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

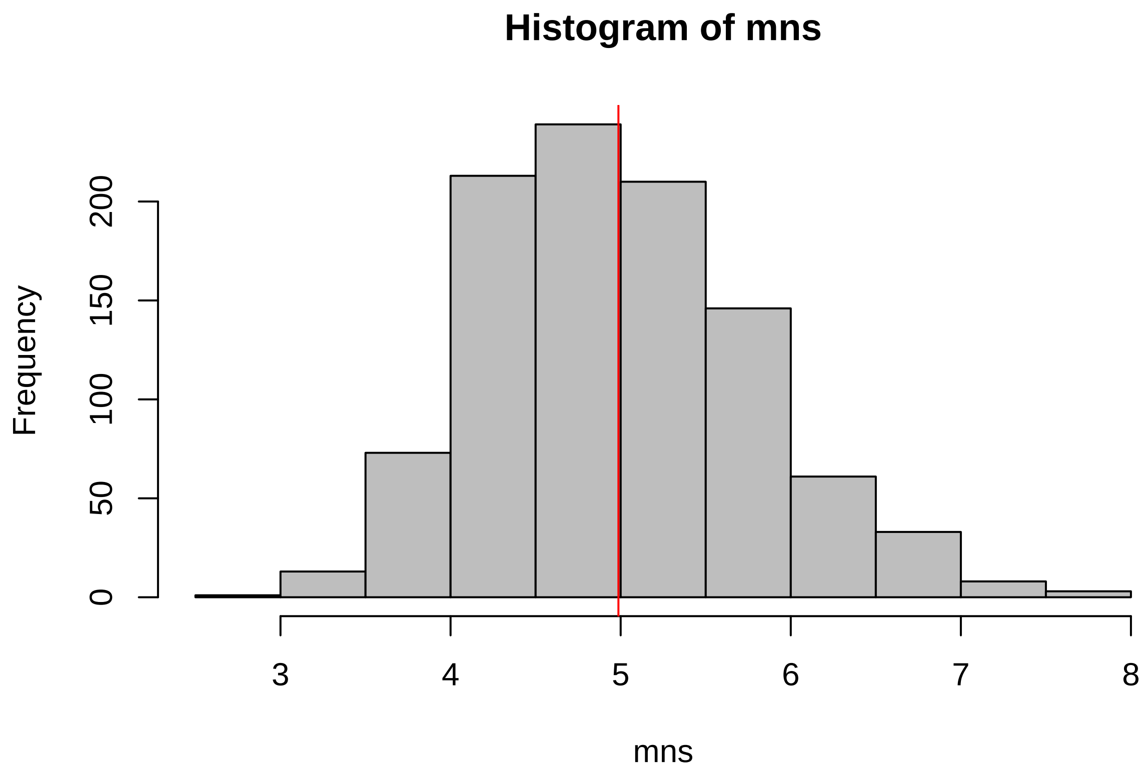
The document will illustrate as follows:
1. The sample mean compared to the theoretical mean of the distribution.
2. The variance of the sample compared to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal. (With focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials)

Sample Mean versus Theoretical Mean

In this test we are trying to show that we get similar results for mean and standard deviation for samples as we get for larger populations.
We know that when we use rexp to generate exponential numbers that the mean of these number will be 1/lambda (or 1/rate). Therefore if we use lambda = 0.2 as the rate parameter for rexp, we would expect to see a mean around 5 and a standard deviation around 5.

In this section we will generate a thousand samples of 40 random exponential numbers. In the histogram we will show the sample means

```
sample_means = NULL
sample_sd    = NULL
for (i in 1 : 1000) sample_means = c(sample_means, mean(rexp(40, rate = 0.2)))
for (i in 1 : 1000) sample_sd    = c(sample_sd, sd(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:

```
library("ggplot2")
```

```
## Warning: package 'ggplot2' was built under R version 3.2.2
```

```r
nosim <- 1000
cfunc <- function(x, n) sqrt(n) * (mean(x) - 3.5) / 1.71
dat <- data.frame(
  x = c(apply(matrix(sample(1 : 6, nosim * 10, replace = TRUE),
                     nosim), 1, cfunc, 10),
        apply(matrix(sample(1 : 6, nosim * 20, replace = TRUE),
                     nosim), 1, cfunc, 20),
        apply(matrix(sample(1 : 6, nosim * 30, replace = TRUE),
                     nosim), 1, cfunc, 30)
        ),
  size = factor(rep(c(10, 20, 30), rep(nosim, 3))))
g <- ggplot(dat, aes(x = x, fill = size)) + geom_histogram(alpha = .20, binwidth=.3, colour = "black", a
g <- g + stat_function(fun = dnorm, size = 2)
g + facet_grid(. ~ size)
```