Simulating Exponential Distributions

Synopsis

The purpose of this data analysis is to investigate the behaviour of random numbers drawn from the Exponential distribution and compare their distribution to values expected when assuming that the Central Limit Theorem applies. The analysis will avail of the many tool available from the R programming language.

This investigation will compare the distribution of averages of 40 exponentials over 1000 simulations. For this study, the lambda will be set to 0.2 for all of the simulations.

Knowledge of the parameters of the Exponential distributon and the Central Limit theorem are assumed herein.

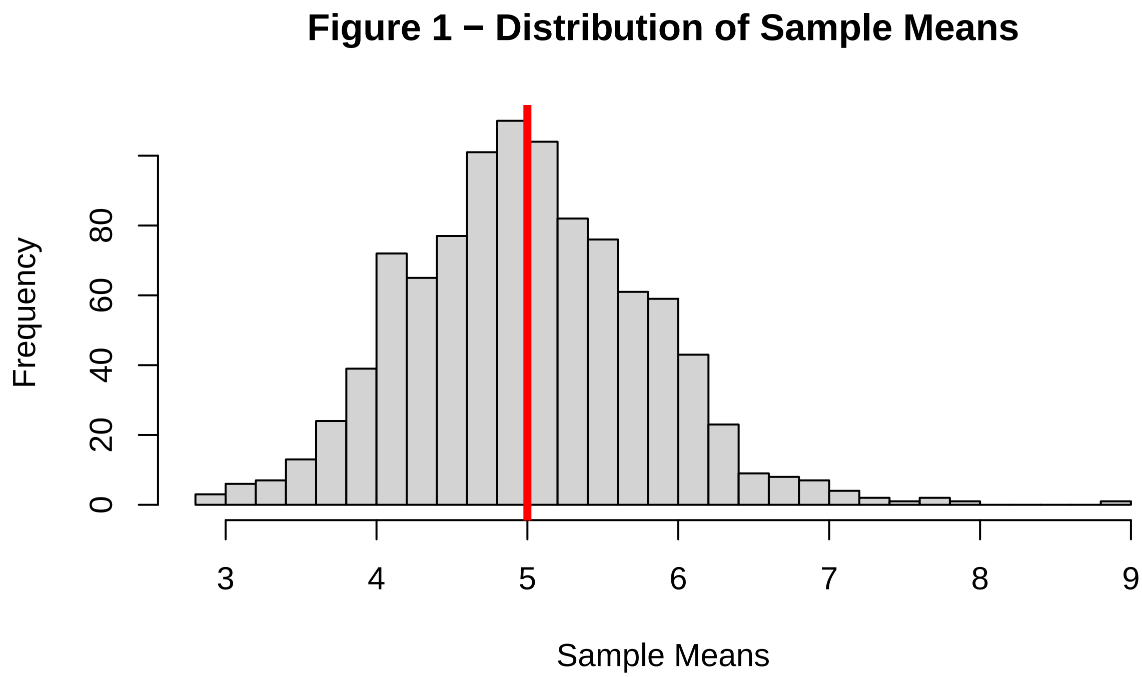
Simulation

We start by simulating 1000 trials with 40 observations each. Thapproach used is with a "for" loop and the rexp() function. The means and standard deviations for each trial are saved in respective vectors.

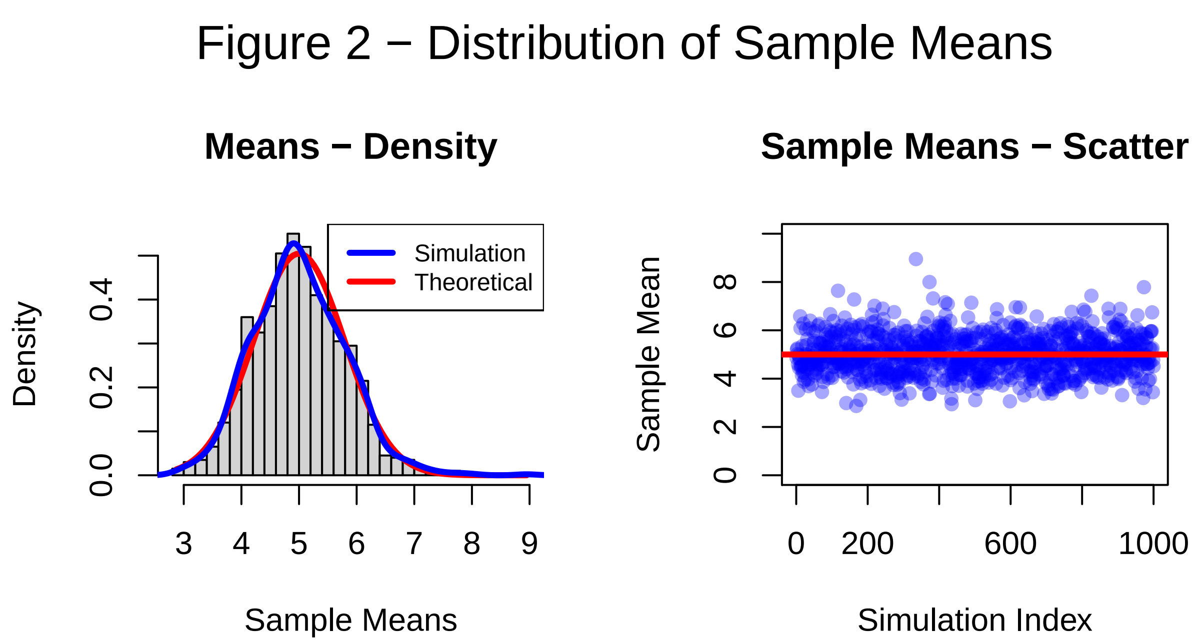
```
lambda <- 0.2 #rate paramter for exponential distribution
n <- 40 #sample size
M <- 1000 #number of iterations
mymeans <-  numeric(M) #container for sample means
mysds <- numeric(M) #container for sample standard deviations

for (i in 1:M) {
       set.seed(i) #use seed corresponding to each iteration index
       mydata = rexp(n, lambda) #generate a sample
       mymeans[i] = mean(mydata) #save the sample mean
       mysds[i] =  sd(mydata) #save the sample standard deviation
}
```

Sample Mean vs Theoretical Mean Using the data we can calculate the mean of all the collected samples, and compare that to the theoretical mean of the exponential distribution.

```
sample_mean = mean(mymeans)
# observed mean for all samples

#theoretical mean : mu
mu = 1/lambda

#sampling fluctuation
fluctuation = mu-sample_mean

hist(mymeans, breaks=25,
     main = "Figure 1 - Distribution of Sample Means",
     xlab = "Sample Means")
abline(v=5,lw=4,col="red")
```



The results show that the sample mean is off by only `-0.002327`. In the figure below we can further see that the distributions of the sample means is approximately normal.



Sample Variance vs Theoretical Variance Next we calculate the standard deviations and variances of the sample data, and their theoretical counterparts.

```
sample_var <- var(mymeans)
sample_var
```

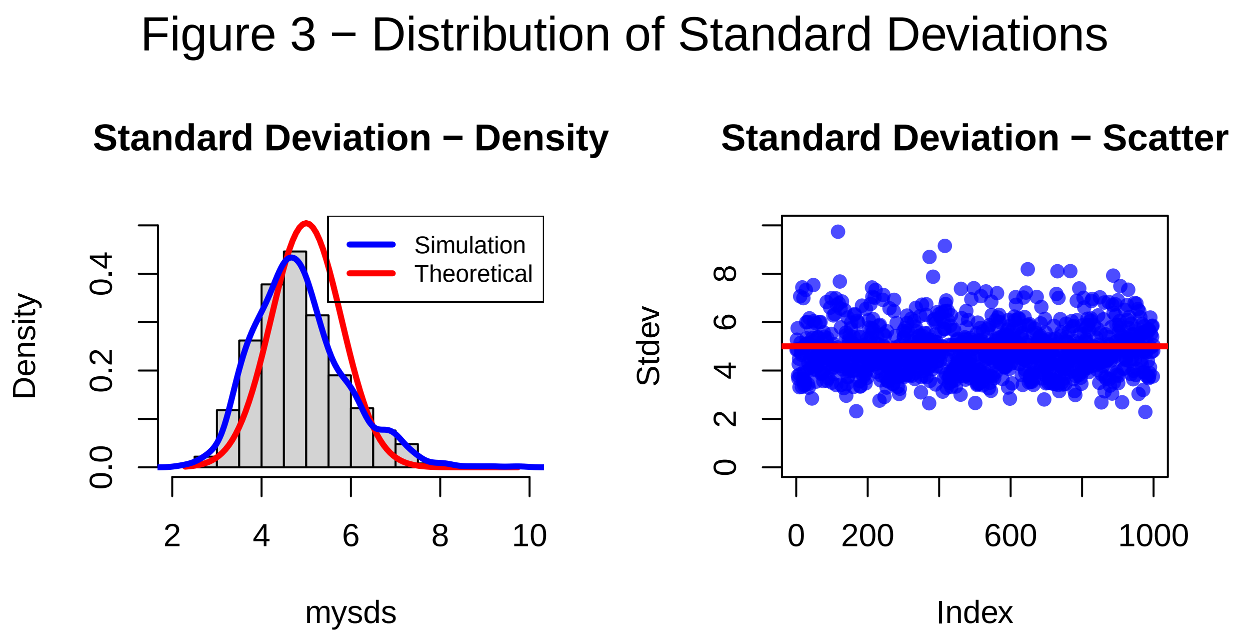
```
## [1] 0.6308244
```

```
sigma <- (mu^2)/n
sigma
```

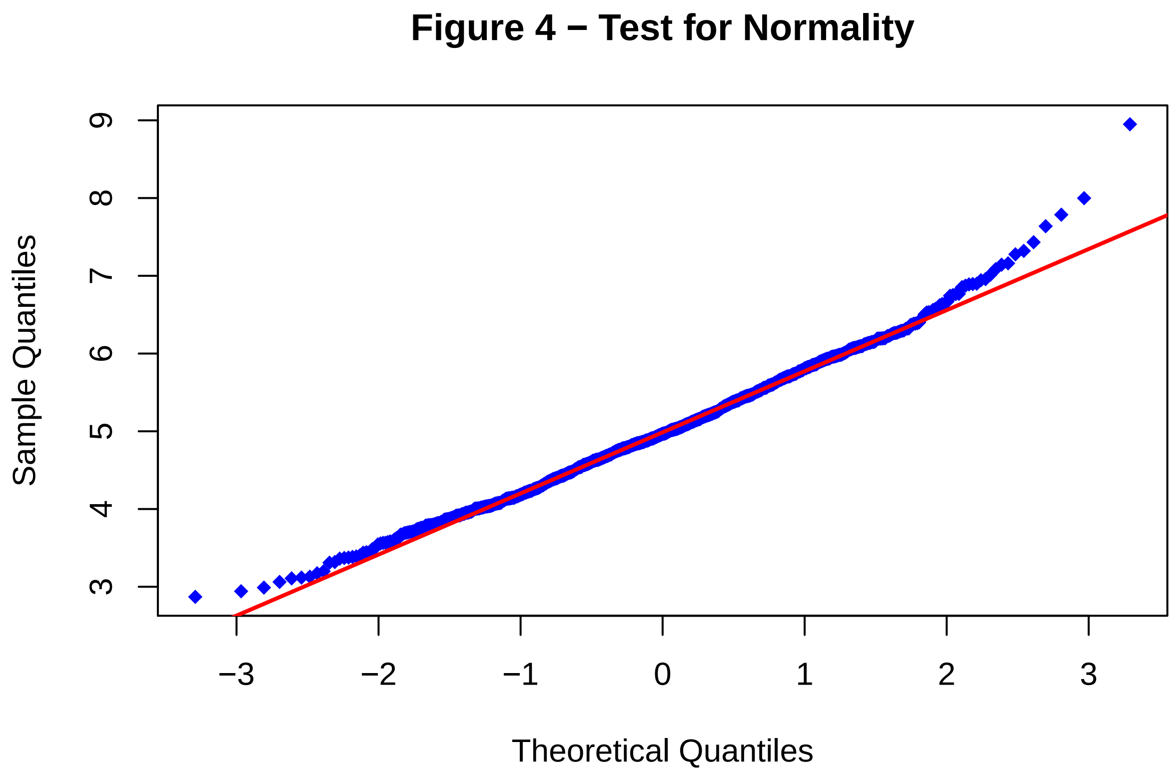
```
## [1] 0.625
```

The variance of sample means is numerically close to the theoretical variance according to the central limit theorem, i.e 0.6308 and 0.625 respectively. The next figure is a chart comparing the distributions of theoretical standard deviation versus that of the standard deviations of the sample data.

The similarity of sample and theoretical distribution is less satisfactory, when compared to that of the sample means. A valid conclusion is that the central limit theorem doesn't hold in the case of sample variances, and that an alternative model should be sought to characterize the distribution.



Finally, we can test the normality using a normaly probability ("q-q") plot.



Appendix - Plot Codes

```
hist(mymeans, breaks=25,
     main = "Figure 1 - Distribution of Sample Means",
     xlab = "Sample Means")
abline(v=5,lw=4,col="red")
```

Figure 1

```
#Setting up graphs side by side
par(mfrow=c(1,2), oma=c(2,0,2,0))

#Generating histogram as a density function
hist(mymeans, breaks = 25,
     freq = FALSE,
     main = "Means - Density",
     xlab = "Sample Means")

#Fitting the theoretical line
xfit <- seq(min(mymeans), max(mymeans), length = 100)
```

```r
yfit <- dnorm(xfit, mean = 1/lambda, sd = mu/sqrt(n))
lines(xfit, yfit, pch=22, lty = 1, lw = 3, col = "red")

#Fitting the line to the data
lines(density(mymeans), lw = 3, col = "blue")
mtext("Figure 2 - Distribution of Sample Means",
      outer = TRUE, cex = 1.5)
#Adding a legend and title
legend('topright', c("Simulation", "Theoretical"),
       col=c("blue", "red"), lw=c(3,3), cex = 0.75)
#plotting the data on the mean for the 1000 simulations

plot(mymeans, main = "Sample Means - Scatter", ylab = "Sample Mean",
     xlim = c(0,M), ylim = c(0,10),
     xlab = "Simulation Index", pch = 16,
     col = adjustcolor("blue", alpha=0.35))
abline(h=5, col = "red", lw =3)
```

Figure 2

```r
par(mfrow=c(1,2), oma=c(2,0,2,0))
hist(mysds, 22, freq = FALSE, ylim = c(0,0.5),
     main = "Standard Deviation - Density")

#Fitting the theoretical line
xfit <- seq(min(mysds), max(mysds), length = 100)
yfit <- dnorm(xfit, mean = 1/lambda, sd = mu/sqrt(n))
lines(xfit, yfit, pch=22, lty = 1, lw = 3, col = "red")

#Fitting the line to the data
lines(density(mysds), lw = 3, col = "blue")

#Adding a legend and title
legend('topright', c("Simulation", "Theoretical"),
       col=c("blue", "red"),
       lw=c(3,3), cex = 0.75)
mtext("Figure 3 - Distribution of Standard Deviations",
      outer = TRUE, cex = 1.5)
plot(mysds, main = "Standard Deviation - Scatter",
     ylab = "Stdev", xlim = c(0,s=M),
     ylim = c(0,10),pch = 16,
     col = adjustcolor("blue", alpha=0.7))
abline(h=5, col = "red", lw =3)
```

Figure 3

```r
qqnorm(mymeans, main= "Figure 4 - Test for Normality",
       pch=18, col="blue")
```

```
qqline(mymeans, col="red", lw = "2")
mtext("Figure 4", outer = TRUE, cex = 1.5)
```

Figure 4