Simulation of the Central Limit Theorem

Bo Liu

March 27, 2017

Overview

This report is about simulation of the Central Limit Theorem.We will investigate the exponential distribution in R and compare it with the Central Limit Theorem.

SIMULATIONS

Set some parameters before we start the simulations. average_number stands for the number of exponentials. sample_number is the simulations times.

```
lambda = 0.2
average_number = 40
sample_number = 1000
```

Sample the distribution of averages of 40 exponentials a thousand times.

```
sample_mean_distribution<-sapply(1:sample_number,
                                 function(i)mean(rexp(average_number,lambda)))
head(sample_mean_distribution)
```

```
## [1] 4.434814 3.731379 4.038770 4.588016 5.505394 4.500214
```

Sample Mean versus Theoretical Mean

Get the sample mean

```
sample_mean <- mean(sample_mean_distribution)
sample_mean
```

```
## [1] 4.997298
```

Get the theoretical mean of the distribution which is 1/lambda.

```
theoretical_mean <- 1/lambda
theoretical_mean
```

```
## [1] 5
```

Comparison of sample mean and theoretical mean

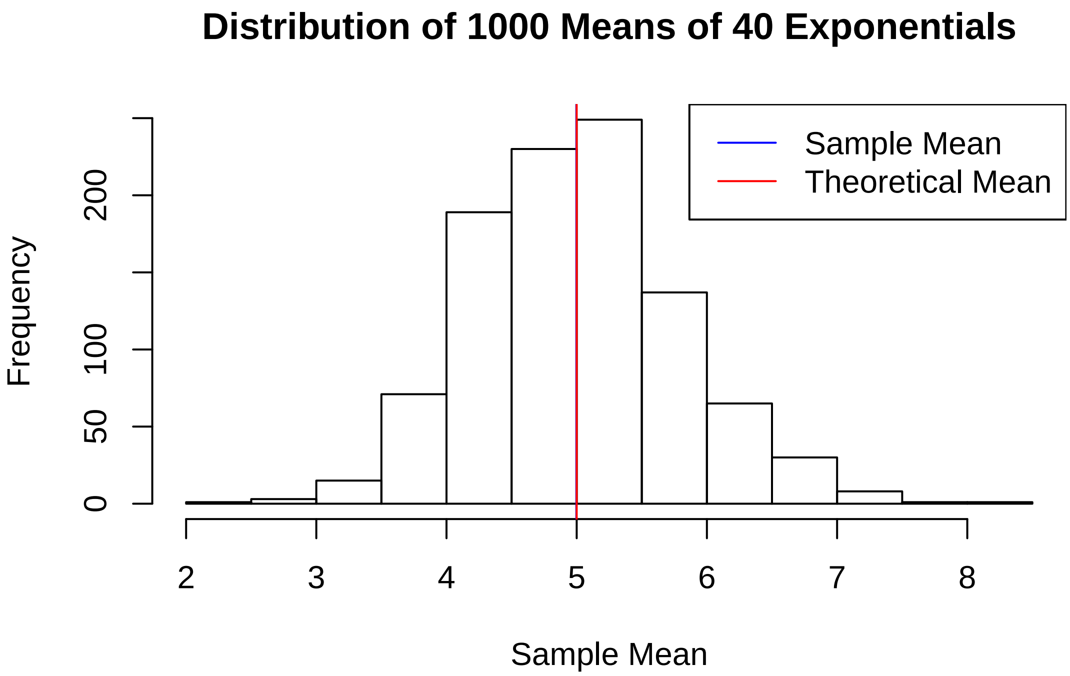
```
sample_mean-theoretical_mean
```

```
## [1] -0.002701766
```

Plot the distribution of average 40 exponentials with sample mean and theoretical mean, the blue line indicates the sample mean and red line stands for the theoretical mean.

```
hist(sample_mean_distribution,
     main = "Distribution of 1000 Means of 40 Exponentials",
     xlab = "Sample Mean")
abline(v= sample_mean, col = "blue", lwd = 1)
```

```
abline(v= theoretical_mean, col = "red", lwd = 1)
legend("topright",c("Sample Mean","Theoretical Mean"),col=c("blue","red"),lwd = c(1,1))
```



It is pretty close, so CLT worked in this experiment.

Sample Variance versus Theoretical Variance

First, get sample variance.

```
sample_variance<-var(sample_mean_distribution)
sample_variance
```

```
## [1] 0.6296397
```

Second get the theoretical variance of the distribution which is sigma^2/n

```
theoretical_variance<-(1/lambda)^2/average_number
theoretical_variance
```

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

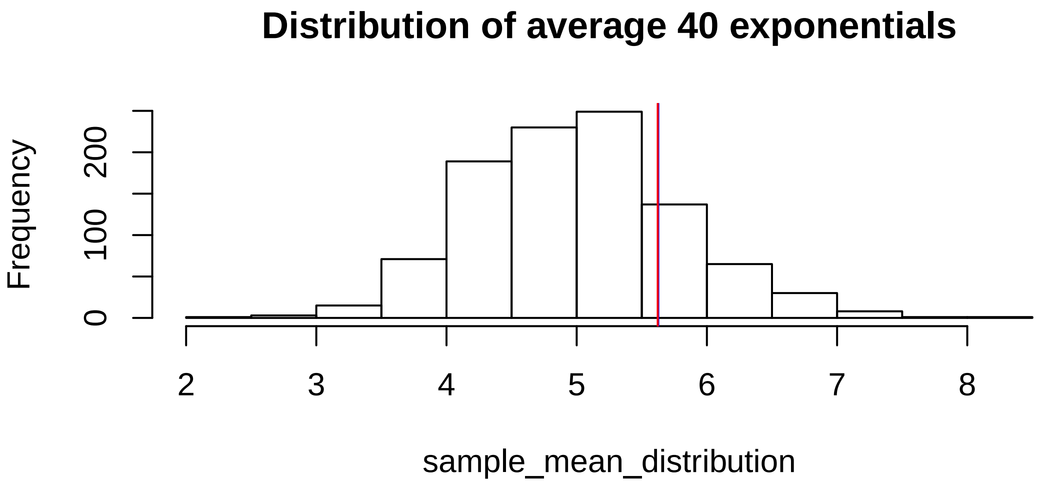
Compare sample variance and theoretical variance

```
sample_variance-theoretical_variance
```

```
## [1] 0.004639671
```

Plot the distribution of average 40 exponentials with sample variance and theoretical variance frome the sample mean, the blue line indicates the sample variance and red line stands for the theoretical variance.

```
hist(sample_mean_distribution,main="Distribution of average 40 exponentials")
abline(v= sample_mean+sample_variance, col = "blue", lwd = 1)
abline(v= sample_mean+theoretical_variance, col = "red", lwd = 1)
```

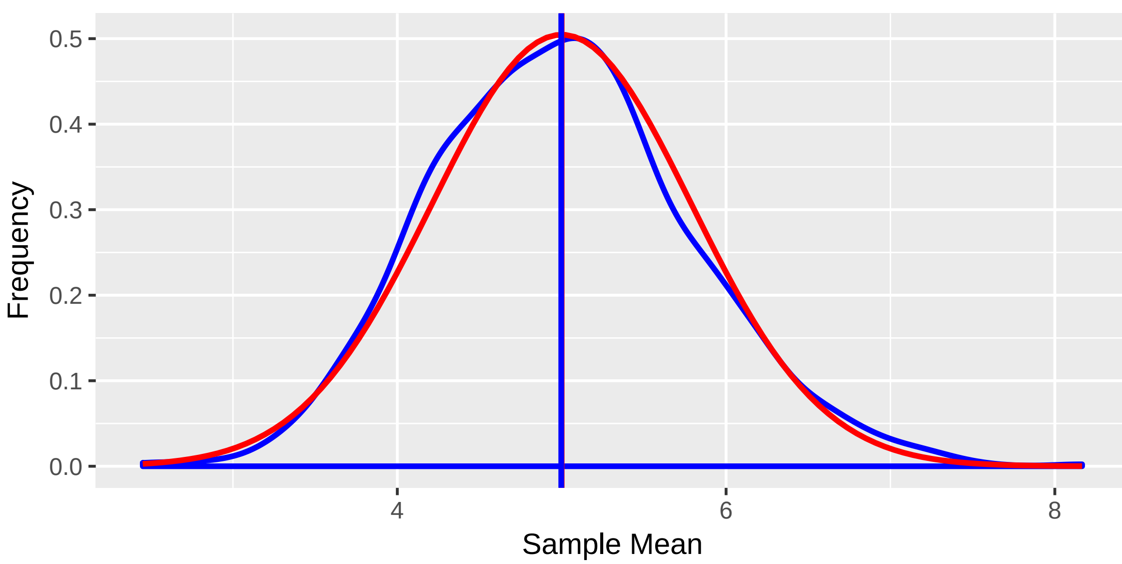


Those two lines are pretty close, so CLT worked again.

Distribution

Plot the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

```
library(ggplot2)
df<-data.frame(sample_mean_distribution)
g <- ggplot(df,aes(x=sample_mean_distribution))
g + geom_density(colour="blue",size=1) +
        labs(y="Frequency") + labs(x="Sample Mean") +
        stat_function(fun=dnorm, args=list(mean=theoretical_mean, sd=sqrt(theoretical_variance)), color
        geom_vline(xintercept=theoretical_mean, color="red", size=1) +
        geom_vline(xintercept=sample_mean, color="blue",size=1)
```



The plot shown above indicates that the distribution of 40 exponentials. The shape of the distribution is like a bell curve and symmetric from the distrition mean. So it is far more Gaussian than the original distribution!