Comparison of Exponential Distribution against Central Limit Theorem

Sooraj Raju

May 8, 2016

1. Overview

The objective of this project is to investigate the Exponential Distribution and compare it against Central Limit Theorem.

Central Limit Theorem: In probability theory, the central limit theorem (CLT) states that, given certain conditions, the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed, regardless of the underlying distribution

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

We will set n = 40 and set lambda = 0.2 for all of the simulations and compare the resulting distribution of the smaple averages to the theory for 1000 simulations.

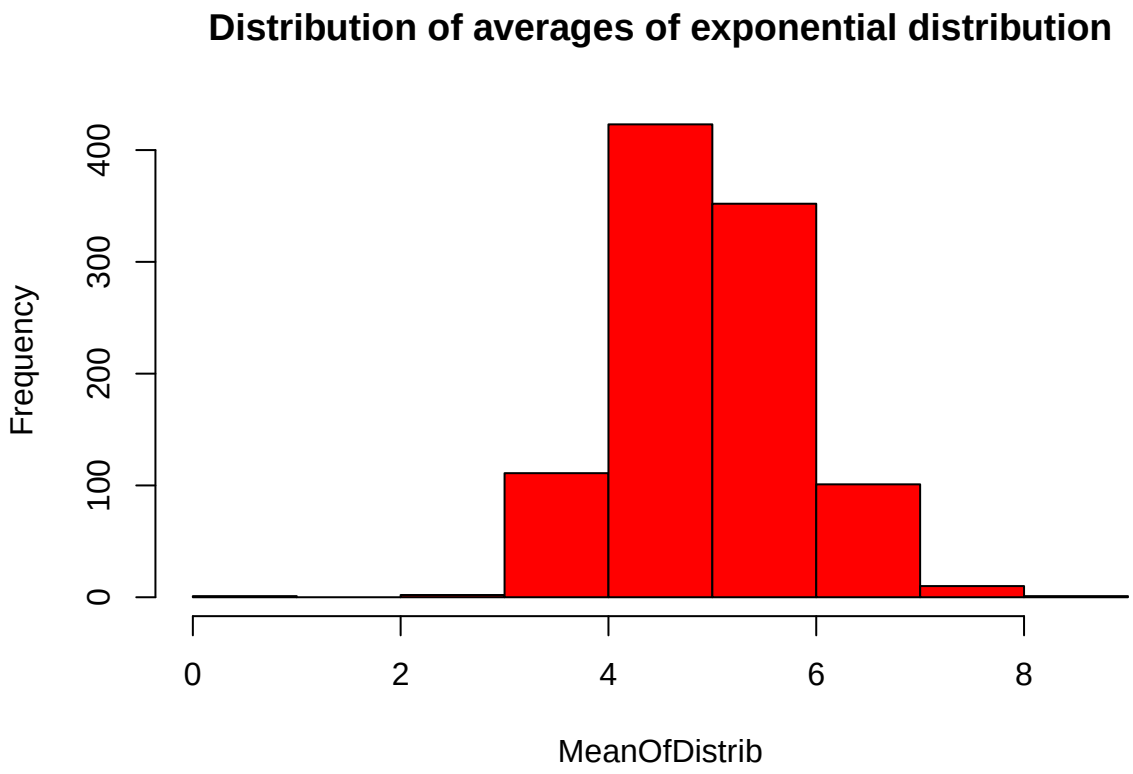
```
library(ggplot2)
library(psych)
```

```
set.seed(1000)
lambda <- 0.2 ## Set lambda as per instructions

## Distibution size
DisSize <- 40

## number of simulations
NoOfSim <- 1000

MeanOfDistrib <- 0
for (i in 1 : NoOfSim)
  MeanOfDistrib <- c(MeanOfDistrib, mean(rexp(DisSize,lambda)))
hist(MeanOfDistrib,col="red",main="Distribution of averages of exponential distribution")
```




```
varxp <- ((1/lambda)^2)/DisSize ## theoretical variance
varmean <- var(MeanOfDistrib) ## variance of the means
Mean_theoritical <- 1/lambda
```

```
## [1] "The theoretical mean is 5.000000 while mean of means is 4.981981"
```

```
## [1] "The theoretical variance is 0.625000 while variance of means 0.678534"
```

```
head(MeanOfDistrib)
```

```
## [1] 0.000000 4.514222 5.050788 3.252216 3.916899 4.898008
```

```
summary(MeanOfDistrib)
```

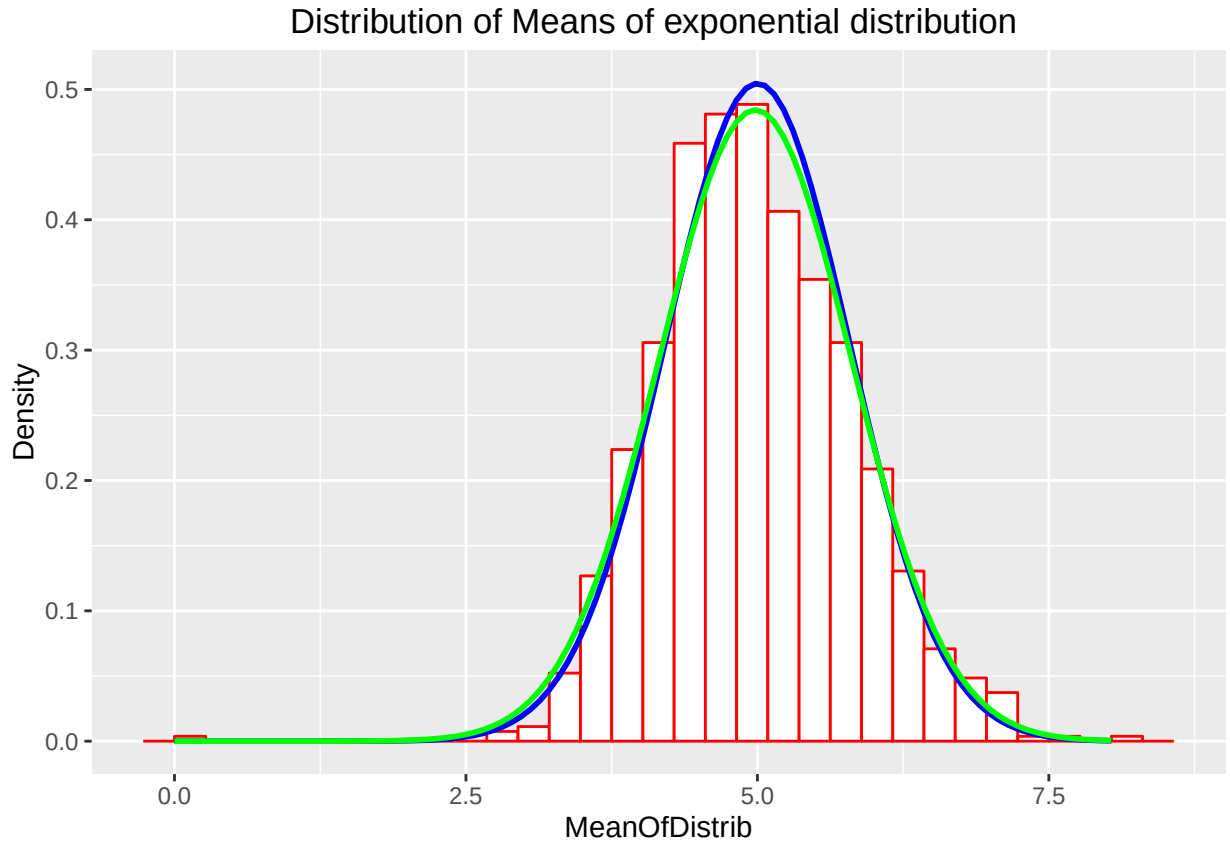
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   4.413   4.902   4.982   5.533   8.036
```

```
describe(MeanOfDistrib)
```

```
##   vars    n mean   sd median trimmed  mad min  max range skew kurtosis
## 1    1 1001 4.98 0.82    4.9    4.96 0.83   0 8.04  8.04 0.09     1.07
##     se
## 1 0.03
```

```
plotdata <- data.frame(MeanOfDistrib)
plot1 <- ggplot(plotdata,aes(x = MeanOfDistrib))
plot1 <- plot1 +geom_histogram(aes(y=..density..), colour="red",fill="white")
plot1<-plot1+labs(title="Distribution of Means of exponential distribution ", y="Density")
plot1<-plot1 +stat_function(fun=dnorm,args=list( mean=1/lambda, sd=sqrt(varxp)),color = "blue", size = 
plot1<-plot1 +stat_function(fun=dnorm,args=list( mean=mean(MeanOfDistrib), sd=sqrt(varmean)),color = "g
print(plot1)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

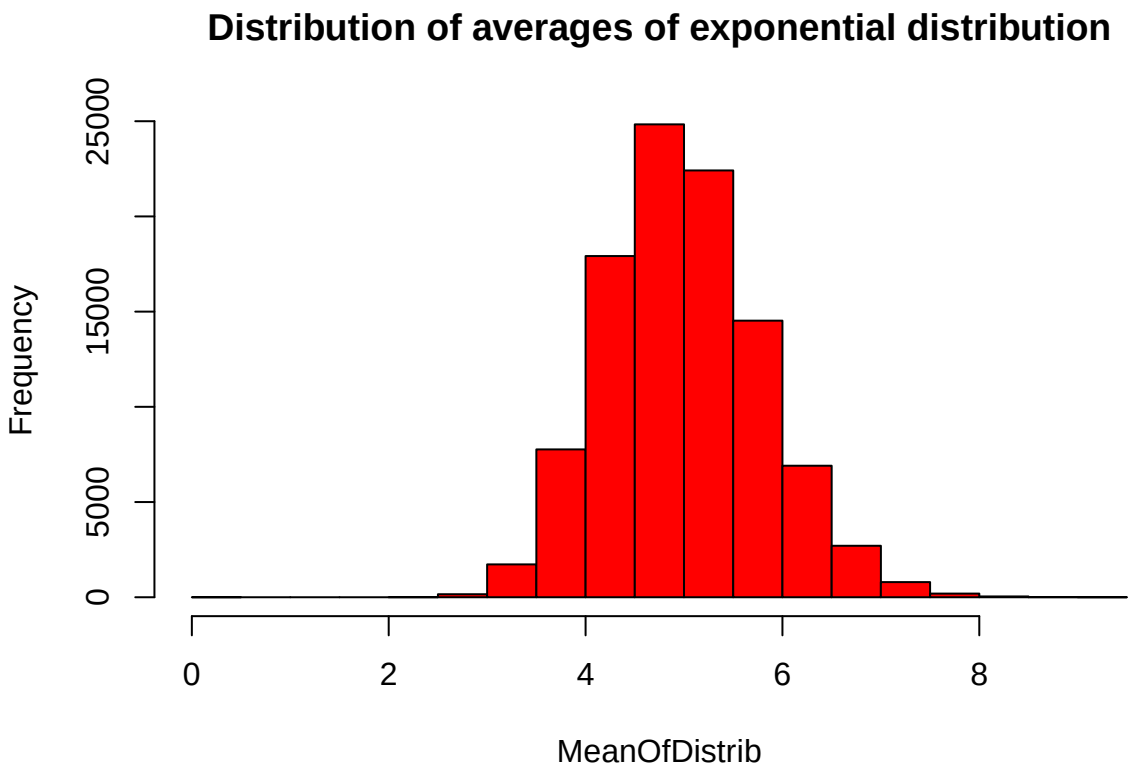



```
set.seed(1000)
lambda <- 0.2 ## Set lambda as per instructions

## Distibution size
DisSize <- 40

## number of simulations
NoOfSim <- 100000

MeanOfDistrib <- 0
for (i in 1 : NoOfSim)
  MeanOfDistrib <- c(MeanOfDistrib, mean(rexp(DisSize,lambda)))
hist(MeanOfDistrib,col="red",main="Distribution of averages of exponential distribution")
```




```
varxp <- ((1/lambda)^2)/DisSize ## theoretical variance
varmean <- var(MeanOfDistrib) ## variance of the means
Mean_theoritical <- 1/lambda
```

```
## [1] "The theoretical mean is 5.000000 while mean of means is 4.995464"
```

```
## [1] "The theoretical variance is 0.625000 while variance of means 0.624027"
```

```
head(MeanOfDistrib)
```

```
## [1] 0.000000 4.514222 5.050788 3.252216 3.916899 4.898008
```

```
summary(MeanOfDistrib)
```

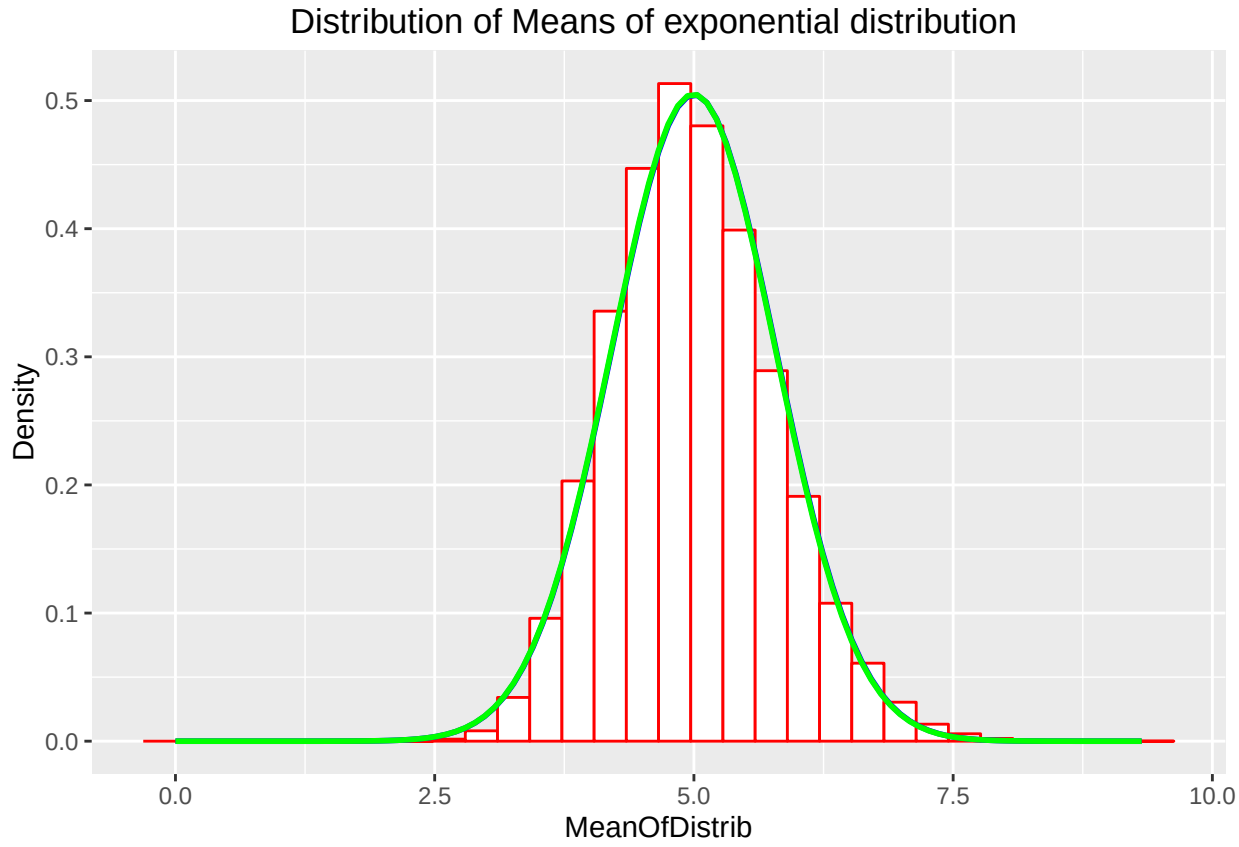
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   4.444   4.952   4.995   5.504   9.318
```

```
describe(MeanOfDistrib)
```

```
##   vars      n mean   sd median trimmed  mad min  max range skew kurtosis
## 1    1 100001    5 0.79   4.95    4.97 0.78   0 9.32  9.32 0.32     0.16
##   se
## 1  0
```

```r
plotdata <- data.frame(MeanOfDistrib)
plot1 <- ggplot(plotdata,aes(x = MeanOfDistrib))
plot1 <- plot1 +geom_histogram(aes(y=..density..), colour="red",fill="white")
plot1<-plot1+labs(title="Distribution of Means of exponential distribution ", y="Density")
plot1<-plot1 +stat_function(fun=dnorm,args=list( mean=1/lambda, sd=sqrt(varxp)),color = "blue", size = 
plot1<-plot1 +stat_function(fun=dnorm,args=list( mean=mean(MeanOfDistrib), sd=sqrt(varmean)),color = "g
print(plot1)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.