

Final Project, Part 1

A. Kramer

Wednesday, February 18, 2015

Assignment

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can 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 $1/\lambda$. Set `lambda = 0.2` for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

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.

Overview

This project investigated the exponential distribution in R and compared it with the Central Limit Theorem. The exponential distribution was simulated in R with `rexp(n, lambda)` where `lambda` was the rate parameter. The mean of exponential distribution was set to be $1/\lambda$ and the standard deviation was also set to be $1/\lambda$. The `lambda` was set to 0.2 for all of the simulations. The distribution of averages of 40 exponentials was used for the investigation. 1000 simulations were executed to get sample data for the project.

Solution

Sections below provide answers to the questions poised for the Final Project 1, Part 1.

Simulation

Running the simulation

```
set.seed(2015)
lambda <- 0.2
n <- 1000
sample.size <- 40
data <- matrix(rexp(n * sample.size, rate=lambda), nrow = n)
means <- rowMeans(data)
```

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

The histogram below shows the distributions of the sample and theoretical means, and sample and theoretical variances. It appears that both of the distributions are really close. This affirms the Central Limit Theorem, vividly showing that both density curves will become looking very similar as the number of simulation increases.

```

# create histogram of averages
hist(means, breaks=50, prob=TRUE, xlab="Exponential Distributions of Means",
     ylab="Density", main="Distribution of Sample Means")

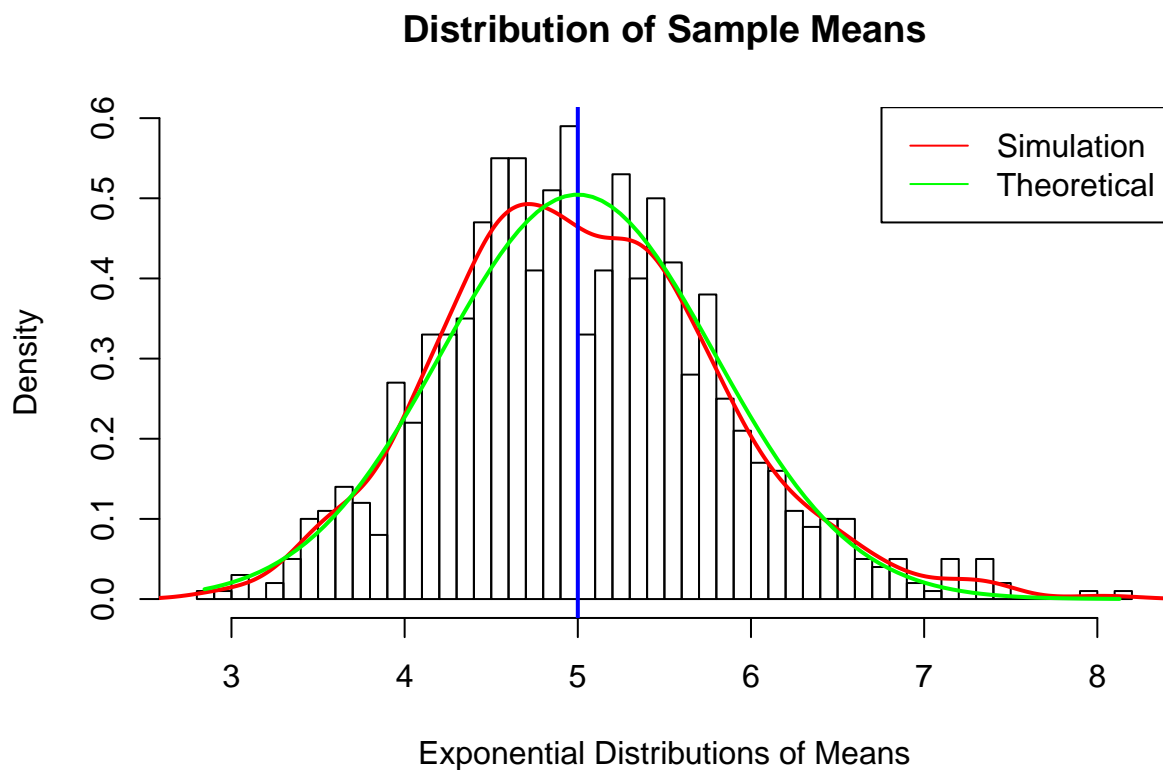
# density of the means for the sample
lines(density(means), col="red", lwd=2)

# show the mean of distribution
abline(v=1/lambda, col="blue", lwd=2)

# theoretical density of the means of samples
x <- seq(min(means), max(means), length=100)
y <- dnorm(x, mean=1/lambda, sd=(1/lambda/sqrt(sample.size)))
lines(x, y, pch=22, col="green", lwd=2)

# Add legend
legend('topright', c("Simulation", "Theoretical"), lty=c(1,1), col=c("red", "green"))

```



Question 2: Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

The list below shows theoretical sample's means and variances along with the calculations used to derive the answers. It looks like theoretical and sample means are really close and theoretical and sample variances are really close. According to Central Limit Theorem, the means and variances will become even closer if the number of simulations will be increased.

- distribution of sample means is 5.0115634 calculated using `mean(means)`
- theoretical center of the distribution is 5 calculated using `1/lambda`
- variance of sample means is 0.6410788 calculated using `var(means)`
- theoretical variance of the distribution is 0.625 calculated using `1/(lambda^2 * sample.size)`

Question 3: Show that the distribution is approximately normal.

According to the Central Limit Theorem, the averages for samples follow normal distribution. The q-q plot below also suggests the normality of the distribution of means for the sample. The same of normality is also affirmed by the histogram in Question 1. See Question 1.

```
qqnorm(means)
qqline(means)
```

