Exponential Distribution: A Simulation Comparison

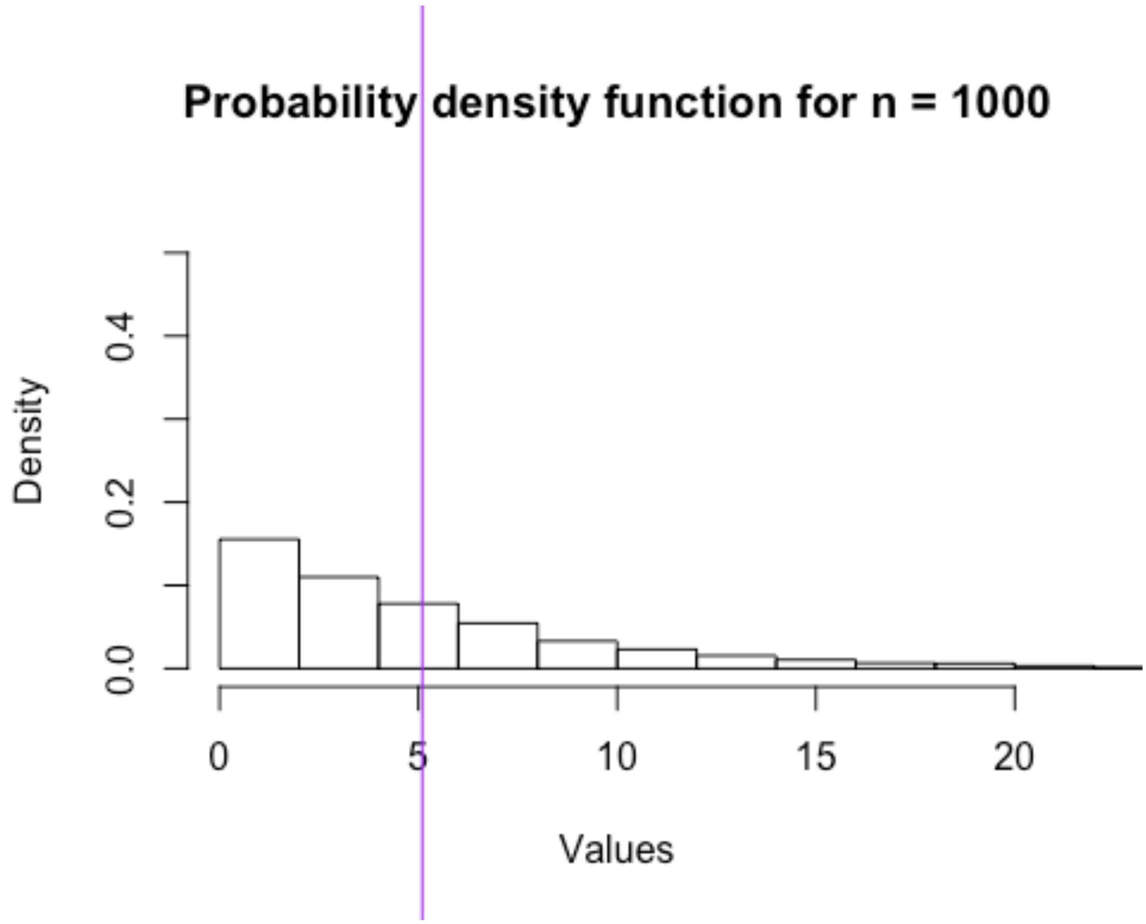
Introduction

This is a report done on the simulations of the exponential distribution in R. The distribution of the mean of the resulting sample exponential distributions is then investigated and compared against theoretical assumptions.

Preliminaries: The Exponential Distribution

We first generate a large collection of randomly selected exponential distribution and plot the outcome so we could inspect the shape of the distribution. For this initial exercise, we set n = 1,000 and $\lambda = 0.2$. We then compute for the mean and standard deviation and compare this with the theoretical mean and standard deviation for an exponential distribution.

```
# Setting the random seed
set.seed(1211)
par(xpd=NA)

# Setting the assumptions
lambda <- .2
n <- 40
nosim <- 1000

# Generating 1000 random exponential distributions
ranex <- rexp(1000, lambda)

par(mfrow = c(1,1))
myhist <- hist(ranex , freq = FALSE, xlim = c(0, 20), ylim = c(0, .55),
               breaks = 25, main = paste("Probability density function
for n = 1000"),
               xlab = "Values")

# Calculating the mean and standard deviation of the 1000 simulations
meanex <- mean(ranex)
sdex <- sd(ranex)

# Plotting the mean value from the 1000 simulations
abline(v = meanex , col = "purple", lwd = 1, lty = 1)
```



The theoretical mean of the exponential distribution is computed as 1/λ and the standard deviation also as 1/λ.

```
emean <-1/lambda
esd <- 1/lambda
```

These are then compared to the sample mean and sample standard deviation of the generated data set.

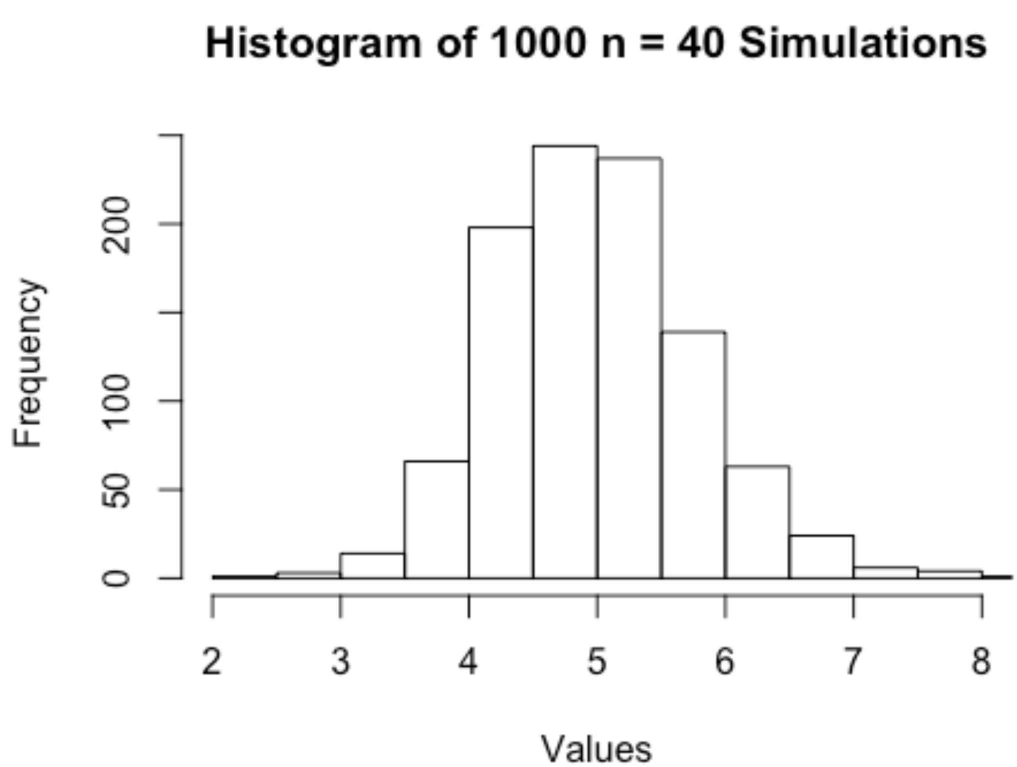
```
emean; meanex

## [1] 5

## [1] 5.096788

esd; sdex

## [1] 5

## [1] 4.881025
```

As we can see, the sample mean and sample standard deviation are close to the expected theoretical values for mean and standard deviation of a exponential distribution.

Simulations: A Thousand Runs of Forty Sample Sized Exponential Distributions

We are now going to do 1000 simulations of exponential distribution with sample size equal to 40 and $\lambda = 0.2$. The mean and standard deviation of each of these iterations are then calculated. We then plot the means of the iterations to show the shape of the distribution.

```
# Clearing the vectors
mean_val <- NULL
mean_sd <- NULL

for (i in 1:nosim){
  # Calculating the mean & sd of the sample means
  val <- rexp(n, lambda)
  means <- mean(val); sd <- sd(val)
  mean_val  <- c(mean_val, means); mean_sd <- c(mean_sd, sd)
}

# Plotting of the means
myhist <- hist(mean_val , freq = TRUE, xlim = c(2, 8),
               main = paste("Histogram of", nosim, "n = 40
Simulations"), xlab = "Values")
```



Comparisons: Sample Mean versus Theoretical Mean; Sample Variance versus Theoretical Variance

We now calculate the mean and variance of the generated sample means as well as the theoretical mean and variance of the exponential distribution given the following computations:

```
# Calculating the mean, standard deviation and variance of the
aggregated sample means
smean <- mean(mean_val)
ssd <- sd(mean_val)
svar <- ssd^2

# Calculating the theoretical mean, standard deviation and variance of
X-bar
tmean <-1/lambda
tsd <- (1/lambda) / sqrt(n)
tvar <- (1/lambda)^2 / n
```

Inspecting the calculated mean of the sample means, we can see that it approximates the theoretical mean of the exponential distribution:

```
# Comparing the Sample Mean and Theoretical Mean of Xbar
smean; tmean

## [1] 4.99286

## [1] 5
```

To show the variability of the sample, we now look at the computed variance of the sample means and compare it with the variance of the exponential distribution. We can see the the sample variance is close to the theoretical variance:

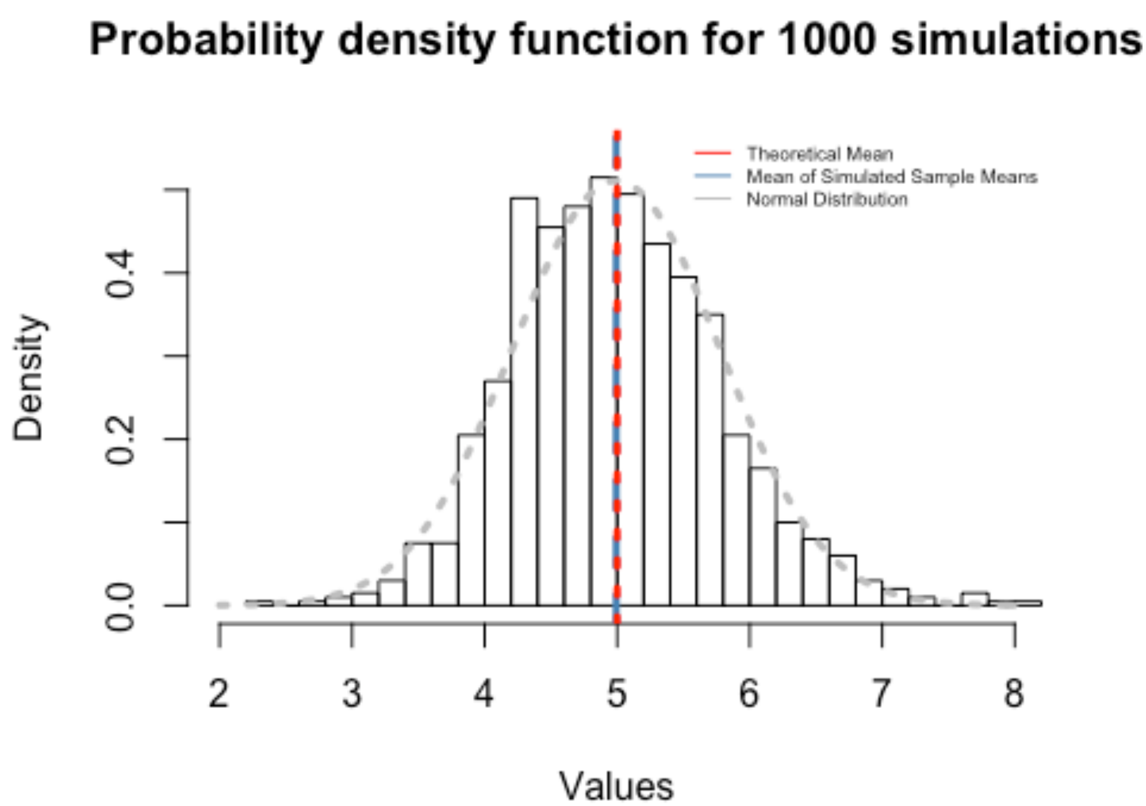
```
# Comparing the Sample Variance and Theoretical Variance of Xbar
svar; tvar

## [1] 0.6092425

## [1] 0.625
```

Is the distribution Gaussian?

Now we take a look at the shape of the distribution to see if it approximates a Normal distribution. Recalling the Central Limit Theorem, the distribution of the averages of independent and identically distributed random variables approximates a Normal distribution, that is $\overline{X} \approx N(\mu, \sigma^2/n)$.

```
# Investigating the Distribution of the Simulated sample means from the
Exponential Distribution
# Plotting the Histogram of the Probability Density of the sample mean
par(mfrow = c(1,1))
```

```
myhist <- hist(mean_val , freq = FALSE, xlim = c(2, 8), ylim = c(0,
.55),
               breaks = 25, main = paste("Probability density function
for", nosim, "simulations"),
               xlab = "Values")

# Plotting the theoretical distribution for simulated sample means
x <- seq(min(mean_val ), max(mean_val ), length = 100)
y <- dnorm(x, mean = smean, sd = ssd)
curve(dnorm(x, mean = smean, sd = ssd),
      col = "gray", lwd = 3, lty = 3, add = TRUE)

# Plotting the mean value from the data set
abline(v = smean , col = "steelblue", lwd = 3, lty = 2)

# Plotting the expected value of an exponential distribution
abline(v = 5, col = "red", lwd = 3, lty = 9)


legend('topright', c("Theoretical Mean", "Mean of Simulated Sample
Means", "Normal Distribution"),
       lty=1, col=c('red', 'steelblue', 'gray'), bty='n', cex=.5)
```



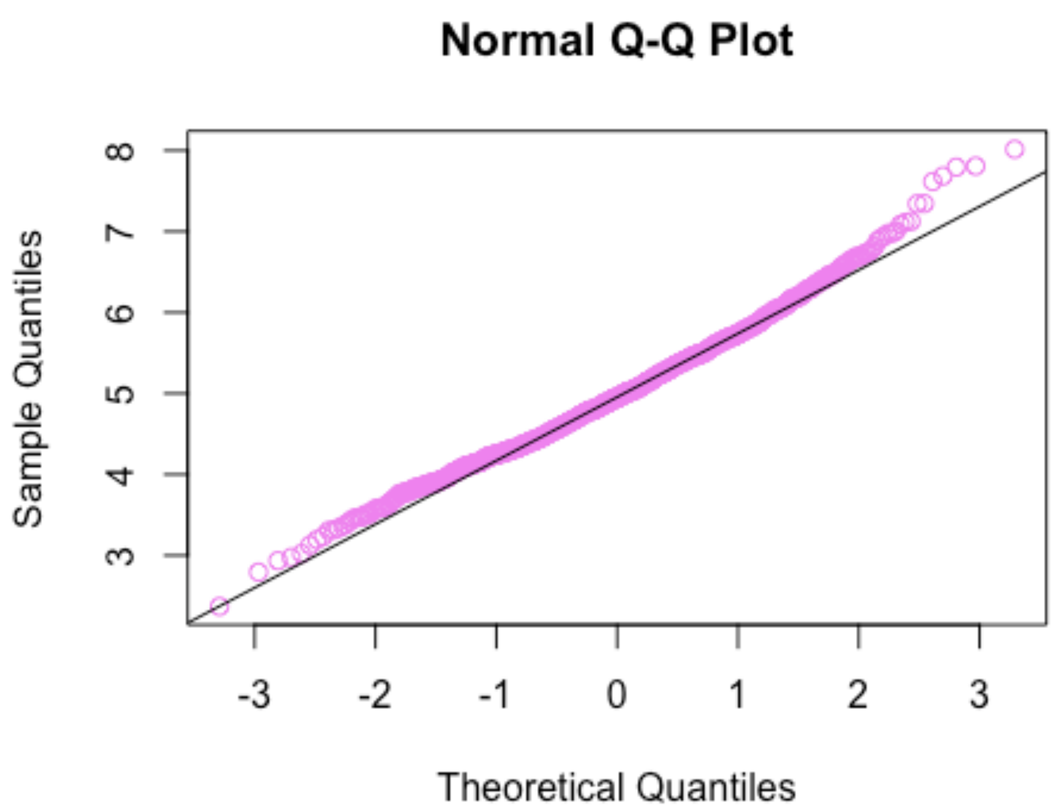
As we can see, the shape of the distribution of the sample means of the exponential distributions is bell curve centered at the mean. This distribution looks more Gaussian than the original distribution which we preliminary plotted.

Moreover, looking back at the standard deviation of the sample means we can notice that it corresponds to the standard error of the mean:

```
# Comparing Sample Standard Deviation and Theoretical Standard Error of Xbar
ssd; tsd

## [1] 0.7805399

## [1] 0.7905694
```

A Q-Q plot of the mean values is also plotted. We can further see that there is little deviation between the actual quantile values and the theoretical.

```
# Generating the Q-Q plot
qqnorm(mean_val, col = "violet")
qqline(mean_val)
```



These evidences highly indicate that the aggregated sample of means generated from simulated exponential distributions is approximately normally distributed.