Comparing Exponential Distribution to the Central Limit Theorem

Dominic Lloyd

Tuesday, September 22, 2015

Overview

In this document we will investigate the exponential distribution in R and compare it with the Central Limit Theorem.
The exponential distribution will be simulated in R with rexp.
We will investigate the distribution of averages of 40 exponentials across a thousand simulations.

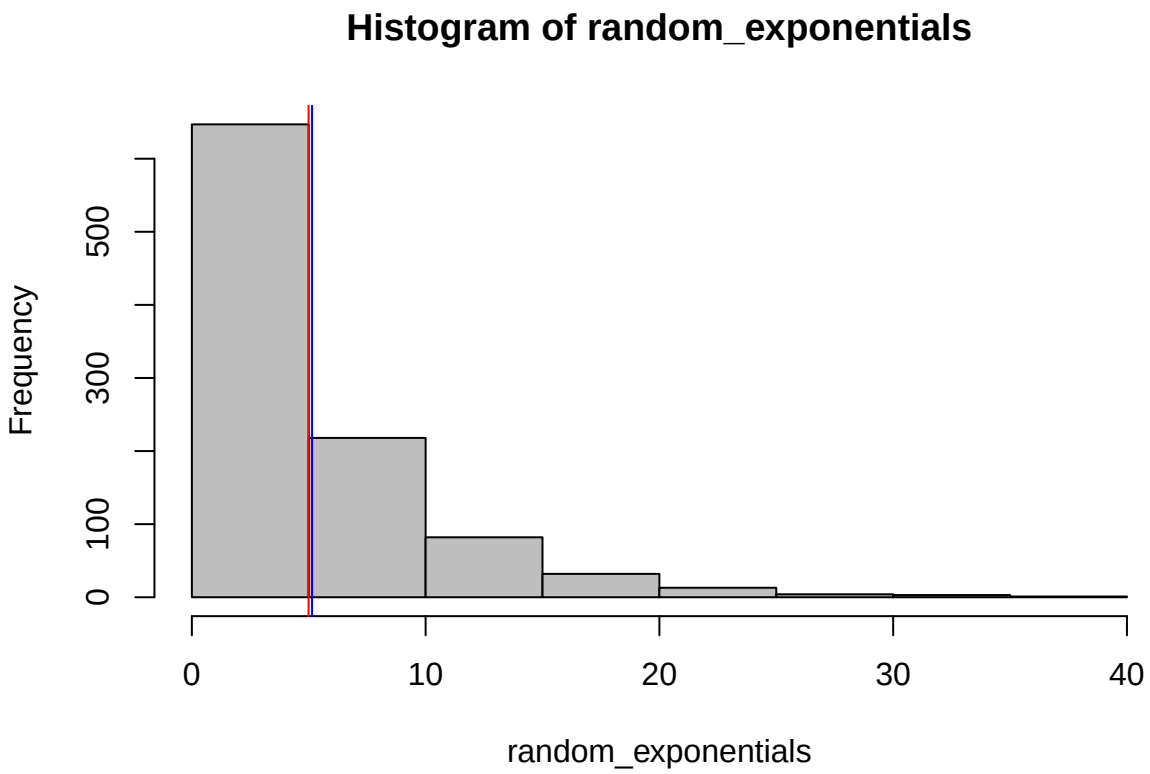
Simulations

The document will illustrate as follows:
1. The sample mean compared to the theoretical mean of the distribution.
2. The variance of the sample compared to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal. (With 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)

Sample Mean versus Theoretical Mean

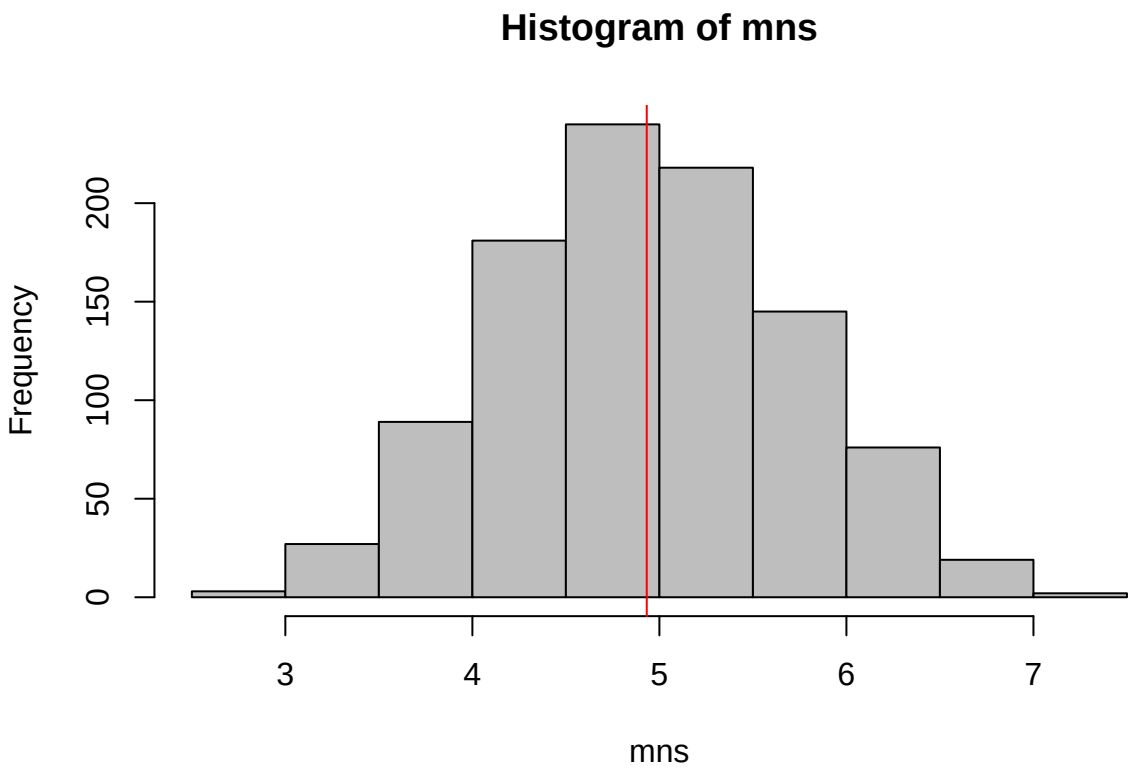
We will use lambda = 0.2 as the rate parameter for rexp. We will illustrate how the mean is 1/lambda (5) and the standard deviation is also 1/lambda (5).
First let us compare the distribution of 1000 exponentials.

```
# store the random exponentials, get the mean and the sd, plot the histogram and add the mean and sd li
random_exponentials <- rexp(1000, rate = 0.2)
mean_random_exp <- mean(random_exponentials)
sd_random_exp   <- sd(random_exponentials)
hist(random_exponentials, col = 'grey')
abline(v = mean_random_exp, col = "red", lwd = 1)
abline(v = sd_random_exp, col = "blue", lwd = 1)
```



and the distribution of 1000 averages of 40 random exponentials.

```
mns = NULL
for (i in 1 : 1000) mns = c(mns, mean(rexp(40, rate = 0.2)))
mean_mns <- mean(mns)
sd_mns   <- sd(mns)
hist(mns, col = 'grey')
abline(v = mean_mns, col = "red", lwd = 1)
abline(v = sd_mns, col = "blue", lwd = 1)
```



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

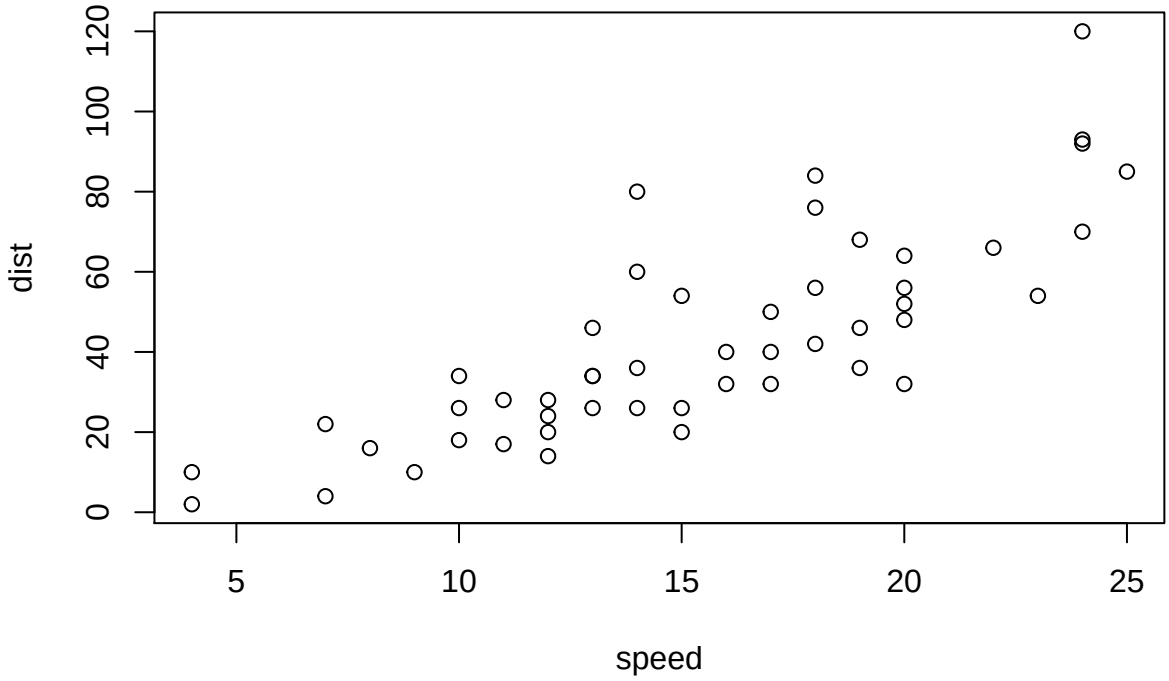
Via figures and text, explain how one can tell the distribution is approximately normal.

```
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:

```
plot(cars)
```


dist
0
20
40
60
80
100
120
5
10
15
20
25
speed