Comparison of the Exponential Distribution with the Central Limit Theorem

Bill Gourley

October 25, 2015

Overview

The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of the exponential distribution is $1/lambda$ and the standard deviation is also $1/lambda$.

In this analysis we will run a thousand simulations to investigate the distribution of the mean of the averages of 40 exponentials. This will illustrate, by the Central Limit Theorem, that the distribution obtained by running many simulations is approximately normal.

We will also illustrate that the distribution of 1000 random exponentials, by comparison, is not normal. lambda will be set to 0.2 for all simulations.

Analysis Requirements

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of averages of 40 exponentials by :

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.

R Setup and Initialization

```
#load required libraries
suppressMessages(library(ggplot2))
```

```
## Warning: package 'ggplot2' was built under R version 3.2.1
```

```
lambda <- 0.2; sample.size <- 40; no.simulations <- 1000
set.seed(123456)
```

Simulation : Collection of 1000 Exponentials of sample size 40

This simulation creates a collection matrix of 1000 exponentials each of size 40.

```
simulation.matrix <- matrix(rexp(no.simulations * sample.size, rate=lambda),
                            no.simulations, sample.size)
```

1. Comparison of Theoretical and Simulation Means

The theoretical mean of an exponential distribution is given by the formula $1/lambda$. Therfore the theoretical mean with lambda $= 0.2$ is equal to 5.

The simulation mean is calculated as shown below.

```
print(paste("Simulation Mean = ",round(mean(simulation.matrix),2),sep=""))
```

```
## [1] "Simulation Mean = 5.02"
```

2. Comparison of Theoretical and Simulation Variances

The theoretical variance of an exponential distribution is given by the formula $(1/lambda)^2$. Therefore the theoretical variance with lambda $= 0.2$, is equal to 25.

The simulation variance is calculated as shown below.

```
print(paste("Simulation Variance = ",round((mean(simulation.matrix)^2),2),sep=""))
```

```
## [1] "Simulation Variance = 25.23"
```

The values for the mean and variance obtained from the simulation are pretty close to the theoretical values thus demonstrating the Central Limit Theorem in action.

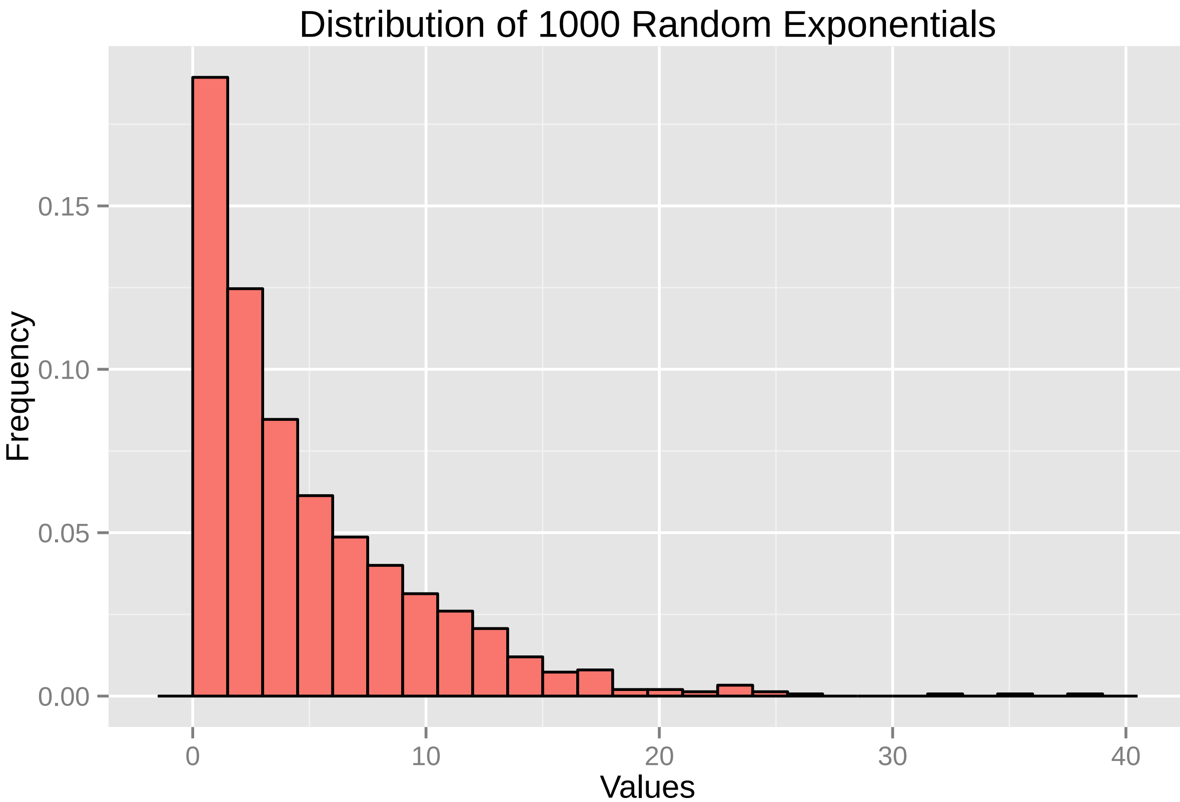
3. Comparison of the Exponential and Simulation Distributions

In this section we will create a collection of 1000 random exponentials and and compare its distribution with the collection of Exponential Sample Averages obtained from the simulation carried out previously.

Random 1000 Exponential Distribution

The collection of 1000 random exponentials can be created as shown below. Distribution is dsiplayed in the plot below.

```
random.exponentials <- data.frame(x=rexp(no.simulations,lambda))
```

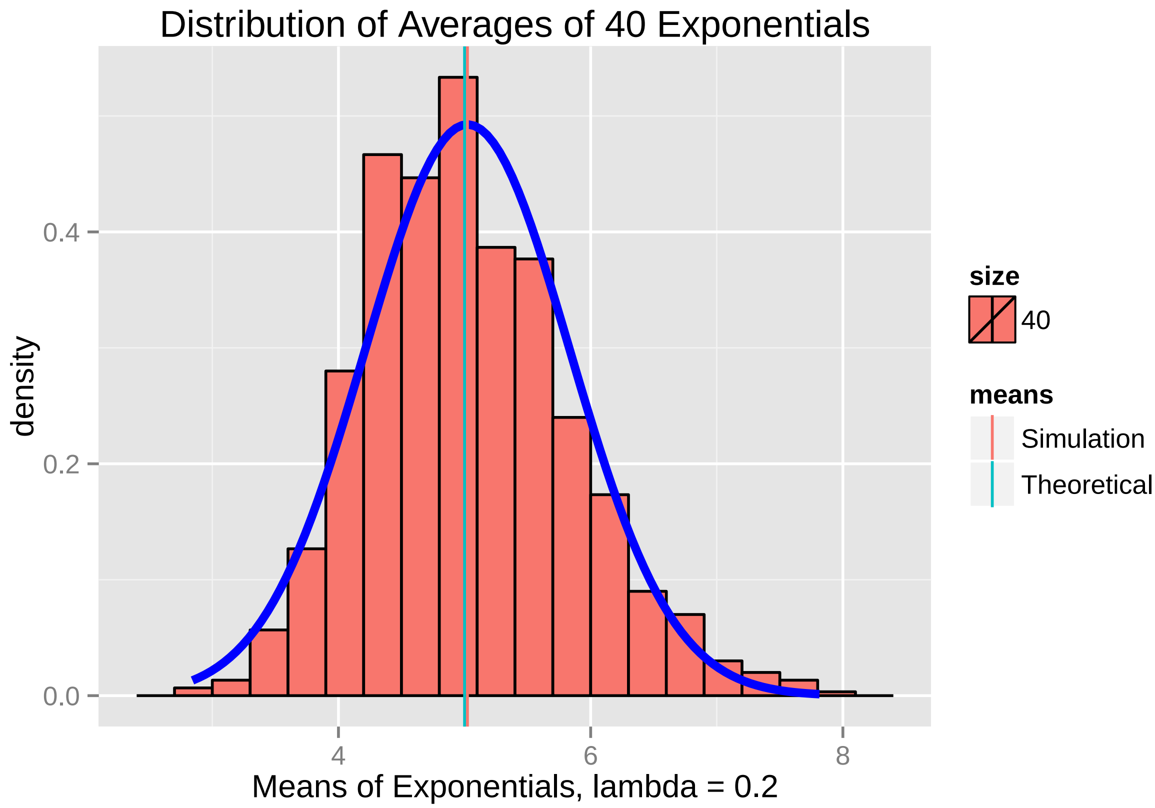


The distribution displays the characteristic exponential curve shape, therefore this distribution is clearly not normal.

Averages of 40 Exponentials Distribution

The Distribution of Averages of 40 Exponentials obtained from our simulation can be created as shown below. The Distribution is displayed in the plot below.

```
dat1 <- data.frame(
  x = rowMeans(simulation.matrix),size = factor(rep(c(sample.size),no.simulations)))
```



This distribution displays characteristics of a normal distribution, as indicated by the overlaid blue bell-shaped density curve. There is strong symmetry about the mean and the Theoretical mean and Simulation mean shown can also be seen to be almost identical in value.

Therefore we could characterize this distribution as 'approximately normal'.

Appendix : R code to produce the Distribution Plots

The R code used to prduce the Distribution Plots is as shown below.

Exponential Distribution

```
random.exponentials <- data.frame(x=rexp(no.simulations,lambda))

g1 <- ggplot(random.exponentials,aes(x = x,fill="red")) +
  geom_histogram(binwidth=1.5, colour = "black",
                 aes(y = ..density..)) +
  labs(title = "Distribution of 1000 Random Exponentials",
       x = "Values", y = "Frequency") +
  theme(legend.position="none")
```

Exponential Averages Distribution

```
dat1 <- data.frame(
  x = rowMeans(simulation.matrix),size = factor(rep(c(sample.size),no.simulations)))

#place theoretical and simulation means in a single data frame so they can be added to the plot legend
means.df <- data.frame(means=c("Theoretical","Simulation"),
                       vals=c(1/lambda,simulation.mean))

g2 <- ggplot(dat1, aes(x = x, fill = size)) + geom_histogram(binwidth=.3, colour = "black",
                aes(y = ..density..)) +
  stat_function(fun = dnorm,
                     size = 1.5,
                     col="blue",
                     show_guide=FALSE,
                     arg = list(mean = simulation.mean,sd = sd(dat1$x))) +
  geom_vline(data=means.df,
                    aes(xintercept=vals,
                        colour=means),
                    show_guide = TRUE) +
  labs(title = "Distribution of Averages of 40 Exponentials",
       x = "Means of Exponentials, lambda = 0.2")
```