# Statistical Inference Project Part 1

## Contents



### 0.1 Synopsis

This is the project for the statistical inference class. In it, you will use simulation to explore inference. This is the first part of the project which is a simulation exercise.

The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also also 1/lambda. Set lambda = 0.2 for all of the simulations. In this simulation, I will investigate the distribution of averages of 40 exponential(0.2)s. Note that I will do a thousand simulated averages of 40 exponentials to illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s.

### 0.2 Simulation

```
# setting lamda the rate
lambda = 0.2

# number of samples
nSamples = 40

# number of simulations
nSimulations = 2000

# simulating
expoMeans = NULL
for( i in 1:nSimulations) expoMeans = c(expoMeans, mean(rexp(40, rate = lambda)))
```

### 0.3 Questions

#### 0.3.1 Q1-Comparing the simulation distribution center to the theoretical center of the distribution.

```
expoDistributionMean = mean(expoMeans)
```

The simulation distribution mean is **5.0015346** which is pretty close to the theoritical distribution mean which is **5**

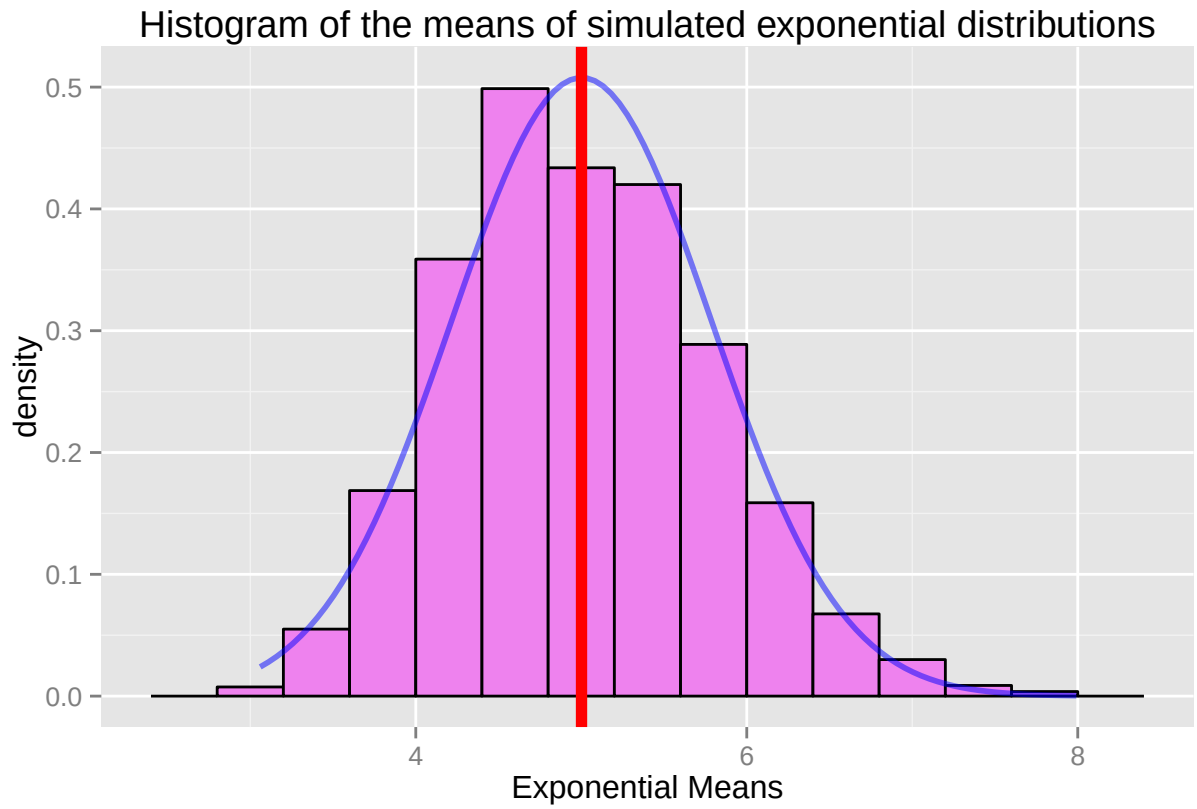
### 0.3.2 Q2-Comparing the simulation distribution variance to the theoretical variance of the distribution.

```
expoDistributionSD = sd(expoMeans)
```

The simulation distribution standard deviation is **0.7856508** which is pretty close to the theoritical distribution standard deviation which is **0.7905694**

### 0.3.3 Q3-Showing that the simulation distribution is approximately normal.

```
library(ggplot2)
expoMeansDataframe = data.frame(expoMeans)
ggplot(expoMeansDataframe, aes(x= expoMeansDataframe$expoMeans)) +
geom_histogram(aes(y=..density..), colour="black", fill="violet",  binwidth = 2 * lambda) +
stat_function(fun = dnorm, colour = "blue", alpha =0.5, size = 1,  fill = "yellow",
      args = list(mean = expoDistributionMean, sd = expoDistributionSD)) +
geom_vline(xintercept = expoDistributionMean,colour="red",size=2) +
xlab("Exponential Means") +
ggtitle("Histogram of the means of simulated exponential distributions")
```




It is obvious that the simulated distribution of exponentials (pink bins) can be approximiated as a normal distribution (blue curve). I have illustrated via simulation and associated explanatory text that the properties of the distribution of the mean of 40 exponential(0.2)s is as follows

- The simulated distribution is approximately centered at the theoretical center of the distribution.
- The simulated distribution is of a variance that is very close to the theoretical variance of the distribution.
- The simulated distribution is can be approximiated as a normal distribution.