Exponential distribution in R and compare it with the Central Limit Theorem

Tejaswi N Rao

26th July 2018

Overview: In this report we 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. We will be Setting lambda = 0.2 for all of the simulations. we will investigate the distribution of averages of 40 exponential.

Through this report we are going to carry out the below given tasks

- Show the sample mean and compare it to the theoretical mean of the distribution.
- Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
- Show that the distribution is approximately normal.

Simulations:

1) Sample Mean versus Theoretical Mean

The sample mean or empirical mean and the sample covariance are statistics computed from a collection (the sample) of data on one or more random variables.

To analyse the issue, as per the given assignment we are taking distribution of 1000 random uniforms, We will run a series of 1000 simulations, each simulation will contain 40 observations and the exponential distribution function will be set to “rexp(40, 0.2)”, so here are the known values and variables.

```
sims = 1000;
n = 40; ## number of distributions
lambda = 0.2; ## number of simulations
means <- vector("numeric")
means_sum <- vector("numeric")
means_cum <- vector("numeric")
```

now we are calculating the mean

```
for (i in 1:sims) { means[i] <- mean(rexp(n, lambda))}
means_sum[1] <- means[1]
for (i in 2:sims) { means_sum[i] <- means_sum[i-1] + means[i] }
for (i in 1:sims) { means_cum[i] <- means_sum[i]/i }
```

now with the above mean calculation

The sample means

```
means_cum[sims]
```

answer

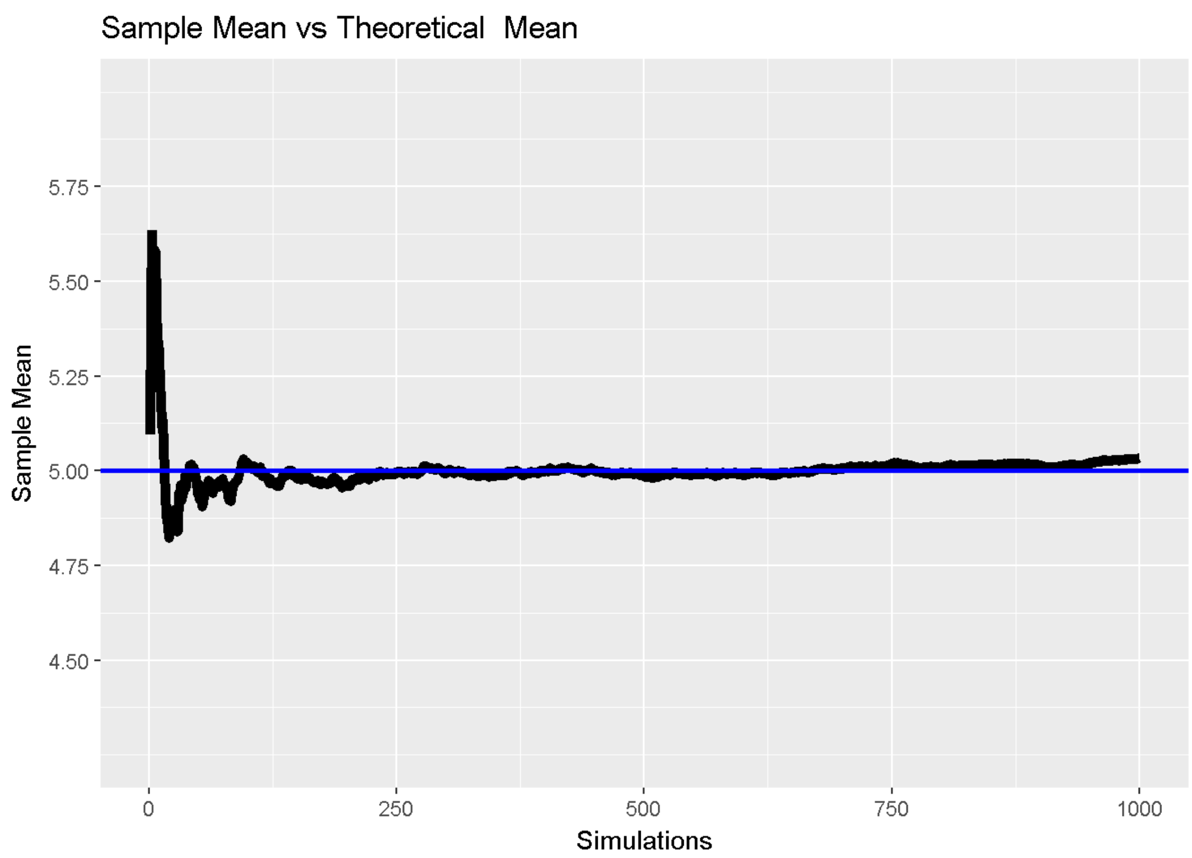
```
## [1] 5.033982
```

The theoretical mean

```
1/lambda
## [1] 5
```

now if we plot both on a graph using ggplot2

```
library(ggplot2)
## Warning: package 'ggplot2' was built under R version 3.3.2
g <- ggplot(data.frame(x = 1:sims, y = means_cum), aes(x = x, y = y))
g <- g + geom_hline(yintercept = 0) + geom_line(size = 2)
g <- g + geom_abline(intercept = 1 / lambda, slope = 0, color = "blue",
size = 1)
g <- g + scale_y_continuous(breaks=c(4.50, 4.75, 5.00, 5.25, 5.50, 5.75),
limits=c(4.25, 6))
g <- g + labs(title="Sample Mean vs Theoretical  Mean")
g <- g + labs(x = "Simulations", y = "Sample Mean")
print(g)
## Warning: Removed 1 rows containing missing values (geom_hline).
```



3)Sample Variance versus Theoretical Variance

We will compare the variance present in the sample means of the 1000 simulations to the theoretical variance of the population.

as per above calculation, the variance of sample mean

```
var(means)*n
## [1] 25.62674
```

theoretical variance

```
(1/lambda)^2
## [1] 25
```

As per the result, we can see variance of sample mean is 25.73227 and theoretical variance is 25 which is almost same and comparable.

4)Distribution of Sample Means vs Normal Distribution

As per above results, lets plot the result on graph

```
library(ggplot2)
g <- ggplot(data.frame(x = means), aes(x = x))
```

```
g <- g + geom_histogram(position="identity", fill="yellow", color="black",
alpha=0.2,binwidth=0.5, aes(y= ..density..))

g <- g + stat_function(fun = dnorm, colour = "red", args=list(mean=5))

g <- g + scale_x_continuous(breaks=c(1, 2, 3, 4, 5, 6, 7, 8, 9),
limits=c(1, 9))

g <- g + scale_y_continuous(breaks=c())

g <- g + theme(plot.title = element_text(size=12, face="bold", vjust=2,
hjust=0.5))

g <- g + labs(title="Distribution of Samle Means vs Normal Distribution")

g <- g + labs(x = "Sample Mean", y = "Frequency")

print(g)
```