Simulation and Basic Inferential Data Analysis

Kent Mok

September 9, 2015

Overview:

The purpose of this experiment is to investigate the exponential distribution in R, and to 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 and the mean of exponential distribution is 1/lambda. This experiment will investigate the distribution of averages of 40 exponentials using a thousand simulations.

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. Include English explanations of the simulations you ran, with the accompanying R code. Your explanations should make clear what the R code accomplishes.

Sample Mean versus Theoretical Mean: Include figures with titles. In the figures, highlight the means you are comparing. Include text that explains the figures and what is shown on them, and provides appropriate numbers.

```
library(ggplot2)

set.seed(1234)

lambda = 0.2
mns = NULL
for (i in 1:1000)
    mns = c(mns, mean(rexp(40, lambda)))

hist(mns, freq = FALSE, main = "Distribution of the Mean of 40 Exponentials")
lines(density(mns))
abline(v = 1/lambda, col = "red")
```

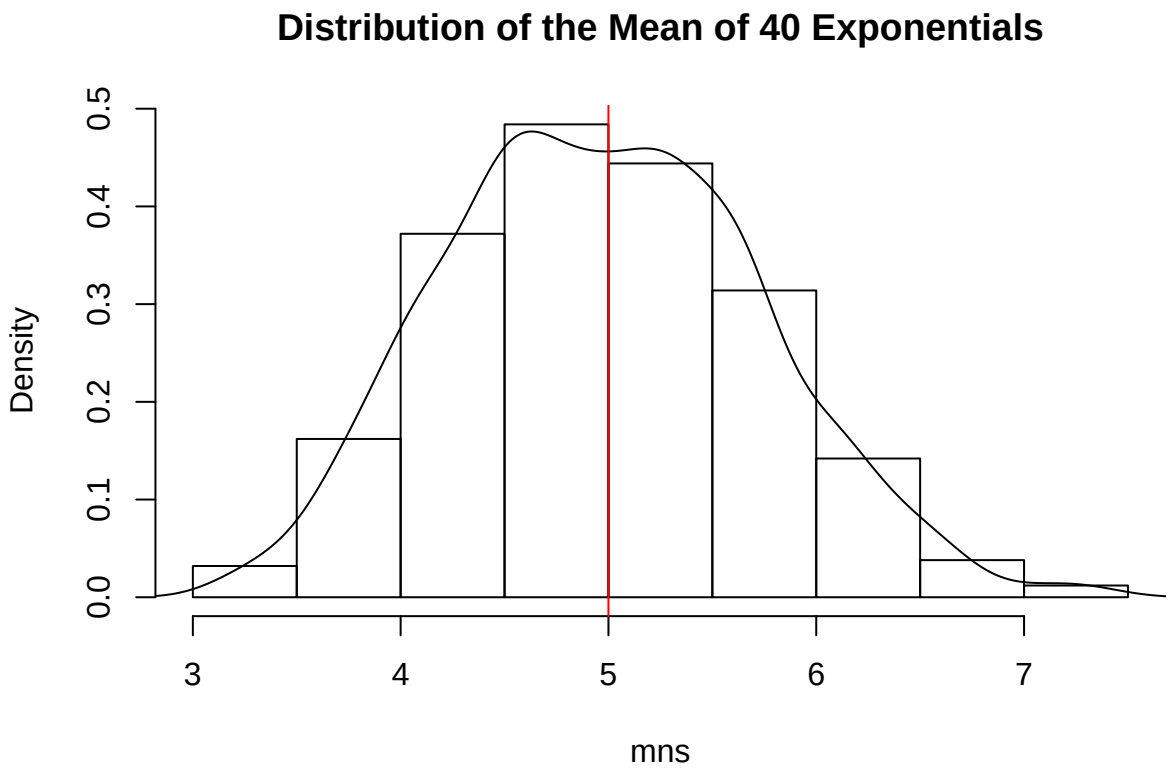


Figure 1 - Distribution of the Mean of 40 Exponentials

The above figure shows the distribution of one thousand simulations of the mean value for 40 exponentials. The theoretical mean of the exponential distribution is given as $1/lambda$.

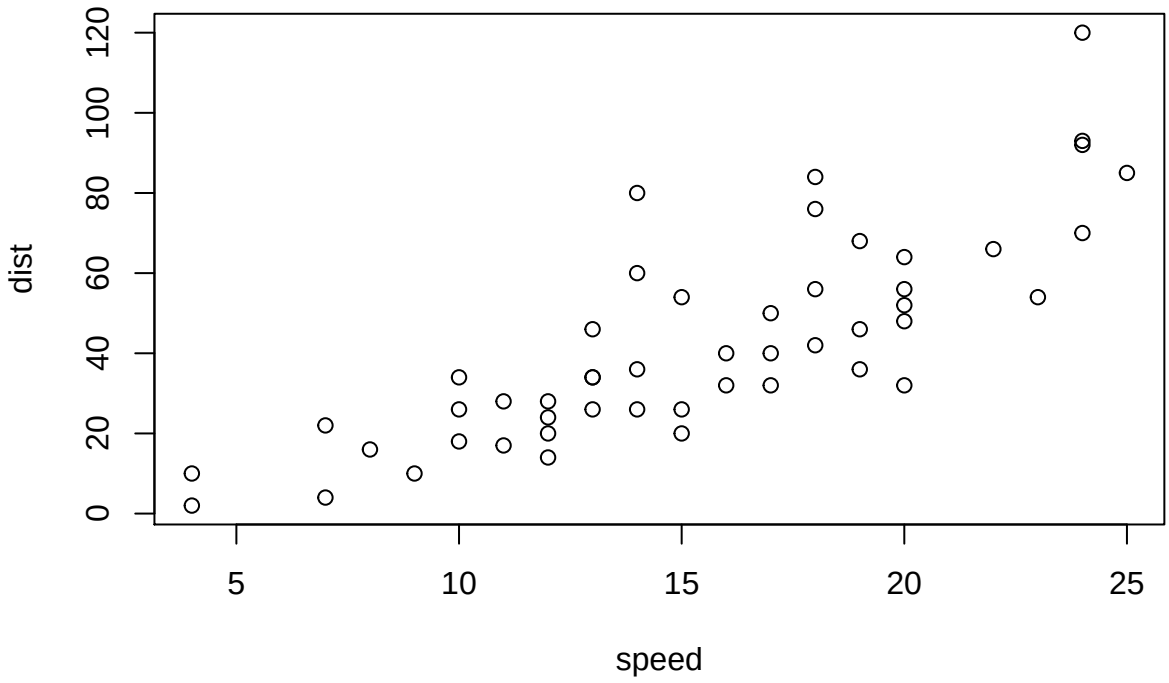
Sample Variance versus Theoretical Variance: Include figures (output from R) with titles. Highlight the variances you are comparing. Include text that explains your understanding of the differences of the variances.

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

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.