Central Limit Theorem vs Exponential Distribution study

Daniel Rodrigues Ambrosio

June 20th, 2015

1. Overview

This is the part 1 of the Project for the Statistical Inference course in Data Science Specialization track from Coursera.

The goal of this assignment is to investigate the exponential distribution in R and compare it with the Central Limit Theorem and illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. The study shall:

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.

1.1 Basis for the study

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. We sill set lambda = 0.2 for all of the simulations and investigate the distribution of averages of 40 exponentials.

1.2 Environment

Being able to reproduce every step of a data analysis is a crucial aspect of the data science. That being said, all the libraries used as support for this analysis are listed below and so is the system information.

```
library(ggplot2)


sessionInfo()
```

```
## R version 3.1.2 (2014-10-31)
## Platform: x86_64-apple-darwin13.4.0 (64-bit)
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
```