Statistical Inference Course Project (Part 1)

Title: A simulation excercise using exponential distribution

Overview

In this project we use the exponential distribution in R and compare it with the Central Limit Theorem by solving three problems. First, we find the mean of a sample from exponential distributions and compare it with the theoretical mean. Then, we compare the vaiance and lastly, we Show that the distribution is approximately normal.

Question 1: Show the sample mean and compare it to the theoretical mean of the distribution.

```
set.seed(1)
lambda = 0.2
n = 40
iterations = 1:1000

sample_means <- data.frame(expcol = sapply(iterations, function(expcol)
{mean(rexp(n, lambda))}))
```

The center of the sample distribution is calculated as:

```
mean(sample_means$expcol)
```

```
## [1] 4.990025
```

The center of the theoretical distribution is calculated as:

```
1/lambda
```

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

Therefore, the sample mean is very close to the theoretical mean.

Question 2: Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution

The standard deviation of the sample distribution is calculated as:

```
sd(sample_means$expcol)
```

```
## [1] 0.7817394
```

The standard deviation of the theoretical distribution is calculated as:

```
(1/lambda)/sqrt(40)
```

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

The variance of the sample distribution is calculated as:

```
var(sample_means)
```

```
##           expcol
## expcol 0.6111165
```

The variance of the theoretical distribution is calculated as:

```
((1/lambda)/sqrt(40))^2
```

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

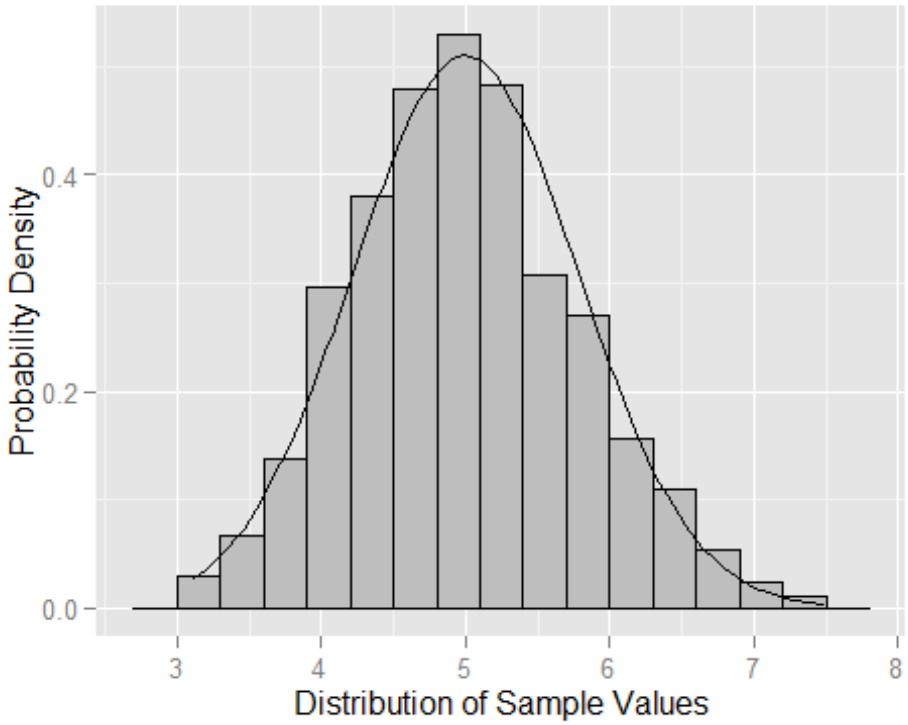
Therefore, the sample variance is very close to the theoretical variance (and so the standard deviations).

Question 3: Show that the distribution is approximately normal.

Below is a histogram plot of the sample means of the 1000 iterations of the exponentially distributed samples.

First plot the histogram of the sample data and overlay the theoretical normal distribution with a mean of 5. They do match closely.

```
library(ggplot2)
ggplot(data = sample_means, aes(x = expcol)) +
        geom_histogram(aes(y = ..density..), fill = I('grey'), binwidth =
0.30, color = I('black')) +
        xlab("Distribution of Sample Values") +
        ylab("Probability Density") +
        stat_function(fun = dnorm, arg = list(mean = 1/lambda, sd =
sd(sample_means$expcol)))
```



We create a Q-Q plot and it shows that the data is approximately normally distributed.

```
qqnorm(sample_means$expcol,col = "gray", main = "Normal Q-Q Plot",xlab =
"Quantiles",ylab = "Quantiles of Sample Data")
qqline(sample_means$expcol,col = "red")
```