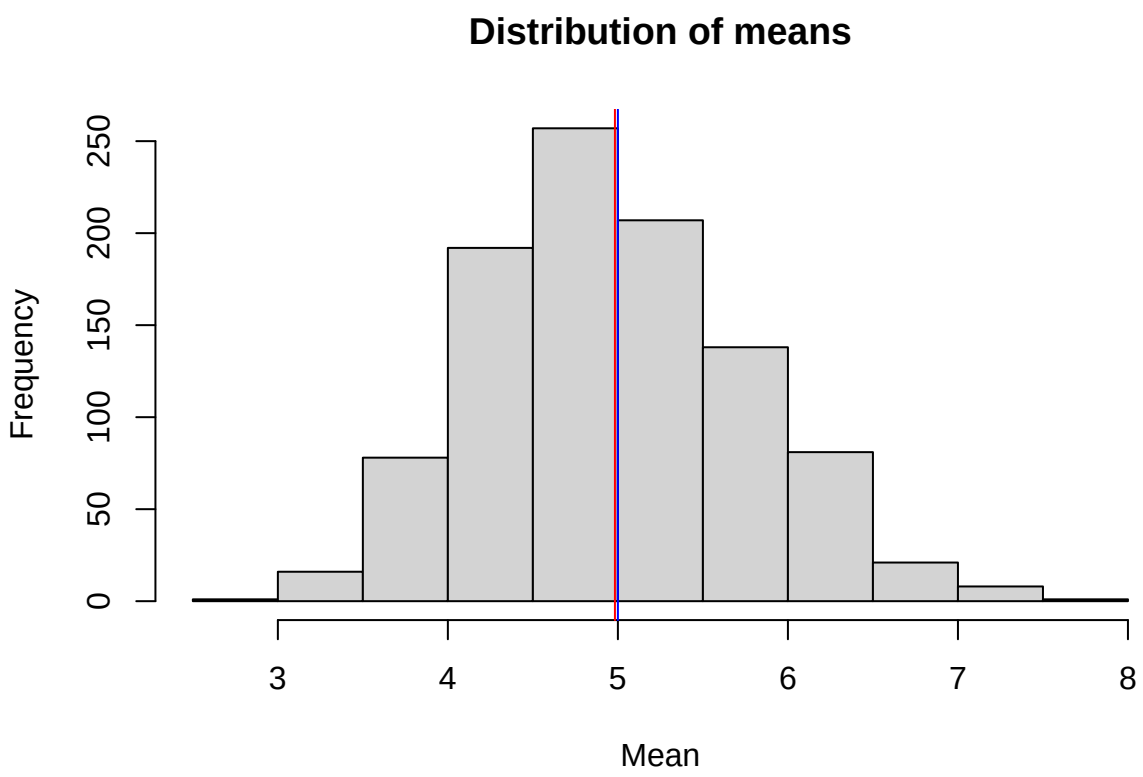
```
theoretical_mean <- 1/lambda
theoretical_mean
```

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

Visualization

```
library(ggplot2)
hist(means_exp,xlab="Mean",main="Distribution of means")
```

```
abline(v=sim_mean,col="red")
abline(v=theoretical_mean,col="blue")
```



The analytics mean is 4.982624 and the theoretical mean is 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 the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
#standard deviation & variance of sample
sd_sim <- sd(means_exp)
sd_sim
```

```
## [1] 0.7830017
```

```
var_sim <- sd_sim ^ 2
var_sim
```

```
## [1] 0.6130916
```

```
#theoretical standard deviation & variance
sd_theory <- (1/lambda)/sqrt(n)
sd_theory
```

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

```
var_theory <- sd_theory^2
var_theory
```

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

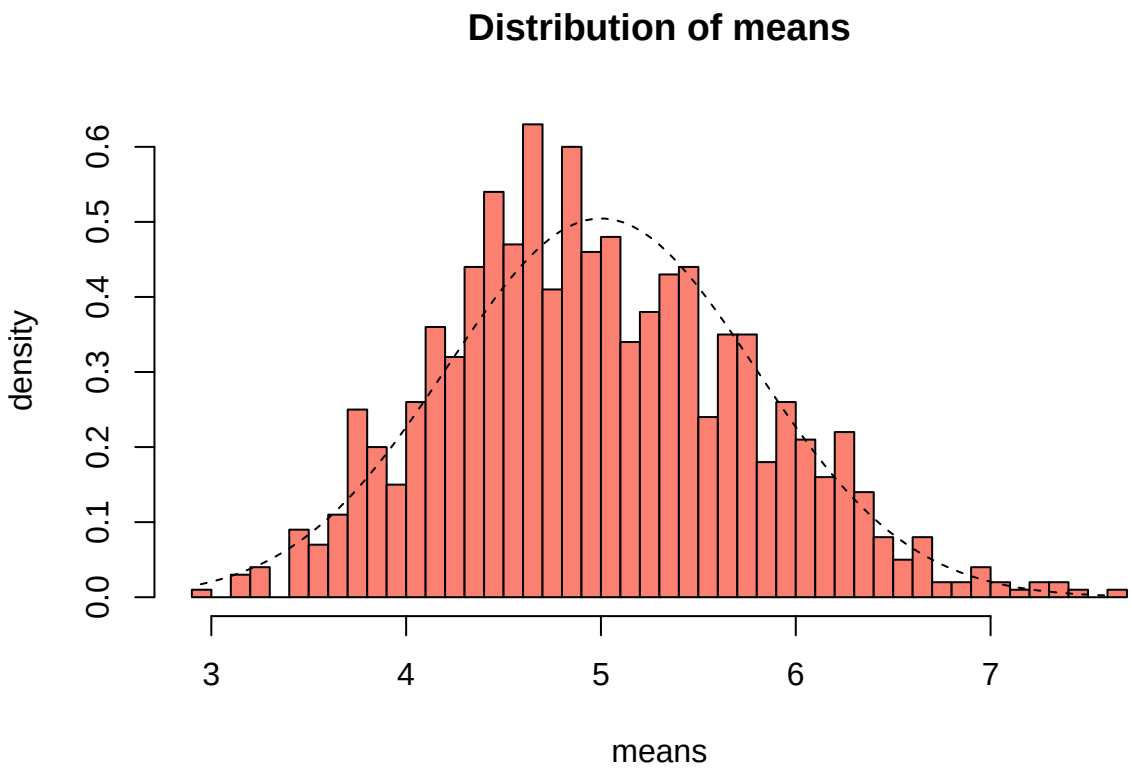
Standard Deviation of the distribution is 0.7830017 while the theoretical standard deviation is 0.7905694.

The theoretical variance is 0.625 while the actual variance of the distribution is 0.6130916

Question 3

Show that the distribution is approximately normal.

```
x_plt <- seq(min(means_exp),max(means_exp),
             length=100)
y_plt <- dnorm(x_plt,mean = 1/lambda,sd=(1/lambda/sqrt(n)))
hist(means_exp,breaks = n,prob=T,col="salmon",xlab="means",ylab="density",main="Distribution of means")
lines(x_plt,y_plt,pch = 18,col="black",lty = 2)
```



From the plot we observe that the distribution of sample means is approximately the same as the standard normal distribution,apparent from the bell-shaped nature of the curve.