Data Science Statistical Inference - Exponential Distribution

Dr Paul Fergus

Introduction

In this report we demonstrate how the exponential distribution can be simulated using the rexp(n, lambda) function in R. Using the exponential distribution the distribution of its averages will be presented.

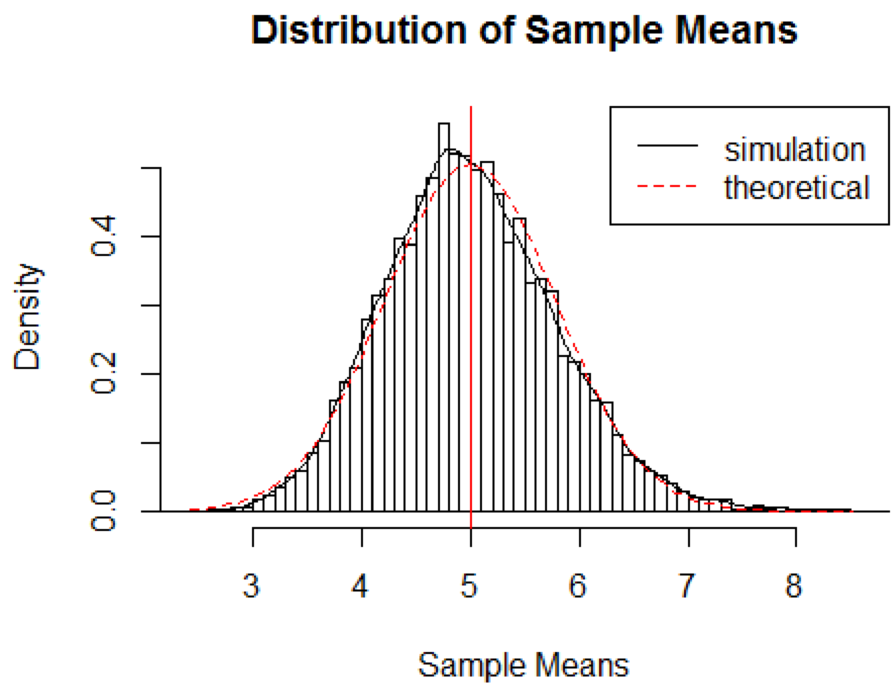
Methodology

The rate parameter for each simulation is %= 0.2$. Within each simulation 40 random variables were selected using the rexp function. Using this configuration the mean and standard deviation for the 40 values are calculated over 1000 simulations.

```
#Set the random seed so that our random values reproducible.
set.seed(1)

#Set our Lambda value
lambda <- 0.2;
#Set our random variables value
n <- 40;
#Set our number of simulations
sim <- 10000;

#Set our plotting parameters to create a 2x2 plot
par(mfrow = c(2,2))

#Run the simulations
sim.results <- matrix(rexp(sim*n, rate=lambda), sim, n)
row.means <- rowMeans(sim.results)
```

The histogram below illustrates the distribution of the sample means.

```
# Display a histrogram showing the distribution of sample means
hist(row.means, breaks=50, prob=TRUE,
     main="Distribution of Sample Means",
     xlab="Sample Means")

# Show the density of the sample means
lines(density(row.means))
# Show the theoretical center of our distribution
abline(v=1/lambda, col="red")
# show the theoretical density of sample means
xfit <- seq(min(row.means), max(row.means), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(n)))
lines(xfit, yfit, pch=22, col="red", lty=2)
# add a legend for clarity
legend('topright', c("simulation", "theoretical"), lty=c(1,2),
col=c("black", "red"))
```

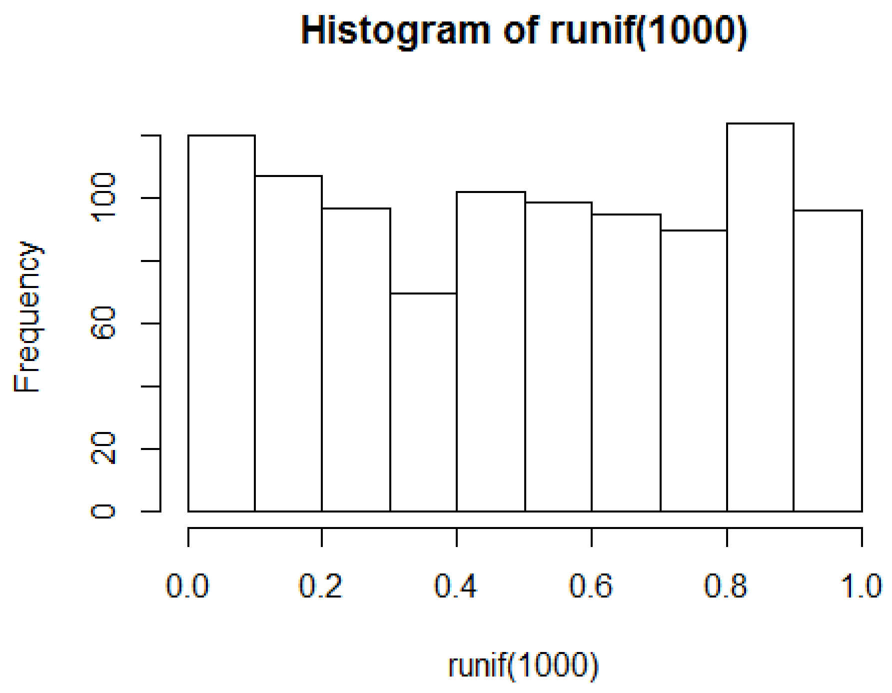


From the above figure it shows the distribution of sample means. The distribution of sample means is centered around 5.0028732, which the theoretical center of the distribution is 5. The variance (or spread) around the mean is 0.6261976 and the theoretical variance is 0.625. We can see that the simulation results (or distribution) is centered and that it closely resembles the theoretical center of the distribution. Futhermore, we find that the simulation variance maps onto the theoretical variance of the distribution. Finally it is clear that the distribution follows a normal distribution (or Gaussian distribution).

The reason that the distribution follows the normal distribution is based on the theoretical central limit theorem, which states that given certain conditions, the mean of a sufficiently large number of simulations of independent random varaibles, will be approximately normally distributed, regardless of the underlying distribution.

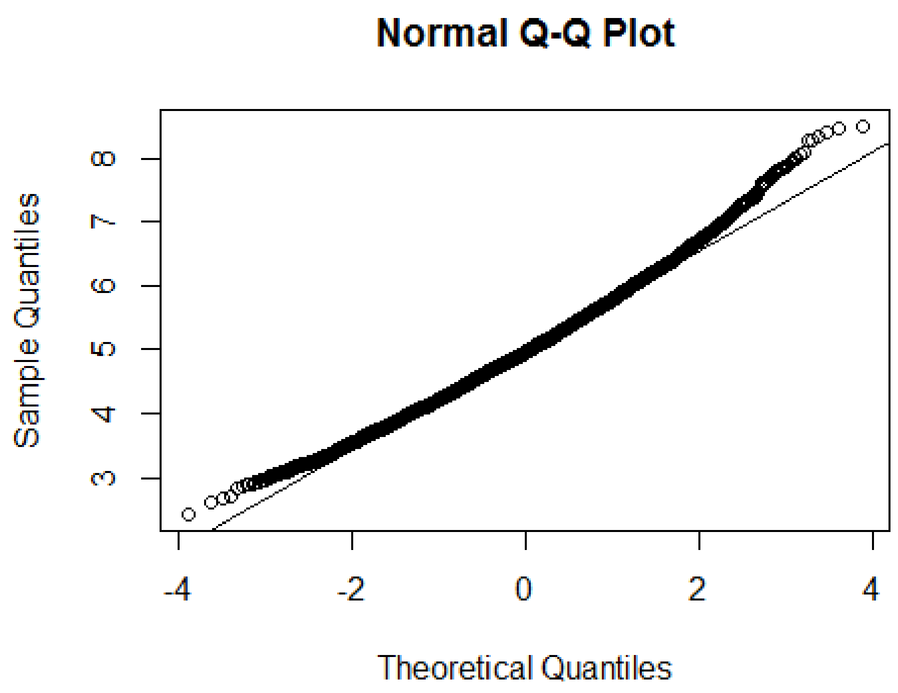
Interestingly the above distribution of 1000 averages of 40 random uniforms can be compared with the below distribution of 1000 random uniforms. We can clearly see that the distribution is not normal. This is due to the fact that we did not apply the central limit theorem, i.e. the mean of 40 random uniforms over a 1000 simulations.

```
hist(runif(1000))
```



Finally, the qq plot below shows us that the distribution is approximaltely normal.

```
qqnorm(row.means); qqline(row.means)
```



The plot shows us that the theoretical and sample quantiles roughly follows a normal distribution apart from some outliers at both tails.