Statistical Inference - Simulation project

H.Harvey

25 December 2015

Introduction

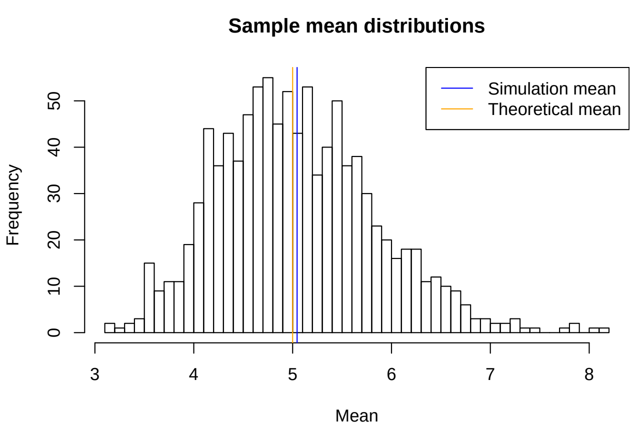
The purpose of this project is to investigate the exponential distribution in R and compare it with the Central Limit Theorem.

Calculations of the sample mean

```
set.seed(10) # R Preparation for reproducible results
lambda <- 0.2 # set lambda to 0.2
Num_Sample <- 40 # 40 samples
Num_Sim <- 1000 # 1000 simulations
Sim_Exp <- replicate(Num_Sim, rexp(Num_Sample, lambda))
mean_Sim_Exp <- apply(Sim_Exp, 2, mean) # Calculate the mean of the sample of exponentials
mean_Sample <- mean(mean_Sim_Exp) # Calculate the mean of the sample mean
mean_Theory <- 1/lambda # Calculate the theoritical mean
```

Graphical representation of the central/theoritical mean

```
# Represent the sample mean distribution
hist(mean_Sim_Exp, breaks=50, xlab = "Mean", main = "Sample mean distributions")
abline(v = mean_Sample , col = "blue")
abline(v = mean_Theory, col = "orange")
legend('topright', c("Simulation mean", "Theoretical mean"),
       lty=c(1,1), col=c("blue", "orange"))
```




```
round(mean_Sample, digits = 3)
```

```
## [1] 5.045
```

```
mean_Theory
```

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

The distribution of sample means is centered at 5.045 and the theoretical center of the distribution is calculated at 5. The center of distribution of averages of 40 exponentials is very close to the theoretical center of the distribution.

Calculation of the variance and standard variation

```
# Calculate the standard deviation of the sample distribution
stddev_Sample <- sd(mean_Sim_Exp)
# Calculate the  theoritical standard deviation
stddev_Theory <- (1/lambda)/sqrt(Num_Sample)
# Calcultate the variance of the sample distribution
var_Sample <- stddev_Sample^2
# Calculate the theoritical variation
var_Theory <- ((1/lambda)*(1/sqrt(Num_Sample)))^2
```

```
round(stddev_Sample, digits = 3)
```

```
## [1] 0.798
```

```
round(stddev_Theory, digits = 3)
```

```
## [1] 0.791
```

```
round(var_Sample, digits = 3)
```

```
## [1] 0.637
```

```
var_Theory
```

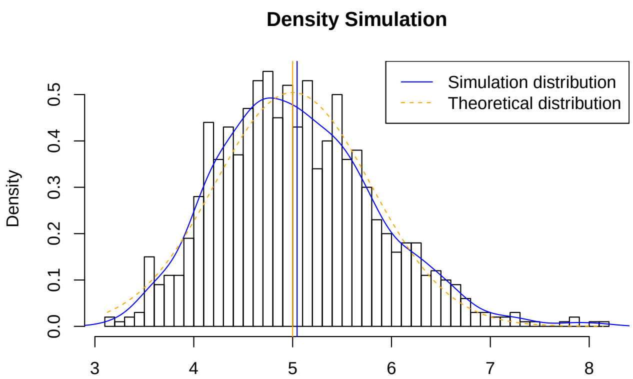
```
## [1] 0.625
```

The standard variation is 0.798 where the theoeritical standard deviation is calculated at 0.791. The variance of sample means is 0.637 where the theoretical variance of the distribution is calculated at 0.625.

Graphical representation of the distribution

```
hist(mean_Sim_Exp, breaks=50, prob=TRUE,
     main="Density Simulation",
     xlab="")
# Calculate and identify the density of the sample mean
lines(density(mean_Sim_Exp), col="blue")
```

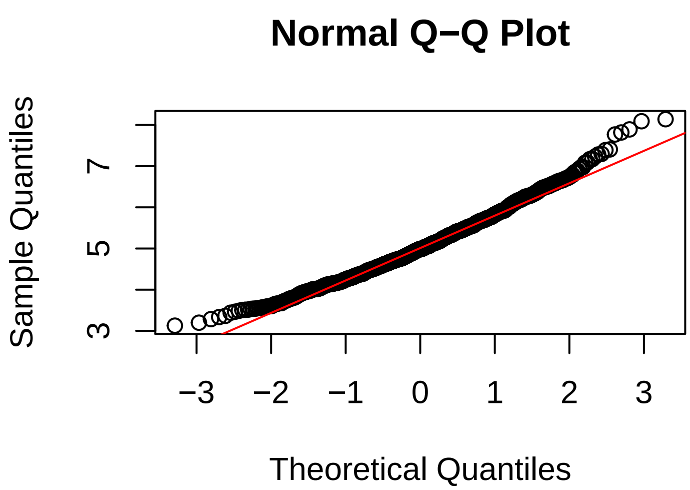
```r
abline(v=mean(mean_Sim_Exp), col="blue")
abline(v=1/lambda, col="orange")

# Calculate the theoretical density of the averages of samples
xfit <- seq(min(mean_Sim_Exp), max(mean_Sim_Exp), length=100)
yfit <- dnorm(xfit, mean=mean_Theory, sd=stddev_Theory)
lines(xfit, yfit, pch=22, col="orange", lty=2)
legend('topright', c("Simulation distribution", "Theoretical distribution"),
       lty=c(1,2), col=c("blue", "orange"))
```



As represented in the graph, the mean of the random sampled exponantials distributions overlaps with the theoritical mean of the distribution, due to the Central Limit Theorem.

Validation

```r
qqnorm(mean_Sim_Exp)
qqline(mean_Sim_Exp, col = 2)
```



Conclusion

Following the central limit theorem (CLT), we can claim that the distribution of averages of 40 exponentials is very close to a normal distribution.