Central Limit Theorem for Exponential Distribution

Y. Zheng

Friday, October 14th, 2016

Overview

We wish to investigate the validity of Central Limit Theorem: *The arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined (finite) expected value and finite variance, will be approximately normally distributed*

1000 simulations of **40** *exponential distributed random variables* were conducted and the eventual distribution was compared against the *normal distribution*.

Density plots of simulation results were compared against theoretical predicted probability density functions.

The following properties of the exponential distribution is assumed:

- The pdf, $\mathrm{f(x)} = \lambda \mathrm{e}^{(-\lambda \mathrm{x})}$
- Mean $= \frac{1}{\lambda}$
- Standard Deviation(sd) $= \frac{1}{\lambda}$

For this project specifically: $\lambda = 0.2$

```
lambda <- 0.2
```

Simulations

1000 simulations were functionally iterated and the means of 40 random samples of exponentially distributed random variables were caquired. The final result is stored in a data.frame called means with a single column labelled x.

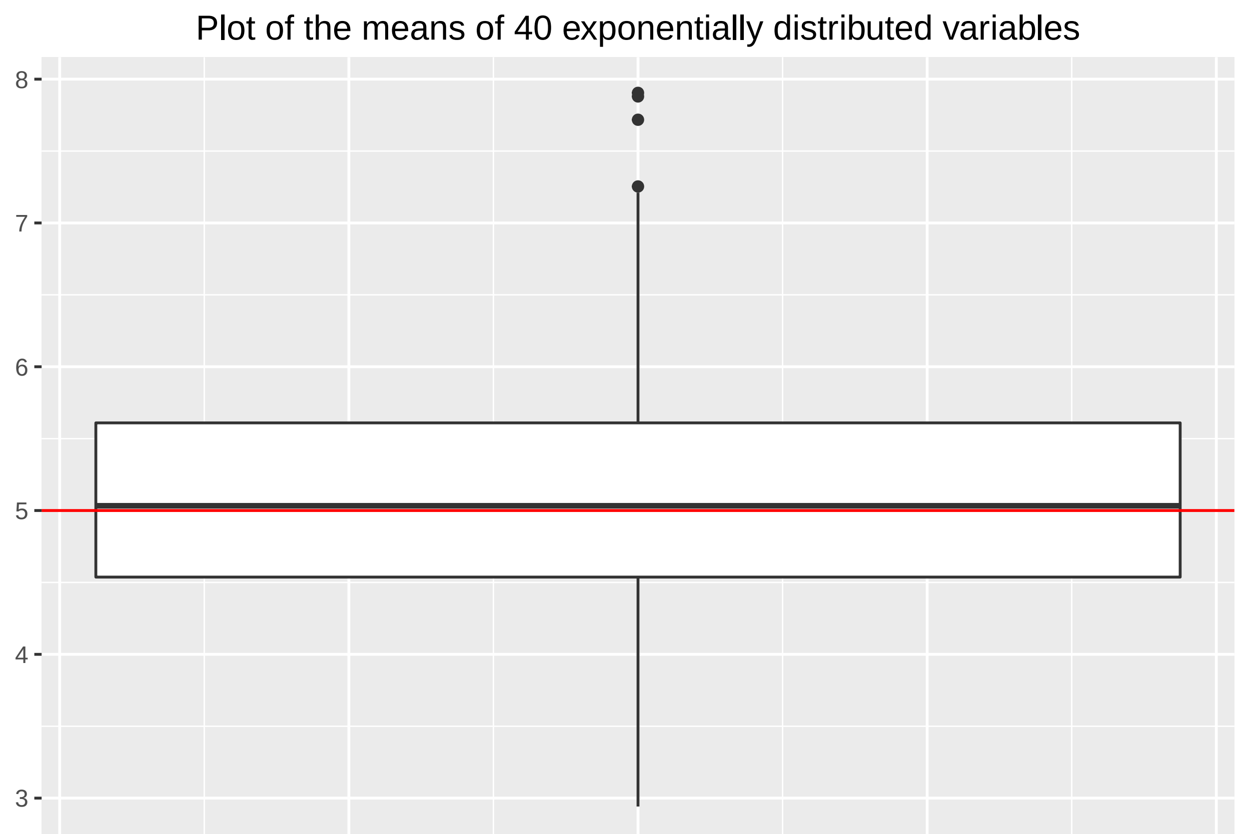
In the spirit of reproducibility, the random seed provided corresponds to the day this investigation was conducted - 1410 (14th of October).

```
set.seed(1410)

means <- data.frame(
    x = sapply(1:1000, function(x){
      mean(rexp(40, 0.2))
    })
)

head(means)
```

```
##          x
## 1 6.736015
## 2 6.172501
## 3 5.306666
## 4 5.359624
## 5 5.009063
## 6 4.564575
```

Sample Mean

```r
# Useful plotting library loaded
library(ggplot2)

# Boxplot of the means of 40 exponentially distributed variables in 1000 simulations
qplot(x= 1, y= means$x, geom = "boxplot") + geom_hline(aes(yintercept = 5), color = "red") +
    theme(axis.title.x=element_blank(), axis.text.x=element_blank(),
          axis.ticks.x=element_blank(), axis.title.y=element_blank()) +
    labs(title = "Plot of the means of 40 exponentially distributed variables")
```



Sample Mean calculation:

```r
mean(means$x)
```

```
## [1] 5.056119
```

The **red line** corresponds to the theoretical mean: $\frac{1}{\lambda} = \mathbf{5}$ From the plot above, we can see the sample mean (Given by the thick black line) roughly corresponds to the theoretical mean. This suggests evidence for the **Law of Large Number (LLN)**.

Sample Variance

The sample variance is evaluated in the below code.

```
var(means$x)
```

```
## [1] 0.6557698
```

The theoretical variance for a *exponentially distributed random variable*, $\sigma^2 = \frac{1}{\lambda^2}$.
We therefore expect the variance of the mean of 40 such variables, ${\sigma_{40}}^2 = \frac{\sigma^2}{40} = \frac{1}{40\lambda^2}$

```
1/(40*lambda^2)
```

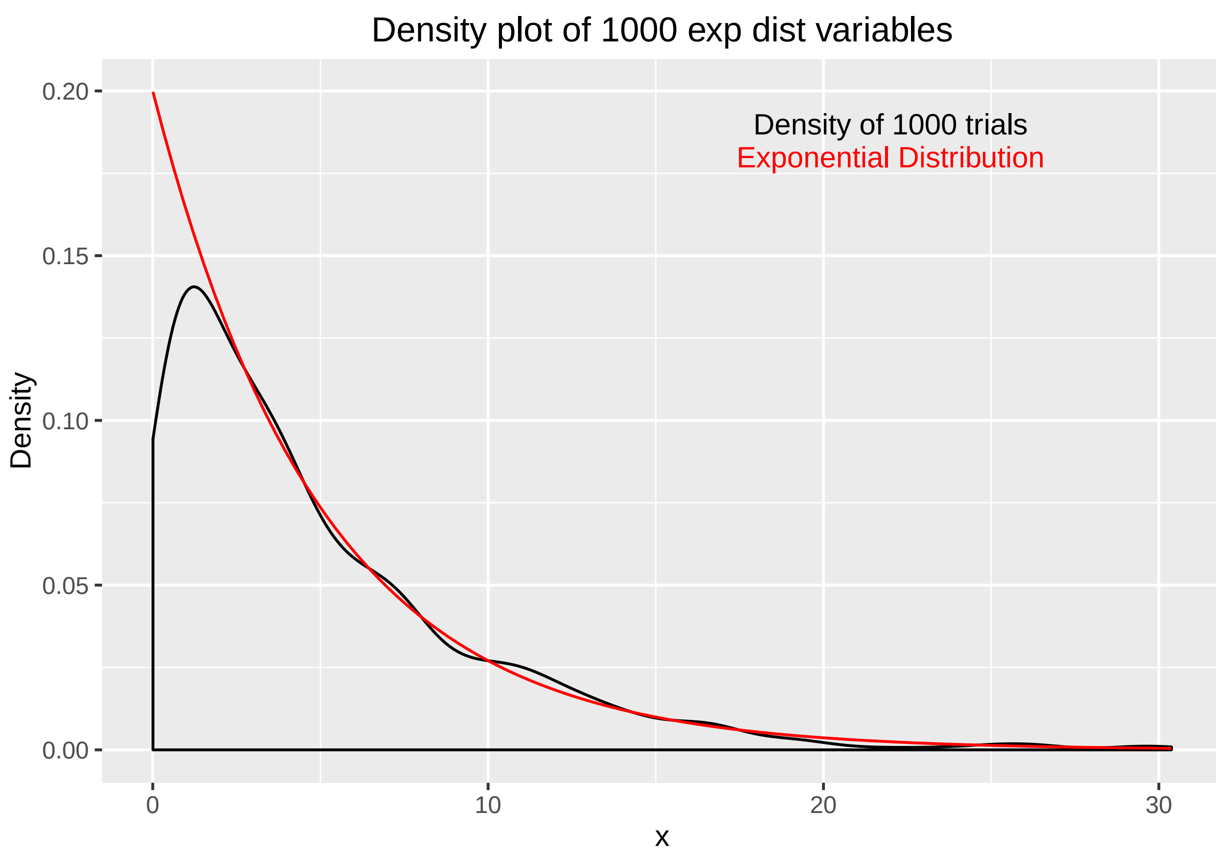
```
## [1] 0.625
```

We are therefore able to see the correspondence between theoretical and sample variance.
However, they do not correspond exactly. The disparity may be explained by the variability of the samples we are taking. Since we are taking a sample of 40 random variables, there exists a non-zero probability of selecting variables such that the mean of 40 samples does not correspond to the exact mean. Pedantically speaking, it is unlikely we are able to recover the exact variance and disparities are well within expectations.

Distribution

We will now show that the means of the 40 exponentially distributed random variables does indeed correspond to the normal distribution. The graphical method is preferred as it takes into account higher order variability such as *skew* and *ketosis*.

```
# Plot of density function of 1000 exponential distributed random variables
# compared to theoretical probability density function
ggplot(data.frame(x = rexp(1000, 0.2)), aes(x)) + geom_density() +
    labs(y = "Density", title = "Density plot of 1000 exp dist variables") +
    stat_function(fun = dexp, colour = "red", args = list(rate = 0.2)) +
    annotate("text", x = 22, y = 0.18, label = "Exponential Distribution", color = "red") +
    annotate("text", x = 22, y = 0.19, label = "Density of 1000 trials")
```

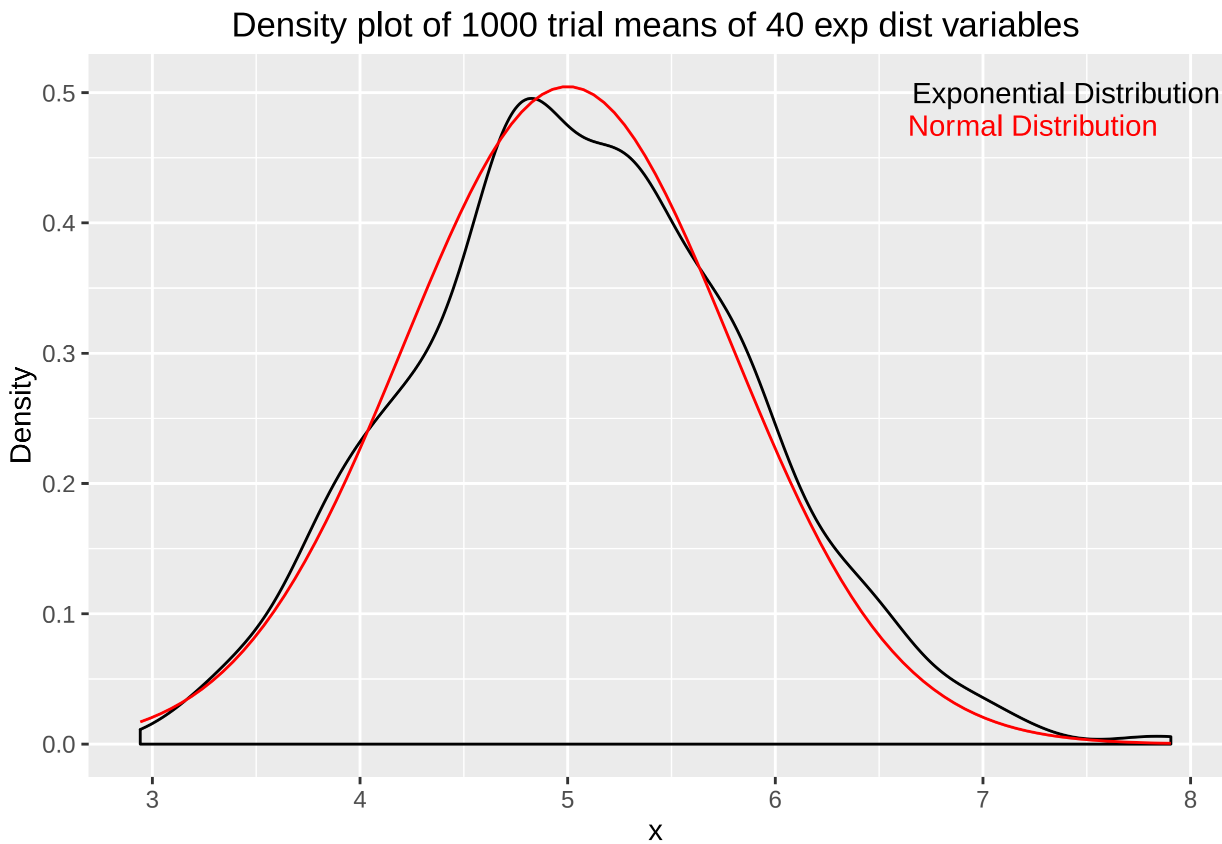


The plot above is used to verify 2 properties we are interested in:

1. The density of 1000 random variables does correspond to the exponential distribution.
2. It does not look anything like the normal distribution.

Property 2 is particularly useful as it depicts the power of **Central Limit Theorem (CLT)**. Despite how drastically different the exponential distribution is from the normal distribution, the means of 40 exponentially distributed random variables still resembles the normal distribution (as we will soon see in the plot below).

```r
# Plot of density function of 1000 trial of the means of 40 exponential distributed random variables
# compared to predicted normal distribution pdf according to Central Limit Theorem
ggplot(means, aes(x)) + geom_density() +
    labs(y = "Density", title = "Density plot of 1000 trial means of 40 exp dist variables") +
    stat_function(fun = dnorm, colour = "red", args = list(mean = 5, sd = 5/sqrt(40))) +
    annotate("text", x = 7.24, y = 0.475, label = "Normal Distribution", color = "red") +
    annotate("text", x = 7.4, y = 0.5, label = "Exponential Distribution")
```



The means of a conservative sample size of 40 have already produced such striking resemblence to the normal distribution. We can therefore see by example that CLT works.