Statistical Inference Course Project

Kaaltho

20/12/2020

Synopsis

The exponential distribution is going to be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. The parameter lambda is set to 0.2 for all of the simulations. This investigation considers the distribution of averages of 40 exponential.

Part 1: Simulation Exercise

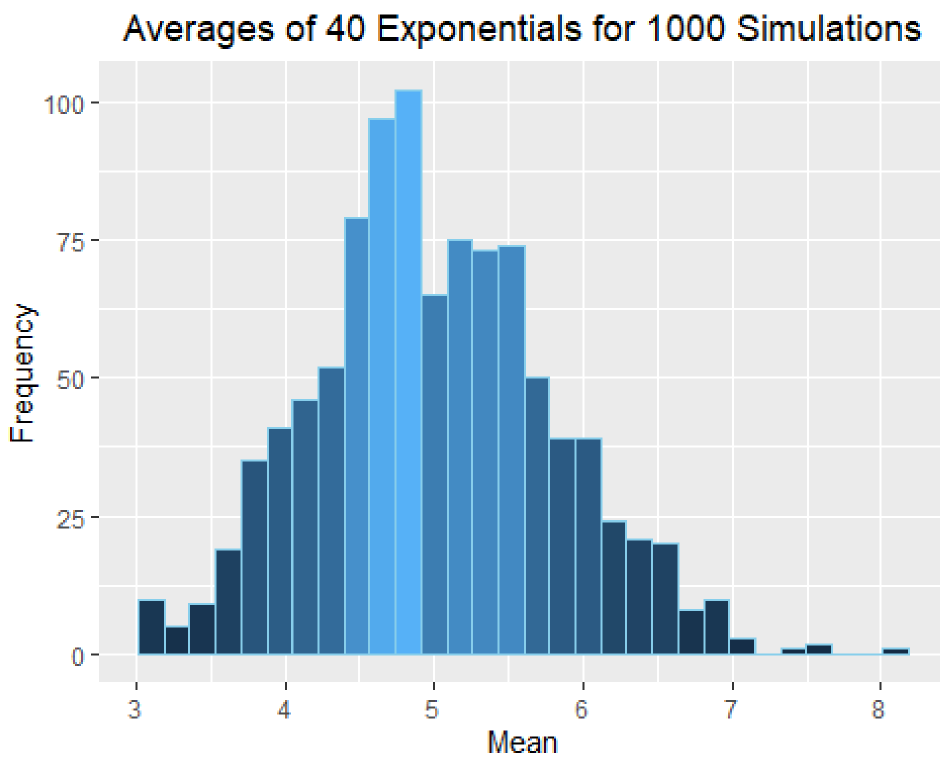
The first step is to simulate the variables that we want to compare.

```
#Oppening libraries
library(ggplot2)

#Assigning variables names, the specified data is the following:
lambda <- 0.2
n <- 40
simulations <- 1:1000
set.seed(100)

# Simulate the population
population <- data.frame(x=sapply(simulations, function(x) {mean(rexp(n,
lambda))}))

#The random varaibles looks like this:
head(population)

##          x
## 1 4.137412
## 2 6.051703
## 3 4.415869
## 4 4.404714
## 5 3.210413
## 6 5.475307
```

The next plot contains the distribution of the variables and it is possible to see if it follows some distribution.

```
ggplot(population, aes(x=x)) +
  geom_histogram(aes(y=..count.., fill=..count..),color="skyblue") +
  guides(fill=FALSE) +
```

```
  labs(title="Averages of 40 Exponentials for 1000 Simulations", y="Frequency", x="Mean")+
  theme(plot.title = element_text(hjust = 0.5))

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



It is possible to see how the data accumulates near 5, and also the shape of the data seems to be similar to a normal distribution.

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

```
# The mean and the theoretical mean of the distribution is as following:
sample.mean <- mean(population$x)
theoretical.mean <- 1/lambda
```

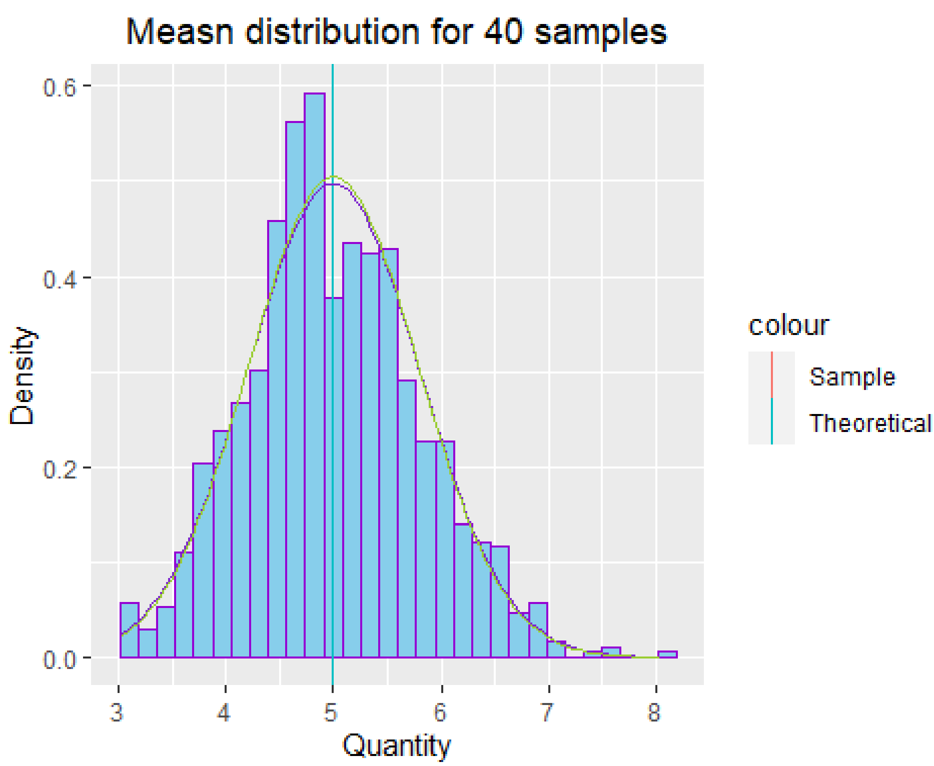
The sample mean is 4.9997019 and compare the theoretical mean is 5

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

```
sample.variance <- var(population)
theoretical.variance  <- (1 / lambda)^2 / (n)


sample.sd <- sd(population$x)
theoretical.sd  <- 1/(lambda * sqrt(n))
```

Show that the distribution is approximately normal.

```r
ggplot(population, aes(x=x)) +
  geom_histogram(aes(y=..density..), colour="darkviolet",
fill="skyblue")+
  guides(fill=FALSE) +
  labs(x="Quantity", y="Density", title=("Measn distribution for 40
samples"))+
  theme(plot.title = element_text(hjust = 0.5)) +
  geom_vline(aes(xintercept=sample.mean, colour="Sample"))+
  geom_vline(aes(xintercept=theoretical.mean, colour="Theoretical"))+
  stat_function(fun=dnorm, args=list(mean=sample.mean,sd=sample.sd),
color="purple3")+
  stat_function(fun=dnorm, args=list(mean=theoretical.mean,
sd=theoretical.sd), color="yellowgreen")

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



As is shown in the plot, the sample mean and the theoretical mean overlaps, from the shape of the plot it can be say that the distribution of means of 40 exponential distributions is close to the normal distribution.