Statistical Inference Course Project: Part 1

Author: Mehul Patel
Date: July 25, 2015

Assignment Description

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 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

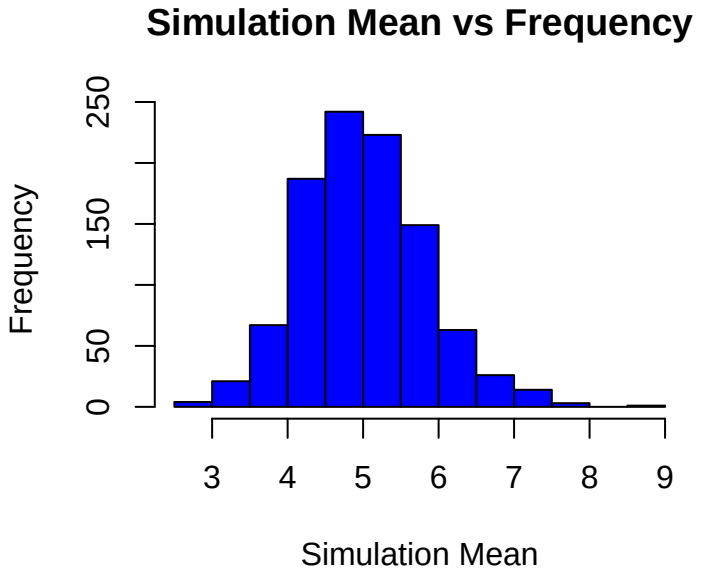
Simulation

The code below performs the simulation to collect/transform data.

```
library(ggplot2)

#variables for simulation
no_sim <- 1000
lambda <- 0.2
n <- 40
set.seed(234)

#Create a matrix of 1000 rows with the columns corresponding to random simulation 40 times
sim_matrix <- matrix(rexp(no_sim * n, rate=lambda), no_sim, n)
sim_mean <- rowMeans(sim_matrix)
hist(sim_mean, col = "blue", xlab = "Simulation Mean", main = "Simulation Mean vs Frequency")
```



Sample Mean Comparison

The actual mean for the sample data and mean through the equation is calculated as:

```
mean_data <- mean(sim_mean)
theo_mean <- 1/lambda
```

Actual center of the distribution based on the simulations is 5.001573 and the theoretical mean for lambda = 0.2 is 5. Hence the actual mean from sample data is very close to the theoretical mean of normal data.

Variance Comparison

Actual mean for the sample data and variance through theoritical equation is calculated as:

```
actual_var <- var(sim_mean); theo_var <- (1/lambda)^2/n
```

Actual variance for the sample data is 0.6631504 and theoretical variance is 0.625. These values are close to each other.

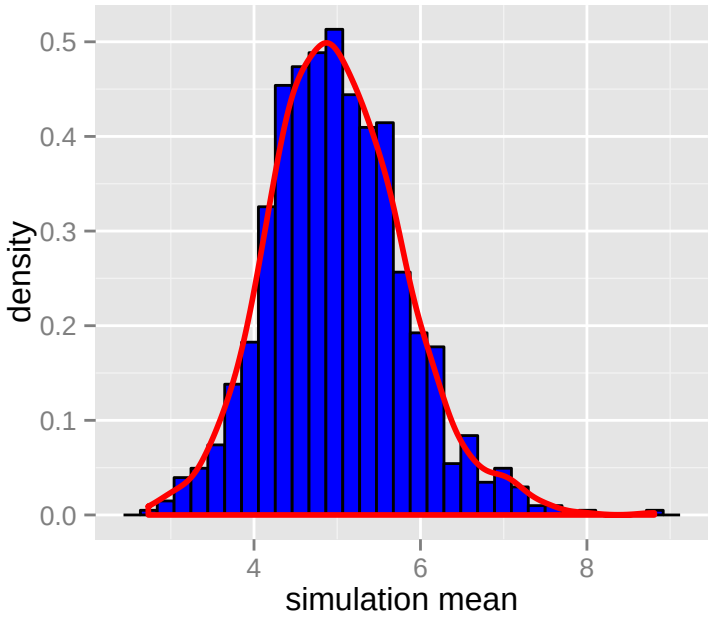
Approximate Normal Distribution

Follow the 3 steps as described below:

- Create an approximate normal distribution and check how the sample data aligns with it.
- Compare the confidence interval along with the mean and variance with normal distribution.
- q-q plot for quantiles.

```
plotdata <- data.frame(sim_mean);
m <- ggplot(plotdata, aes(x =sim_mean)) + xlab("simulation mean")
m <- m + geom_histogram(aes(y=..density..), colour="black", fill = "blue")
m + geom_density(colour="red", size=1)
```

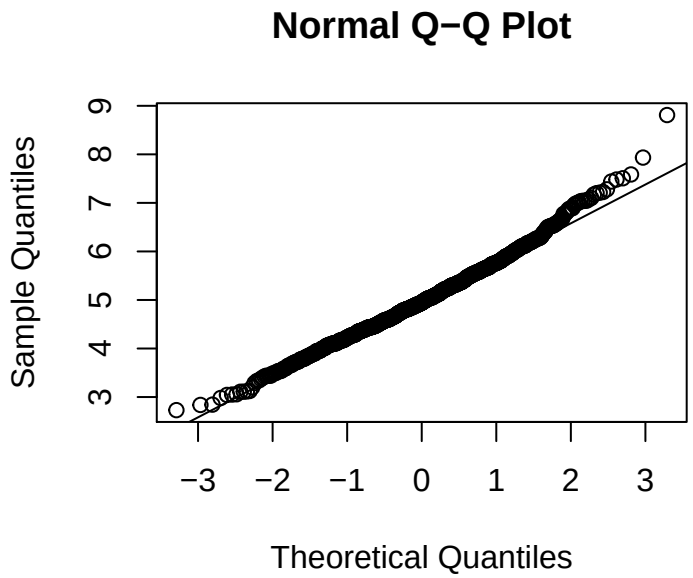
```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```



```
actual_conf_interval <- round (mean(sim_mean) + c(-1,1)*1.96*sd(sim_mean)/sqrt(n),3)
theo_conf_interval <- theo_mean + c(-1,1)*1.96*sqrt(theo_var)/sqrt(n)
```

Actual 95% confidence interval [4.749, 5.254]. Theoretical 95% confidence interval [4.755, 5.245].

```
qqnorm(sim_mean)
qqline(sim_mean)
```



The theoretical quantiles match closely with the actual quantiles. These three evidences prove that the distribution is approximately normal.