Statistical Inference Course Project

Edilmo Palencia

Overview

In this report we analize the exponetial distribution looking at the behaviour of the mean, the variance and the distribution. In order to achive this, we are going to run multiple simulations and compare their behaviors with theorical one.

Simulations

In the next chunk of code we are going to: - Set all the parameters of the evaluation. – The rate or lambda value to use for all the simulations. – The size of the large sample. – The size of short samples. – The amount of short samples to generate. - Generate all the simulations. – One large sample of 1000 simulations. – One thousand short samples of 40 simulations. - Compute the mean and the standar deviation for all the samples

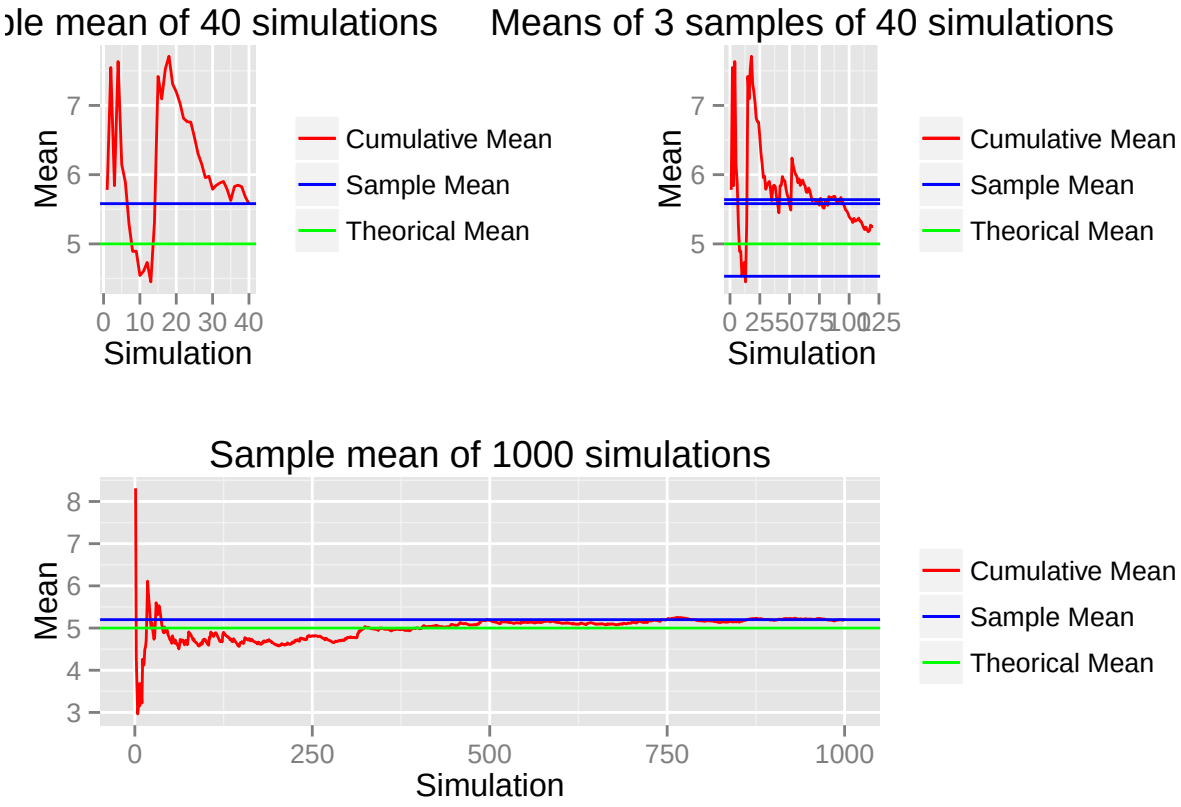
```
# Set value of lambda for all the experiments
lambda <- 0.2
# Set the size of a large sample
n.th <- 1000
# Set the size for short samples
n.sa <- 40
# Set the amount of short samples
n.me <- 1000
# Set the theorical mean for the exponential distribution
mean.theorical <- 1/lambda
# Set the theorical standar deviation for the exponential distribution
sd.theorical <- 1/lambda
# Generate the large sample
simulation.th <- rexp(n.th, lambda)
# Compute the mean of the large sample
simulation.th.mean <- mean(simulation.th)
# Compute the standar deviation of the large sample
simulation.th.sd <- sd(simulation.th)
# Generate a list of short samples
simulations.me <- lapply(rep(lambda,n.me), function(l){ rexp(n.sa,lambda)})
# Compute the mean of the short samples
simulations.me.means <- sapply(simulations.me, mean)
# Compute the mean of the short samples
simulations.me.sd <- sapply(simulations.me, sd)
```

Sample Mean versus Theoretical Mean

The next chunk of code is a example of how was generated the 3 figures showed below. Specifically, the code showed correspond to the first figure.

```
library(ggplot2)
# Compute the cumulative average of the simulation
y <- cumsum(simulations.me[[1]])/(1:n.sa)
```

```r
# Create the ggplot object with the data
g.svm.1 <- ggplot(data.frame(x = 1 : n.sa, y = y), aes(x = x, y = y))
# Add a red line for the cumulative average of the simulation
g.svm.1 <- g.svm.1 + geom_line(aes(colour = "Cumulative Mean"))
# Add a blue lines for the sample mean
g.svm.1 <- g.svm.1 + geom_hline(aes(yintercept = simulations.me.means[[1]], colour = "Sample Mean"), sh
# Add a green line for the theorical mean
g.svm.1 <- g.svm.1 + geom_hline(aes(yintercept = mean.theorical, colour = "Theorical Mean"), show_guide
# Add the labels of the axis and the title
g.svm.1 <- g.svm.1 + labs(x = "Simulation", y = "Mean") + ggtitle("Sample mean of 40 simulations")
# Add the legend
g.svm.1 <- g.svm.1 + scale_colour_manual("", breaks=c("Cumulative Mean", "Sample Mean", "Theorical Mean
```



These figures allows see how the mean of a sample becomes more and more equal to the theorical mean when we increase the amount of simulations present in the sample. The red lines ilustrate this, they represent the cumulative average of the sample (Law of Large Numbers). The blue and green line show the Sample and the Theorical means respectivaly.

- The first figure show the behaviour of a short sample of Means of 5.581282 and a difference with the theorical of 0.5812818.
- The second figure show the behaviour of tree short sample of Means of 5.581282, 5.639541, 4.532369, and a difference with the theorical of 0.5812818, 0.639541, -0.4676314 respectivaly.
- The third figure show the behaviour of a large sample of Means of 5.198975 and a difference with the theorical of 0.1989755.

Sample Variance versus Theoretical Variance

—Include figures (output from R) with titles. Highlight the variances you are comparing. Include text that explains your understanding of the differences of the variances.

Distribution

—Via figures and text, explain how one can tell the distribution is approximately normal.