

DEMONSTRATIVE SIMULATION OF THE CENTRAL LIMIT THEOREM

Oluwadare, Margaret

8/27/2020

SYNOPSIS

This project is to investigate the exponential distribution and compare it to the Central Limit Theorem. For this particular case the distribution parameter `lambda` will be set to 0.2 for all of the simulations. We will be comparing the distribution of 40 averages of exponentials distribution over 1000 simulations.

SIMULATIONS:

The exponential distribution can be simulated in R with `rexp(n, lambda)`, where `lambda` is the rate parameter and `n` is the number of observations. For the purpose of this project value of `lambda` is set to 0.2. First we load the `ggplot2` plotting library.

```
library(ggplot2)
```

```
numcases <- 1000           #how many samples to take?
lambda <- 0.2
ntimes <- 40
```

Set the seed of the random number generator for reproducibility.

```
set.seed(1234)
```

Perform simulations

```
i2 <- 1           #initialize counters

for (i in 1:ntimes)           #repeat n times
{ sample <- 0 ;
  k <- 0           #start off with an empty set of counters
  for (j in 1:i2)           # inner loop
  {
    sample <- sample + matrix(rexp(ntimes * numcases, rate = lambda), numcases)
    k <- k+1 }
  x <- sample/k
  out <- c(k, mean(x), sd(x))
}
```

compute the mean for each of the 1000 simulations(rows)

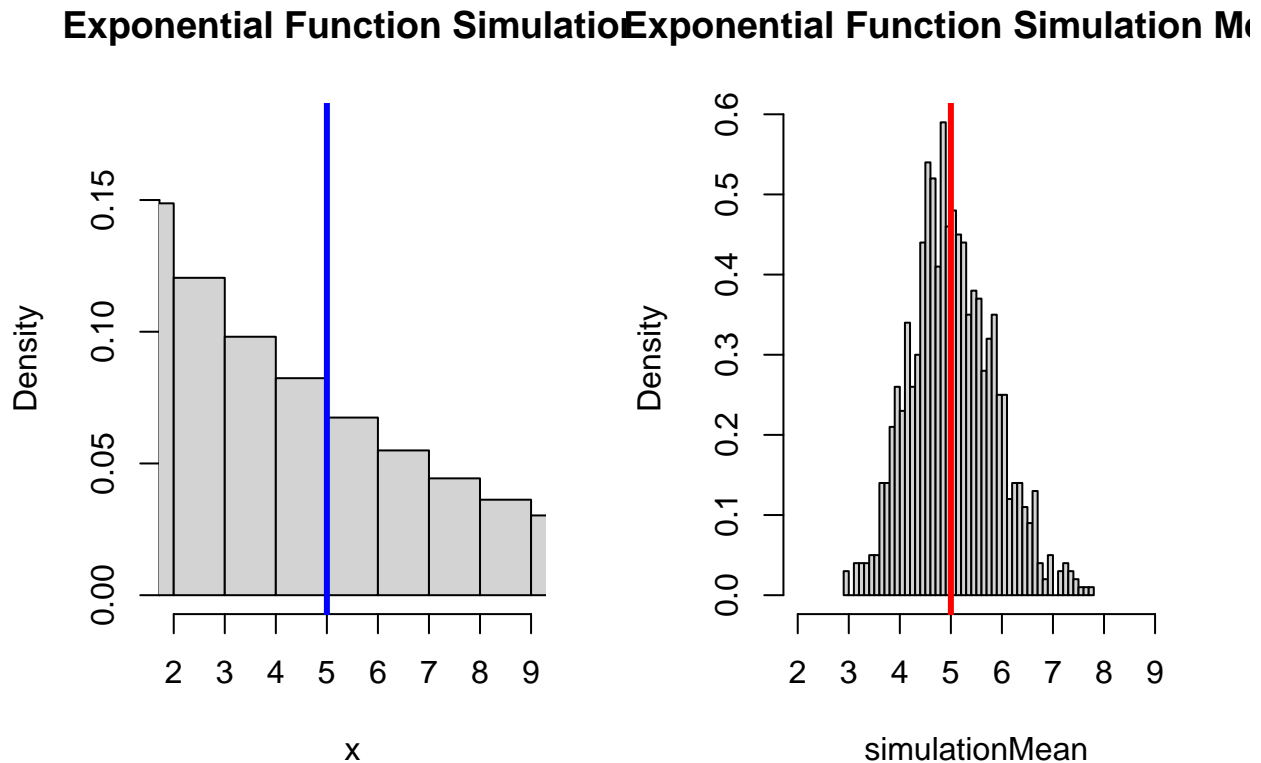
```
simulationMean <- rowMeans(x)
```

We plot the simulation data to visualize it

```
par(mfrow = c(1,2))
hist(x, prob = T, breaks = 40, xlim = c(2,9), main = paste( "Exponential Function Simulation" ))
```

```
abline(v = lambda^-1, col = "blue", lwd = 3)

hist(simulationMean, prob = T, breaks = 40, xlim = c(2,9), main = paste( "Exponential Function Simulation Mean", lambda))
abline(v = lambda^-1, col = "red", lwd = 3)
```



SAMPLE MEAN VS THEORETICAL MEAN

Sample mean is simply the mean of the simulated average values of the exponentials. Theoretical mean of the average values of exponentials is $1/\lambda$. We will plot the distribution of the average exponentials using histogram.

```
sample_mean <- mean(simulationMean)
sample_mean
```

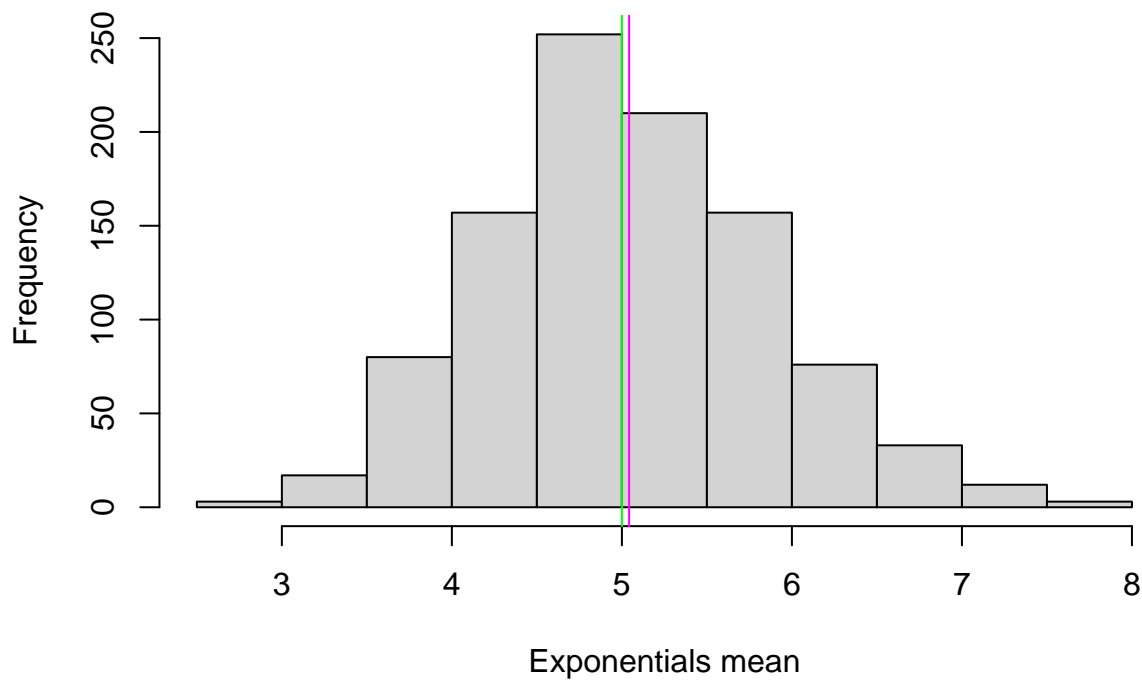
```
## [1] 5.042438
```

```
theo_mean <- 1/lambda
theo_mean
```

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

```
hist(simulationMean, xlab = "Exponentials mean", main = "Distribution of the Exponentials mean")
abline(v = sample_mean, col = "magenta")
abline(v = theo_mean, col = "green")
```

Distribution of the Exponentials mean



The sample mean is 5.042438 while the theoretical mean is 5. As seen in the histogram, the sample mean of the average exponentials (magenta vertical line) is close to the theoretical mean of an exponential distribution (green vertical line).

VARIANCE COMPARISON

The following code will compute the actual variance of the simulated mean

```
sample_var <- var(simulationMean)
sample_sd <- sd(simulationMean)
theo_var <- (1 / lambda)^2 / (ntimes)
theo_sd <- 1/(lambda * sqrt(ntimes))
```

```
sample_var
```

```
## [1] 0.6788525
```

```
sample_sd
```

```
## [1] 0.8239251
```

```
theo_var
```

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

```
theo_sd
```

```
## [1] 0.7905694
```

From our result, we can see that the actual variance (0.6789) of the simulated mean sample data is very close

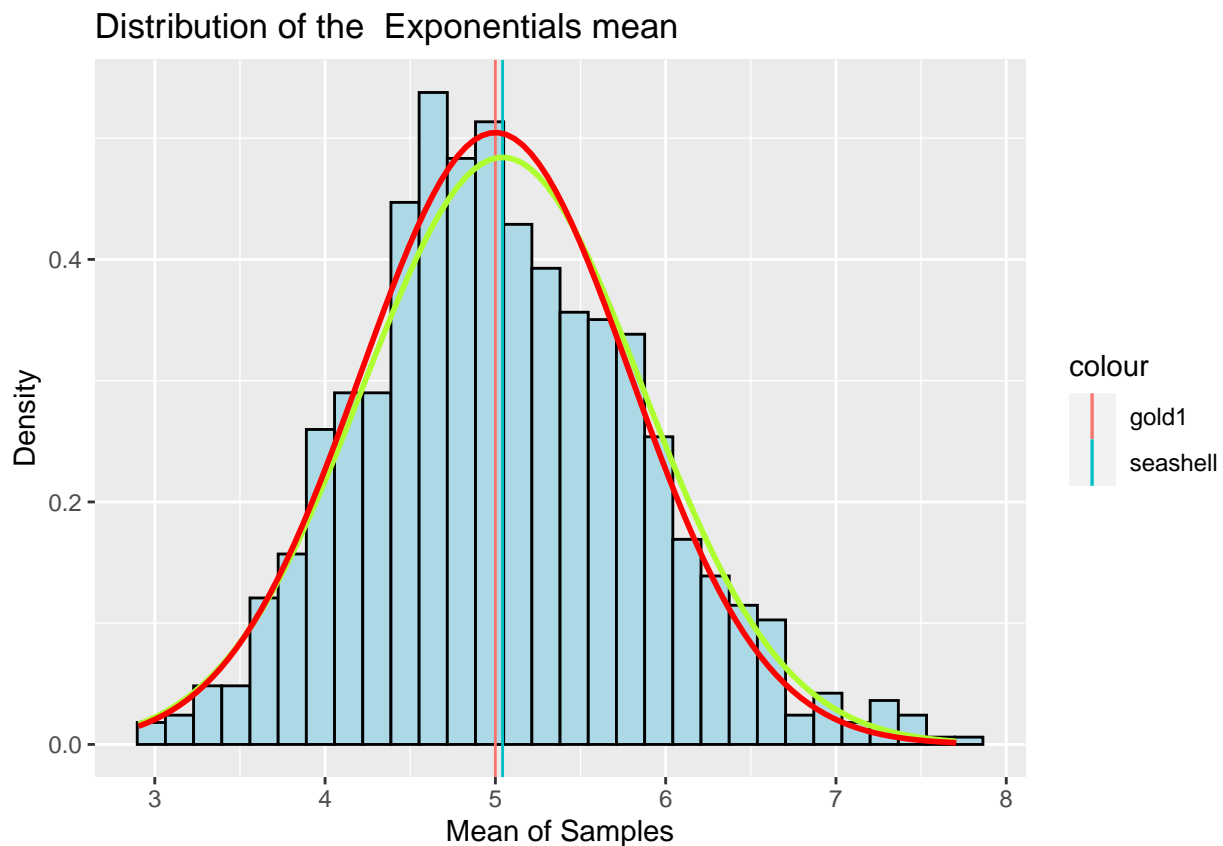
to the theoretical variance (0.625) of original data distribution.

EXAMINE THE NORMALITY OF THE EXPONENTIALS AVERAGES DISTRIBUTION

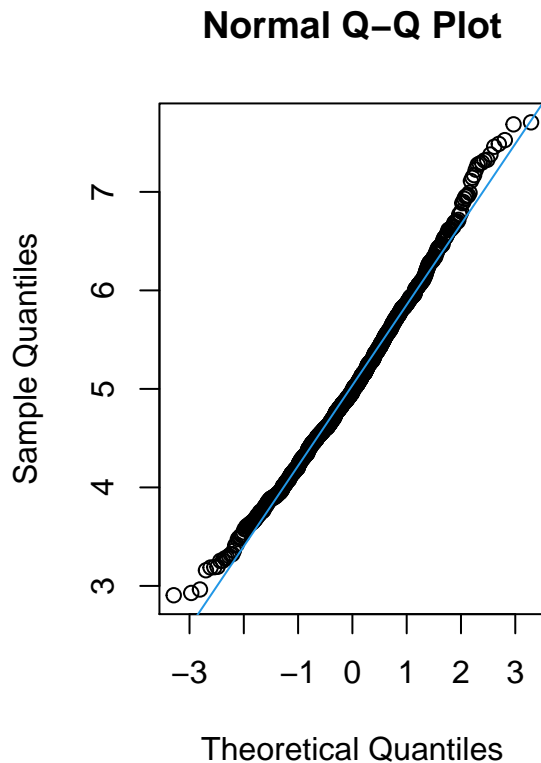
We will consider the minimum and maximum of the average exponentials. Then recreate a histogram of average exponentials with more bins. Draw the density of the average exponentials and draw the normal qqplot. The following investigates whether the exponential distribution is approximately normal. Due to the Central Limit Theorem, the means of the sample simulations should follow a normal distribution.

```
par(mfrow = c(1,2))
plotdata <- data.frame(simulationMean)
distplot <- ggplot(plotdata, aes(x = simulationMean))
distplot <- distplot + geom_histogram(aes(y = ..density..), colour = "black",
  fill = "lightblue")
distplot <- distplot + labs(title = "Distribution of the Exponentials mean", x = "Mean of Samples", y = "Density")
distplot <- distplot + geom_vline(aes(xintercept = sample_mean, colour = "seashell"))
distplot <- distplot + geom_vline(aes(xintercept = theo_mean, colour = "gold1" ))
distplot <- distplot + stat_function(fun = dnorm, args = list(mean = sample_mean, sd = sample_sd), colour = "red")
distplot <- distplot + stat_function(fun = dnorm, args = list(mean = theo_mean, sd = theo_sd), colour = "green")
distplot
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.



```
qqnorm(simulationMean)
qqline(simulationMean, col = 4)
```



The density of the actual data is shown by the seashell bars. The theoretical mean and the sample mean are so close that they nearly overlap. The “red” line shows the normal curve formed by the theoretical mean and standard deviation. The “greenyellow” line shows the curve formed by the sample mean and standard deviation. As you can see from the graph, the distribution of means of 40 exponential distributions is close to the normal distribution with the expected theoretical values based on the given lambda.

CONFIDENCE INTERVAL COMPARISON

Next we check the confidence interval levels to see how they compare.

Sample CI

To Calculate the sample confidence interval using $\text{sampleCI} = \text{mean of } x \text{ plus or minus the } 97.5\text{th normal quantile times the standard error of the mean standard deviation of } x \text{ divided by the square root of } n \text{ (the length of the vector } x\text{)}$.

```
sampleCI <- round(mean(simulationMean) + c(-1,1)*1.96*sample_sd/sqrt(ntimes),3)
sampleCI
```

```
## [1] 4.787 5.298
```

Theoretical CI

To Calculate the theoretical confidence interval we use $\text{theoCI} = \text{theoMean of } x \text{ plus or minus the } 97.5\text{th normal quantile times the standard error of the mean standard deviation of } x \text{ divided by the square root of } n \text{ (the length of the vector } x\text{)}$.

```
theoCI <- theo_mean+ c(-1,1) * 1.96 * sqrt(theo_var)/sqrt(ntimes)
theoCI
```

```
## [1] 4.755 5.245
```

The sample confidence interval is 4.787 5.298 and the theoretical confidence level is 4.755 5.245. The confidence levels also match closely. Again, proving the distribution is approximately normal.

CONCLUSION

As shown above, the distribution of means of the simulated exponential distributions follows a normal distribution due to the Central Limit Theorem. If the number of samples increase (currently at 1000), the distribution should be even closer to the standard normal distribution.