Exponential Distribution

hoaihuongbk

June 26, 2019

Overview

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.

In point 3, focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

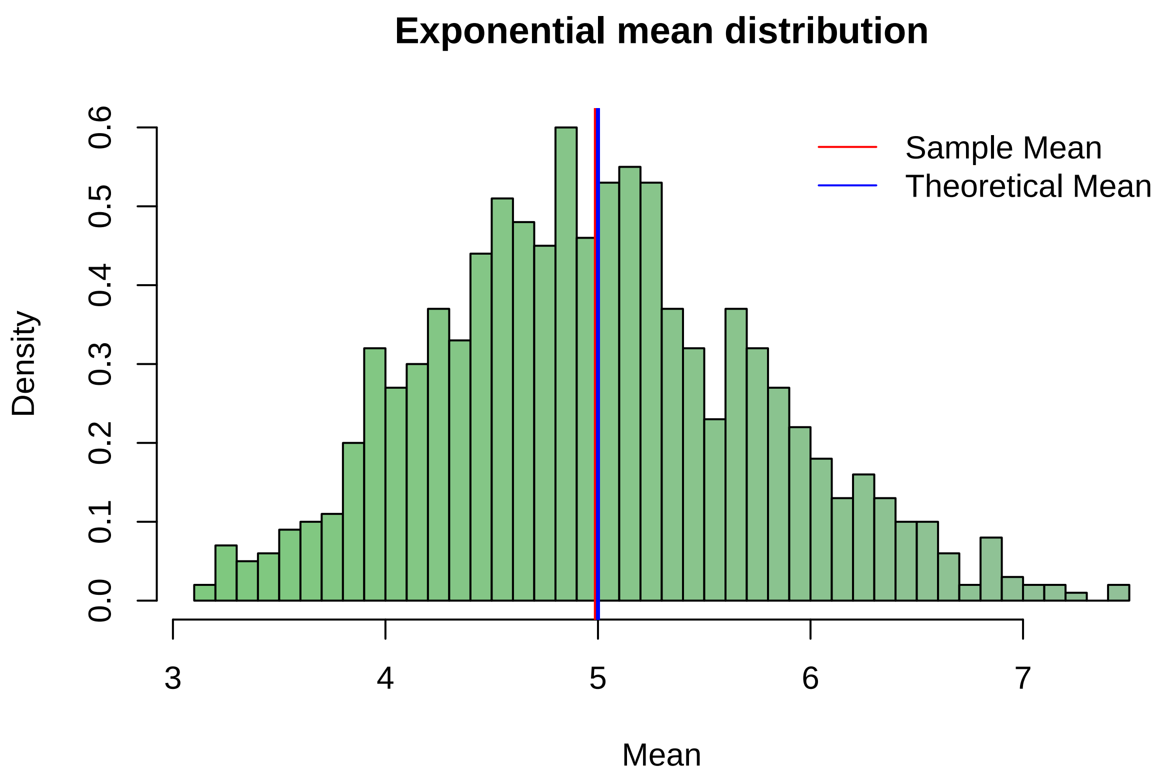
Preparation data

```
set.seed(1)
lambda <- 0.2
sd <- 1 / lambda
tm <- 1 / lambda ## Theoretical mean
nexp <- 40 ## Number of distributions
nsim <- 1000 ## Number of simulations
mns = NULL ## Set mean vector to NULL
for (i in 1 : nsim) mns <- c(mns, mean(rexp(nexp,lambda)))
sm <- mean(mns) ## Sample mean
sv <- var(mns) ## variance of the means
tv <- (sd^2)/nexp ## theoretical variance
```

Sample Mean vs Theoretical Mean

Sample mean is: 4.9900252. Theoretical mean is: 5

```
library(RColorBrewer)
cols <- colorRampPalette(brewer.pal(8,"Accent"))(length(mns))
hist(mns, breaks=40, prob = TRUE, main = "Exponential mean distribution", xlab="Mean", col=cols)
abline(v=mean(mns), col="red", lwd=2)
abline(v=tm, col="blue", lwd=2)
legend('topright', c("Sample Mean", "Theoretical Mean"),
       bty = "n", lty = c(1,1), col = c("red", "blue"))
```

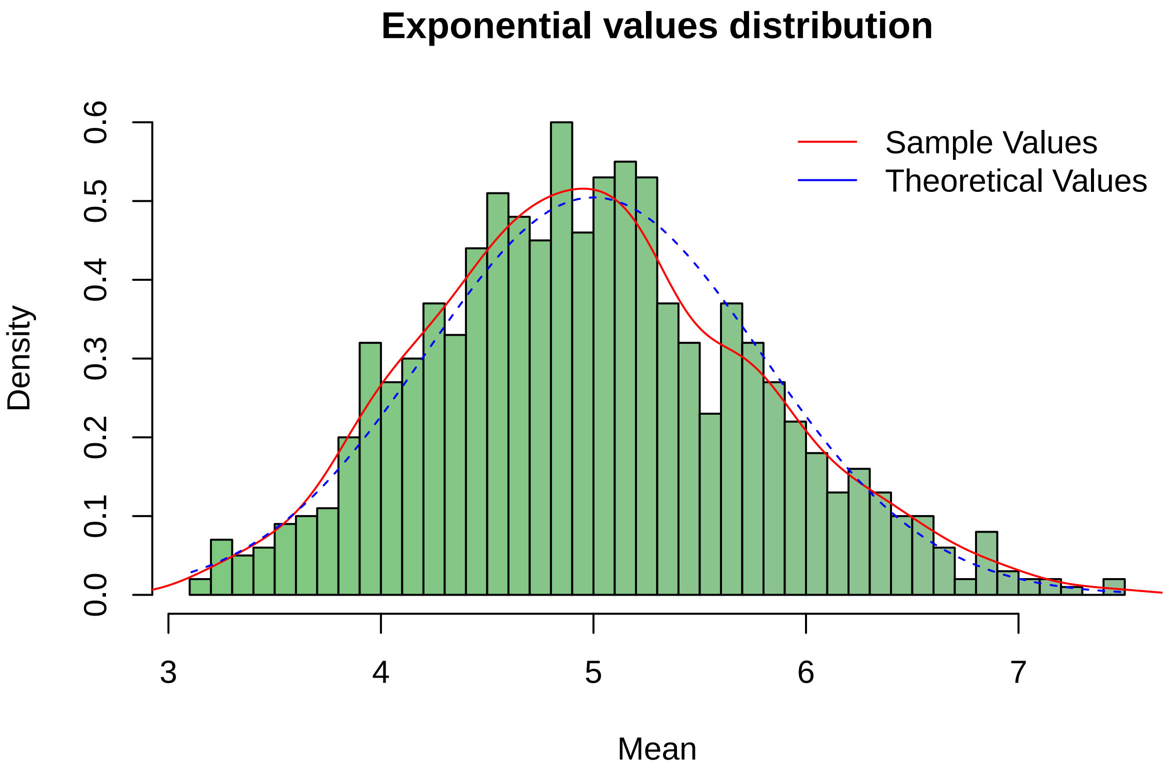


The sample mean (red line) and theoretical mean (blue line) is close.

Sample Variance vs Theoretical Variance

Variance of the mean is: 0.6111165. Theoretical variance is: 0.625

```
hist(mns, breaks = 40, prob = TRUE, main="Exponential values distribution", xlab="Mean", col=cols)
lines(density(mns), col="red")
xfit <- seq(min(mns), max(mns), length = 100)
yfit <- dnorm(xfit, mean = tm, sd = sqrt(tv))
lines(xfit, yfit, pch = 22, col = "blue", lty = 2)
legend('topright', c("Sample Values", "Theoretical Values"),
       bty = "n", lty = c(1,1), col = c("red", "blue"))
```



You can see the sample variance distribution follow theoretical variance distribution.

Is approximately normal distribution ?

Due to the Central Limit Theorem: "*The Central Limit Theorem (CLT) is a statistical theory states that given a sufficiently large sample size from a population with a finite level of variance, the mean of all samples from the same population will be approximately equal to the mean of the population*"

The figures above shows the density computed using the histogram and the normal density plotted with theoretical mean and variance values follow normal distribution.

Appendix

Source code: https://github.com/hoaihuongbk/Statistical_Inference