Exponential Distribution Compare to Central Limit Theorem

Lau Heng Kar

December 25, 2015 (Statistical Inference Course Project)

Overview

In this project you will investigate the exponential distribution in R and 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. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand 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.

Simulations

```
# load library
require(ggplot2)

# set lambda is 0.2
lambda = 0.2

# distribution of 40 exponentials
n = 40

# repeat 1000 simulations
nsims = 1:1000

# set a seed to reproduce the data
set.seed(20151226)

# gather the means
meansDF <- data.frame(x = sapply(nsims, function(x) {mean(rexp(n, lambda))}))
```
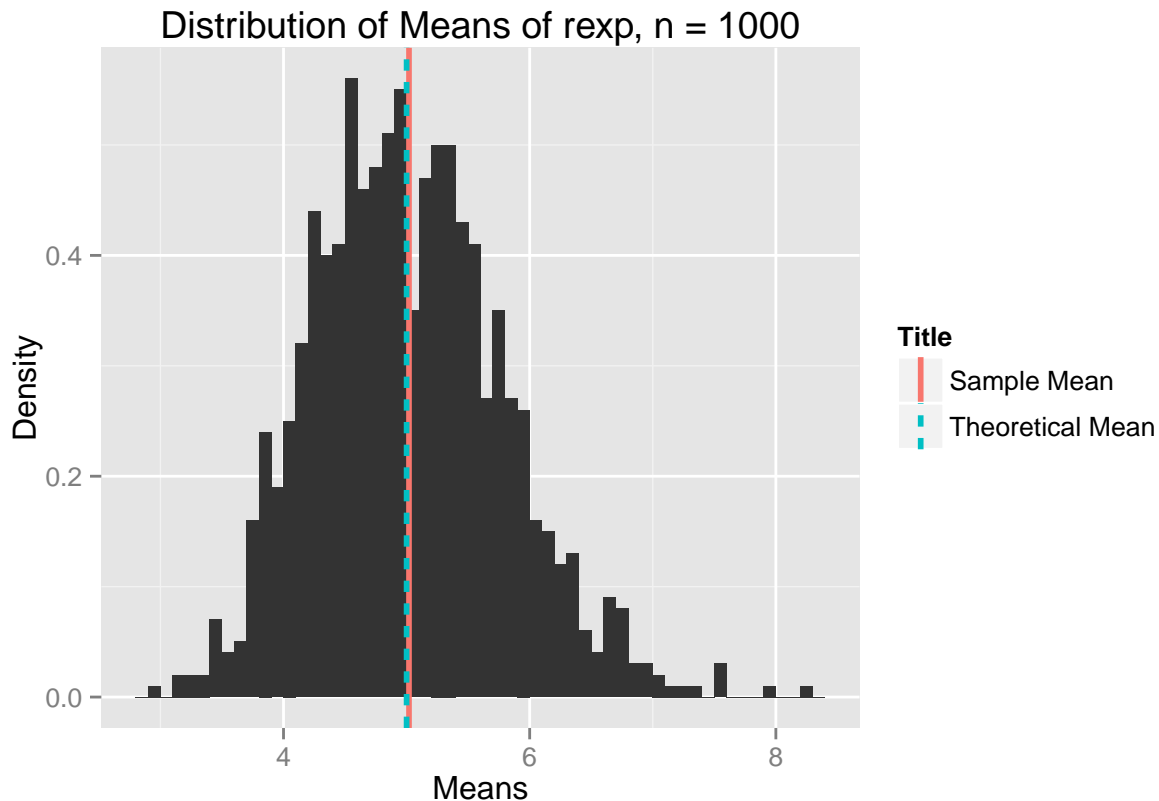
Sample Mean versus Theoretical Mean

```
# calculate Sample mean and Theoretical mean
meanData <- data.frame(Title=c("Sample Mean", "Theoretical Mean"), Vals=c(colMeans(meansDF), 1/lambda))

# plot vertical line for both Sample mean and Theoretical mean
ggplot(data = meansDF, aes(x = x)) +
    geom_histogram(binwidth=0.1, aes(y=..density..)) +
    geom_vline(data=meanData, mapping=aes(xintercept=Vals, linetype=Title, colour=Title), size=1, show_g
    labs(title = "Distribution of Means of rexp, n = 1000") + labs(x="Means") + labs(y="Density")
```


Conclusion: The sample mean of this distrubution is **5.0193258** and the theoretical mean is **5**. The actual center of the distribution of the average of 40 exponetials (red dotted line) is very close to its theoretical center of the distribution (green solid line).

Sample Variance versus Theoretical Variance

```
# Theoretical Standard Deviation and Variance:
theoreticalSD <- (1/lambda)/sqrt(n)
theoreticalVariance <- theoreticalSD ^2

# Simulated Standard Deviation and Variance:
sampleSD <- sd(meansDF$x)
sampleVariance <- sampleSD ^2
```

Conclusion: The variance of this sample distribution is **0.5977213** and the theoretical variance is **0.625**. Both of them are close. The actual standard deviation of the sample distribution is **0.7731244**. The theoretical standard deviation is **0.7905694**. The difference of actual sd and theoretical sd is small.

Distribution

```
# calculate Sample mean and Theoretical mean
df <- data.frame(Title=c("Sample", "Normal Distribution"), Vals=c(colMeans(meansDF), 1/lambda))

# plot Distribution of Averages of Sample vs Theoretical Mean
ggplot(data = meansDF, aes(x = x)) +
    geom_histogram(binwidth=0.1, aes(y=..density..)) +
    geom_density(color="red", size=1, linetype=2) +
    geom_vline(data=df[df$Title == "Sample", ], mapping=aes(xintercept=df[df$Title == "Sample", ]$Vals,
    stat_function(fun=dnorm, arg=list(mean=df$Vals[df$Title == "Normal Distribution"], sd=theoreticalSD)
    geom_vline(data=df[df$Title == "Normal Distribution", ], mapping=aes(xintercept=df[df$Title == "Norm
    guides(linetype=guide_legend(override.aes=list(colour = c("green","red"))))+
    labs(title = "Distribution of Averages of Samples vs Theoretical Mean") + labs(x="Means") + labs(y=
```


Distribution of Averages of Samples vs Theoretical Mean
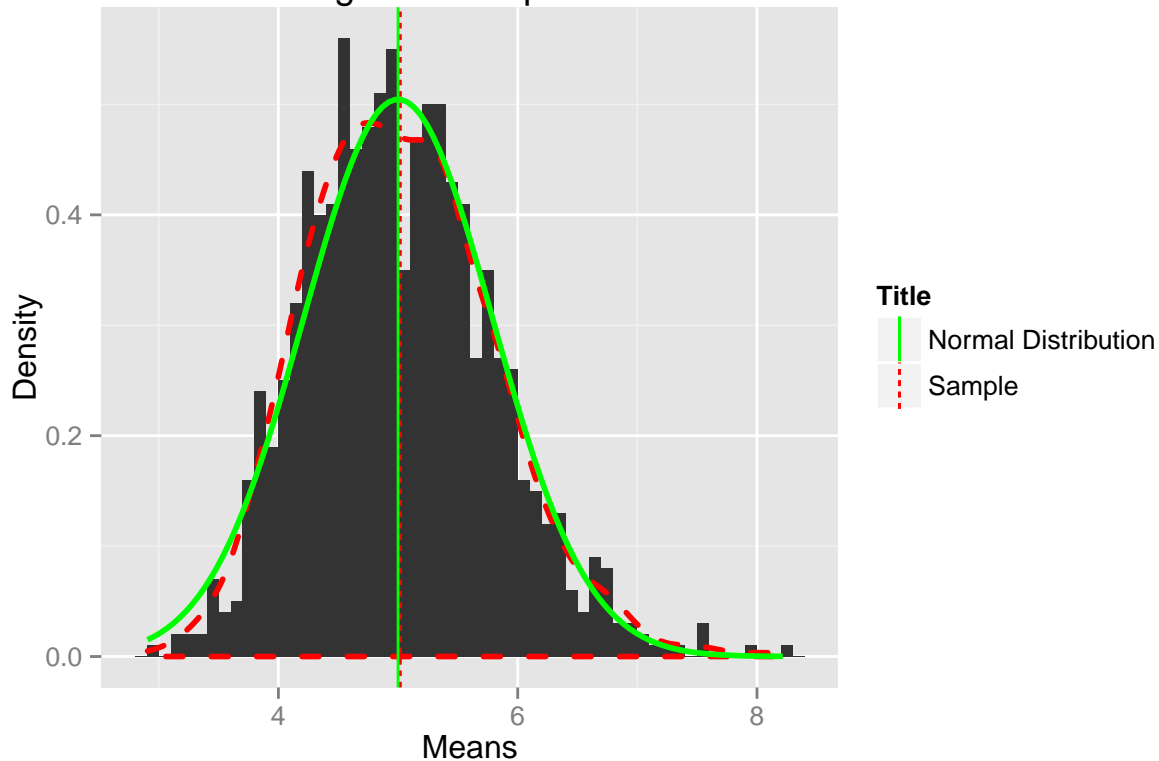
Conclusion: The green line is the normal distribution with lambda = 0.2. The red line is the distribution of averages of the simulated samples. The figure shows that the 2 distribution lines, green and red, are well aligned thus the distribution of simulated data is approximately normal.