Statistical Inference - Simulation

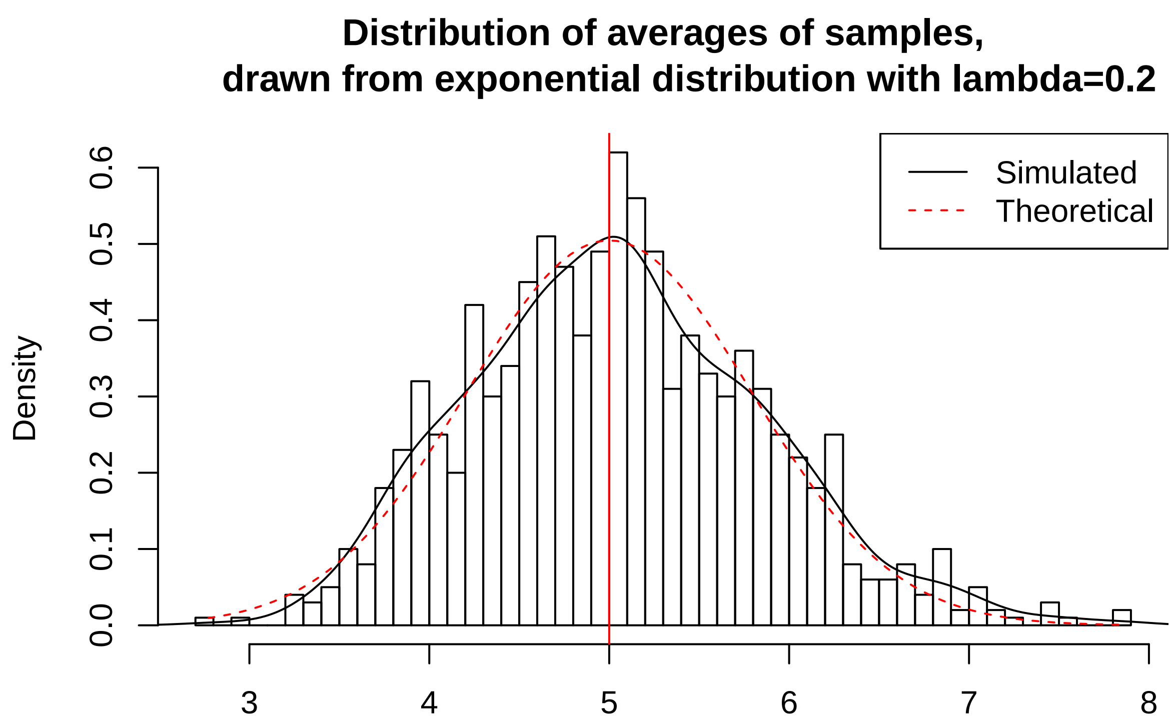
Gaurav Tripathi

Mean and Variance of Sample Mean for Exponential Distribution

To understand the distribution of mean of 40 randomly chosen samples from an exponential distribution, we shall simulate 1000 times (where the number 1000 has been randomly chosen, such that it is high enough to be able to give a proper distribution curve, and yet small enough to be computed on a personal computer). The code for simulating that is as follows:

```
set.seed(10)
lambda <- 0.2
sampleSize <- 40
simulations <- 1000
row_means <- rowMeans(matrix(data = rexp(simulations*sampleSize, rate=lambda),
                  nrow = simulations, ncol = sampleSize))
```

On plotting it on a graph, we get the following distribution:



The following code gives the comparison of theoretical vs. actual mean and variance of the graph

```
# Comparison of center of graph with theoretical center
paste("Graph is centered at: ", round(mean(row_means), 3))
paste("Graph should have been theoretically centered at: ", round(1/lambda, 3))
```

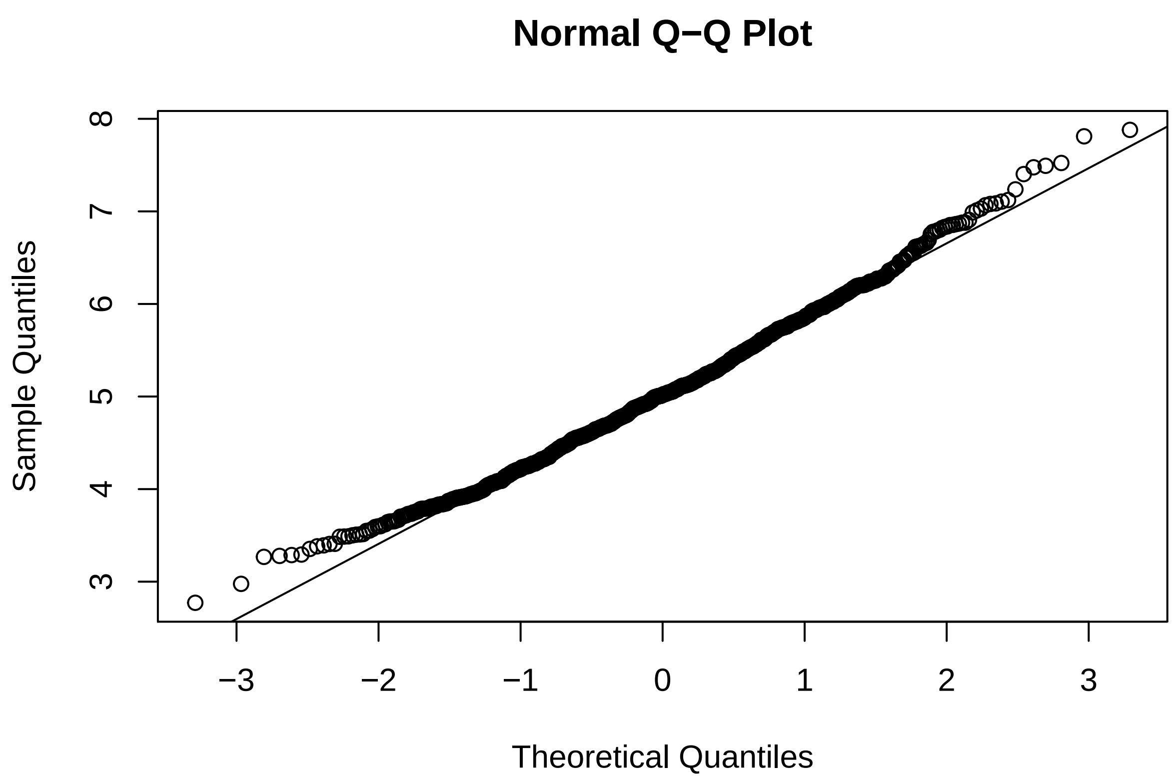
```r
# Comparison of variance of graph with theoretical variance
paste("Variance of Graph: ", round(var(row_means), 3))
paste("Theoretical variance: ", round(1/(lambda*lambda*sampleSize), 3))
```

- Graph is centered at: 5.045, while theoretically it should have been at: 5 (which is 1/lambda)
- Variance of Graph: 0.654, while theoretically it should have been: 0.625 (which is 1/(lambda x lambda x sampleSize))

Normality of Sample Mean for Exponential Distribution

The following code tests for normality of the sample distribution:

```r
# Use Q-Q plot to test for Normality
qqnorm(row_means)
qqline(row_means)
```



Since the points fall very close to the line due to Normal Distribution, we can say with some confidence that sample means follow normal distribution (which is the same as what Central Limit Theorem also states).

```r
# Use Shapiro Test to indicate that the curve just an approximation of normal distribution
shapiro.test(row_means)
```

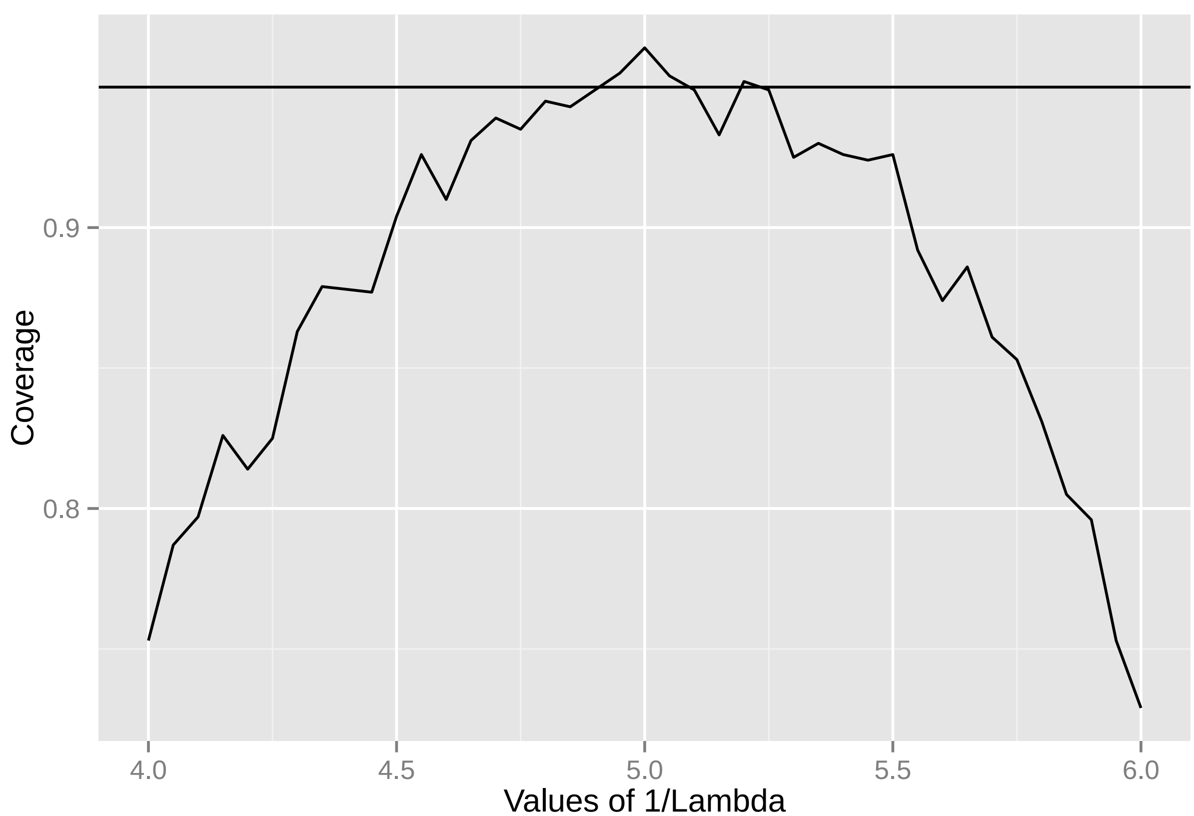
Shapiro-Wilk normality test data: row_means W = 0.993, p-value = 0.0001246

Since p-value < 0.05, NULL Hypothesis (which states that the data comes from a perfect Normal Distribution) is rejected

Coverage of confidence interval for 1/lambda

The following code evaluates the coverage of the confidence interval for 1/lambda:

```
library(ggplot2)
# Calculation of coverage of the confidence interval for 1/lambda
lambda_vals <- seq(4, 6, by=0.05)
coverage <- sapply(lambda_vals, function(lamb) {
    means <- rowMeans(matrix(data = rexp(sampleSize*simulations, rate=lambda),
                               nrow = simulations, ncol = sampleSize))
    ll <- means - qnorm(0.975) * sqrt(1/lambda**2/sampleSize)
    ul <- means + qnorm(0.975) * sqrt(1/lambda**2/sampleSize)
    mean(ll < lamb & ul > lamb)
})
qplot(lambda_vals, coverage, xlab = "Values of 1/Lambda"
      , ylab = "Coverage", geom = "line") + geom_hline(yintercept=0.95)
```



It shows that if 1/lambda is chosen around 5, the average of the sample mean falls within the confidence interval at least 95% of the time. Note that the theoretical value of 1/lambda is 5 given that lambda = 0.2.