Exponential Distribution

Axel Alejandro Garcia Fuentes

21 de junio de 2015

Overview

This document compares the standard normal distribution with the exponential distribution. That is performed by using simulation of data with those two probability distributions usin R. The comparison is supported with figures.

For more information about the exponential distribution, I found this YouTube video useful: Lecture 16: Exponential Distribution | Statistics 110

Simulations

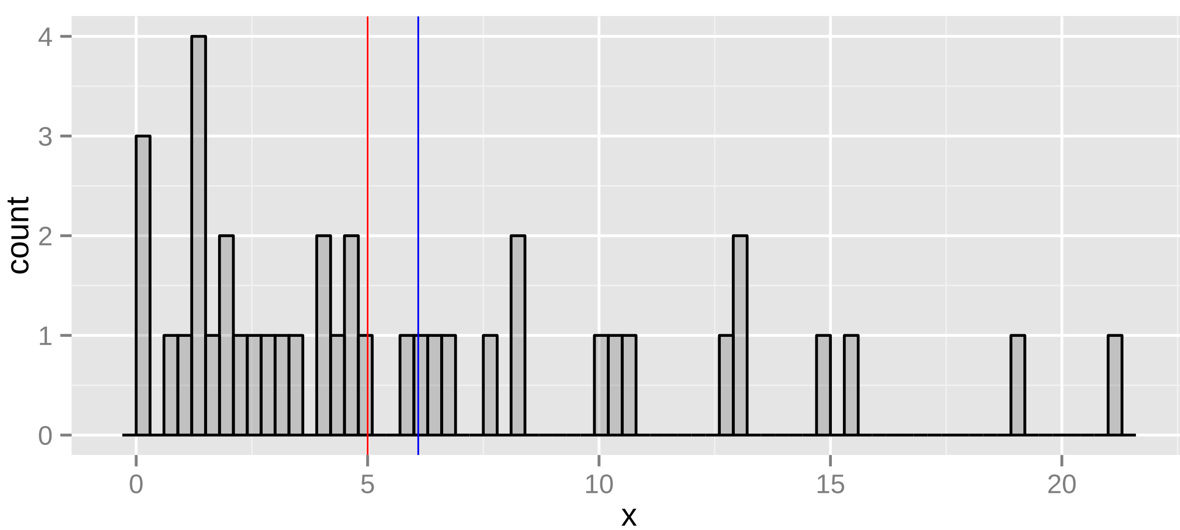
The exponential distribution can be simulated in R with the rexp(n, lambda) function, where lambda is the only parameter that distribution requires. It represents a rate in the data.

Sample Mean versus Theoretical Mean

The mean of exponential distribution is 1/lambda. So for data exponentially distributed with lambda equals to 0.2 the theoretical mean is 1/0.2 which is equal to 5.

The chart below shows how 40 random exponential samples distributes, along with the sample mean(in blue) in comparison with the theoretical mean(red).

```
data <- data.frame( x=rexp( 40, rate=0.2 ) )

ggplot(data, aes(x = x)) +
    geom_histogram(alpha = .20, binwidth=.3, colour = "black") +
    geom_vline(xintercept = 1/0.2, size = 0.3, color="red" ) +
    geom_vline(xintercept = mean(data$x), size = 0.3, color="blue")
```



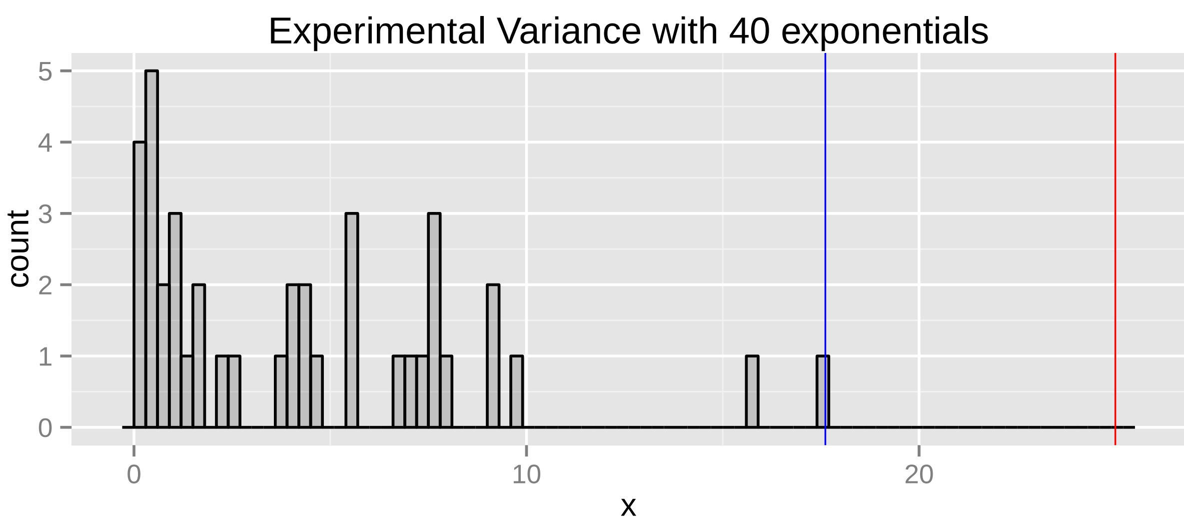
Notice that both lines are not exactly on the same spot. That could be explained by the fact that 40 samples may not be enough to have an exact match. In the other hand, this data simulation shows that theoretical mean and experimental mean aligns on approximately the same spot.

Sample Variance versus Theoretical Variance

In the same way than with the sample mean the sample standard deviation and the theoretical standard deviation of the exponential distribution should tend to be the same.

The standard deviation of exponential distribution is 1/lambda(same than the mean). Thus the variance is 1/lambdaˆ2. This means that for data exponentially distributed with lambda equals to 0.2 the theoretical variance is 25.

```
data <- data.frame( x=rexp( 40, rate=0.2 ) )

ggplot(data, aes(x = x) ) +
    geom_histogram(alpha = .20, binwidth=.3, colour = "black") +
    geom_vline(xintercept = (1/0.2)^2, size = 0.3, color="red" ) +
    geom_vline(xintercept = var(data$x), size = 0.3, color="blue") +
    ggtitle( "Experimental Variance with 40 exponentials" )
```



In that chart the experimental variance is shown in blue and the theoretical variance is shown in red.

The separation between both lines is explained by the fact that sample size may not be large enough to have an exact match. Using the law of large numbers it is possible to hypothesize that a large sample will yield a variance that is closer to the theoretical varince of the exponential distribution.

The table below shows the result of computing the mean and the variance for sample sizes of 10, 100, 1000 and 10000 exponentials respectivelly. Notice how the difference between the expected and actuals is slowly reduced as the sample size is increased:

```
samsize <- c(10,100,1000,10000)
samples <- sapply( samsize, function(x)rexp(x, rate=0.2) )

data <- data.frame(
    Size=c( sapply(samples, length) ),
```

```
    Exp.Mean=rep( 5,each=length(samples) ),
    Act.Mean=c(sapply(samples, mean) ),
    Exp.Var=rep( 5^2,each=length(samples) ),
    Act.Variance=c(sapply(samples, var) )
    )

data$Diff.Mean <- data$Exp.Mean-data$Act.Mean
data$Diff.Var <- data$Exp.Var-data$Act.Variance

data
```

```
##    Size Exp.Mean Act.Mean Exp.Var Act.Variance   Diff.Mean    Diff.Var
## 1    10        5 5.447395      25     38.49020 -0.44739534 -13.4901987
## 2   100        5 4.574878      25     20.30638  0.42512196   4.6936195
## 3  1000        5 4.906946      25     25.83302  0.09305420  -0.8330233
## 4 10000        5 4.920139      25     23.99760  0.07986147   1.0024016
```

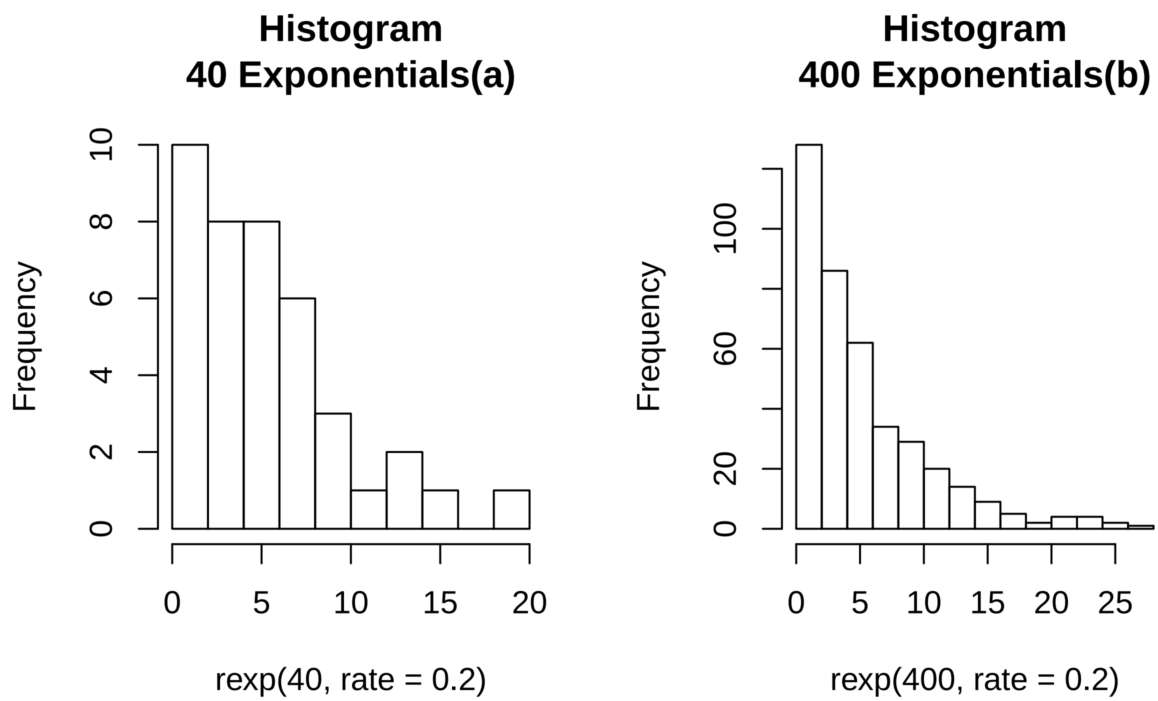
(In the table “Exp.” and “Act.” stand for “Expected” and “Actual”. They were typed like that so the table fit in the page)

This data simulation shows that theoretical mean and experimental mean aligns on approximately the same spot.

Distribution

The following charts show the shape that exponential distribution has. Note that its shape looks very similar despite the number of the used random exponentials:

```
par(mfrow=c(1,2))
hist( rexp( 40, rate=0.2 ), main="Histogram\n40 Exponentials(a)" )
hist( rexp( 400, rate=0.2 ), main="Histogram\n400 Exponentials(b)" )
```



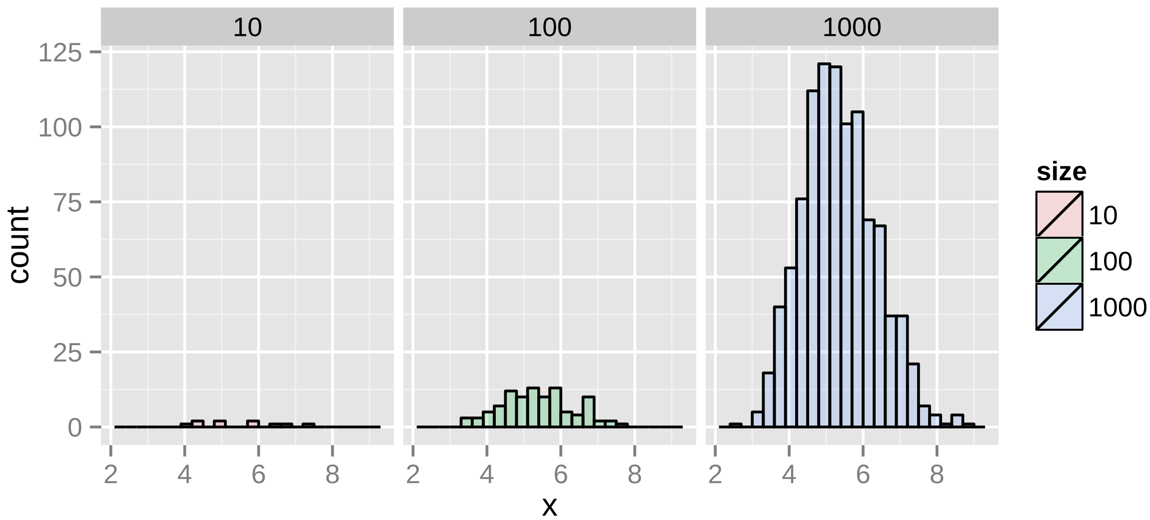
Nevertheless, the Central Limit Theorem can still be used for exponential distribution. The following function helps to simulate the means of n samples of 40 exponentials each; this is n sets of values like in (a) above:

```
simulateExpSampleMeans <- function( n=10 )
{
    means <- NULL
    #------------------------------
    # These are properties of the exponential distribution
    # Mean = SD = 1/lambda, this is using lambda=0.2
    populationMean <- 1/0.2
    populationSD <- 1/0.2
    #------------------------------

    for( i in 1:n )
    {
        sample <- rexp( 40, rate=0.2 )

        means <- c( means,
            # Central Limit Theorem formula applied to current data:
            (sqrt(40)*mean(sample)-populationMean)/populationSD )
    }
    means
}
```

That function will be used to generate means of 10, 100 and 1000 samples. Those samples are plot below.

```
nosam <- c( 10, 100, 1000 )

data <- data.frame(
```

```
x = unlist( c(sapply( nosam, simulateExpSampleMeans )) ),
size = factor(
    c(rep(10, each=10),
      rep(100, each=100),
      rep(1000, each=1000)) )
)
```



The plot clearly shows that as the set of exponentials grows, the distribution resambles the bell curve of the normal distribution. That experiment along with the CLT demonstrates that the exponential distribution can be approximate to standard normal; i.e. ~N(sample mean, standard error)