Statistical inference assignment 1 - Simulation Exercise

preethi

9/27/2019

Overview

In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. We will use the function rexp(n,λ) where λ is the rate parameter. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. The λ is set to 0.2 for all simulations. The sample size is 40 exponentials and 1000 simulations will be run.

We will be illustrating via the simulation the properties of the distribution of the mean of 40 exponentials by answering the following questions

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

```r
##setting the seed
set.seed(10)

lambda <- 0.2
sample_size <- 40
simulations <- 1000

##running the simulation
simulation_1 <- NULL

for(i in 1:simulations)
  {
  simulation_1 <- c(simulation_1, mean(rexp(sample_size, lambda)))
  }
##calculating the mean of the simulation
mean(simulation_1)
```

```
## [1] 5.04506
```

Calculating the mean of the exponential distribution is $1/\lambda$ which is

```r
lambda <- 0.2
##mentioned in the question
mean_exp <- 1/lambda

mean_exp
```

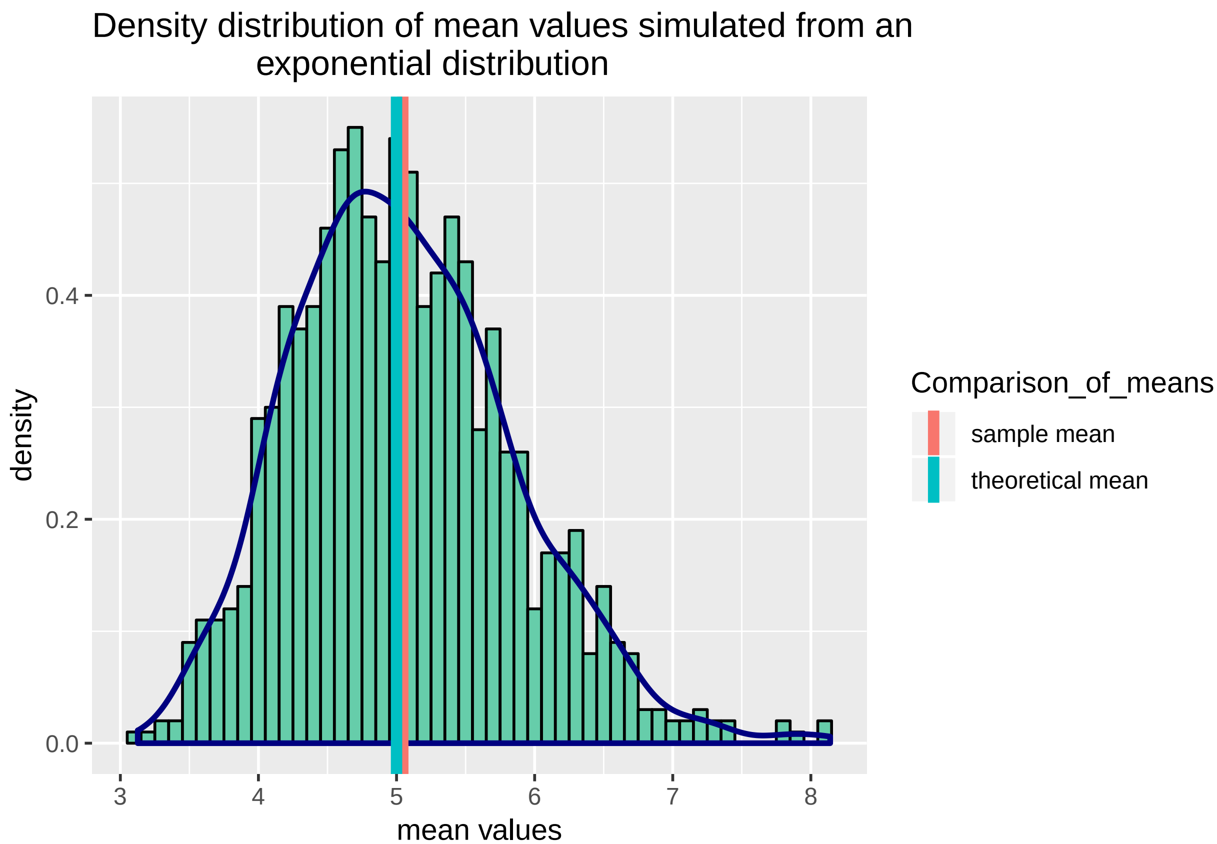
```
## [1] 5
```

As you can see the mean of the simulation is very close to the theoritical mean of the exponential distribution

Given below are the code and plot for the density distribution of the mean values

```r
##creating a data frame with the mean from simulation and theoretical mean
means <- data.frame (Comparison_of_means = c("sample mean","theoretical mean"),
                     values = c(mean(simulation_1), 1/lambda))
## creating a histogram with density distribution of means along with the intercepts for the two means
g2 <- ggplot(data.frame(simulation_1),aes(x=simulation_1))

g2 <- g2+ggtitle("Density distribution of mean values simulated from an
```

```r
                exponential distribution")+xlab("mean values")

g2 <- g2 + geom_histogram(binwidth = 0.1, colour = "black", fill = "aquamarine3",
                         aes(y=..density..))

g2 <- g2+geom_density(color="navyblue", size = 1)

g2 <- g2+geom_vline(data = means,aes(xintercept=values,
                                     col=Comparison_of_means),size=2)
print(g2)
```



Q2: Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution?

```r
var(simulation_1)
```

```
## [1] 0.6372544
```

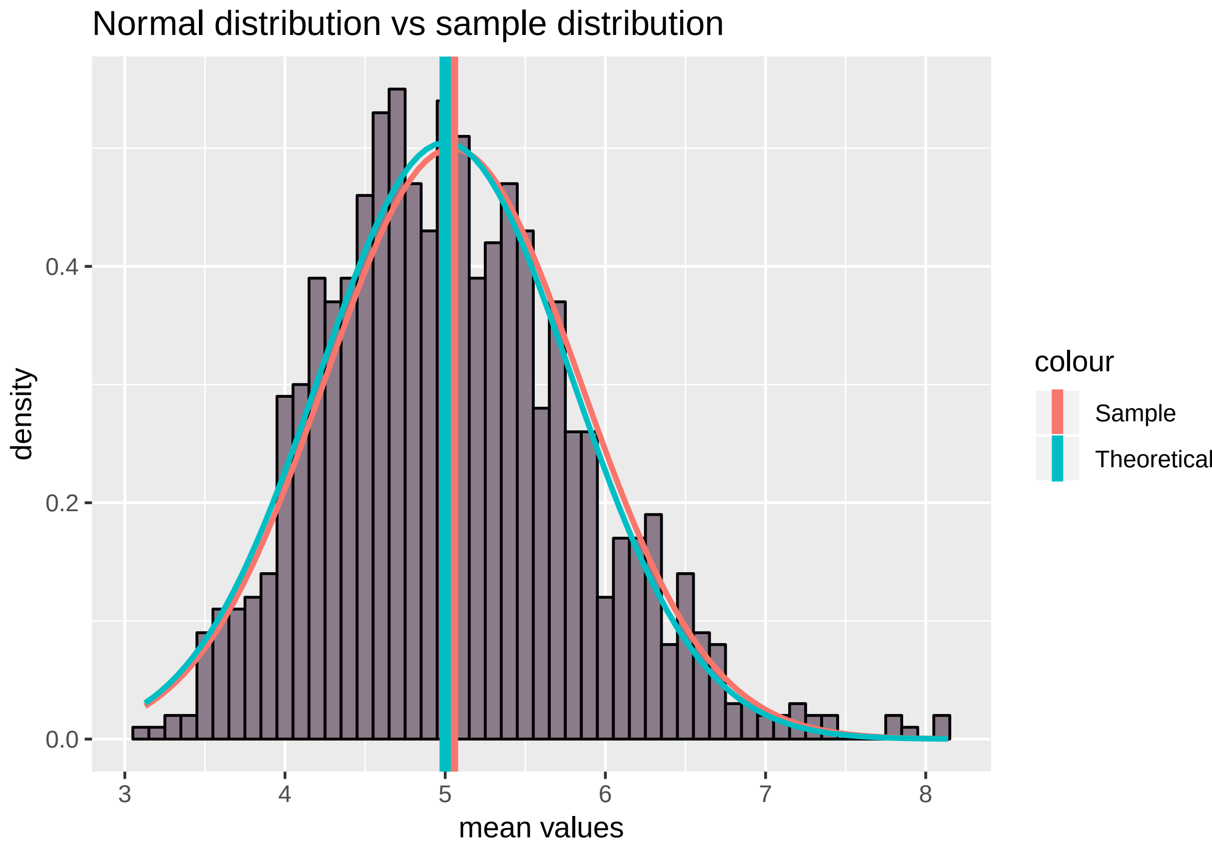
```r
std_dev <- 1/lambda/sqrt(sample_size)
variance_exp <- std_dev^2
variance_exp
```

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

As can be observed above the variance of the sample is very similar to the theoretical variance of the distribution

Q3) Show that the distribution is approximately normal

```r
##setting seed
set.seed(10)

## include all the data supplied
lambda <- 0.2
sample_size <- 40
simulations <- 1000

##creating a for loop to get all the means of the simulations
simulation_1 <- NULL

for(i in 1:simulations)
{
  simulation_1 <- c(simulation_1, mean(rexp(sample_size, lambda)))
}

##creating a data frame with the given means
means <- data.frame (Comparison_of_means = c("Sample","Theoretical"),
                     values = c(mean(simulation_1), 1/lambda))

g3 <- ggplot(data.frame(simulation_1),aes(x=simulation_1))
g3 <- g3+xlab("mean values")
g3 <- g3 + geom_histogram(binwidth = 0.1, colour = "black", fill = "thistle4",
                          aes(y=..density..))
g3 <- g3 + stat_function(fun = dnorm, args = list(mean = mean(simulation_1),
                                                  sd = sqrt(var(simulation_1))),
                         col = "#F8766D",size = 1)
g3 <- g3 + stat_function(fun = dnorm,args = list(mean=
                                                    1/lambda,sd =
                                                sqrt((1/lambda)^2/sample_size)),
                          col = "#00BFC4",size = 1, aes(colour = "Theoretical
                                                      mean"))
g3 <- g3 + ggtitle("Normal distribution vs sample distribution")
g3 <- g3+geom_vline(data = means,aes(xintercept=values,
                                       col=Comparison_of_means),size=2)
print (g3)
```



The sample & theoretical means and distribution are shown in the above plot.