Statistical Inference Project Report Assignment Part1

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, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials. Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s.

Simulate Data

```
set.seed(357) ; lambda <- .2 ; n<-40 ; nsims = 1:1000
simdata <- data.frame(x = sapply(nsims, function(x) {mean(rexp(n, lambda))}))
simmean <- mean(simdata$x) ; simsd<- sd(simdata$x)
```

Report

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.

```
simmean # Simulated Mean
```

```
## [1] 5.031
```

```
1/.2    # Theoretical mean
```

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

As seen above the expected value of the sample mean is equal to the mean it is trying to estimate. The distribution of the sample mean is gausian, centered at 5 and concentrated very close to the center as shown below in a histogram plot.

2. Show how variable it is and compare it to the theoretical variance of the distribution.

```
(1/lambda)/sqrt(n) # Theoretical Standard deviation for n <- 40
```

```
## [1] 0.7906
```

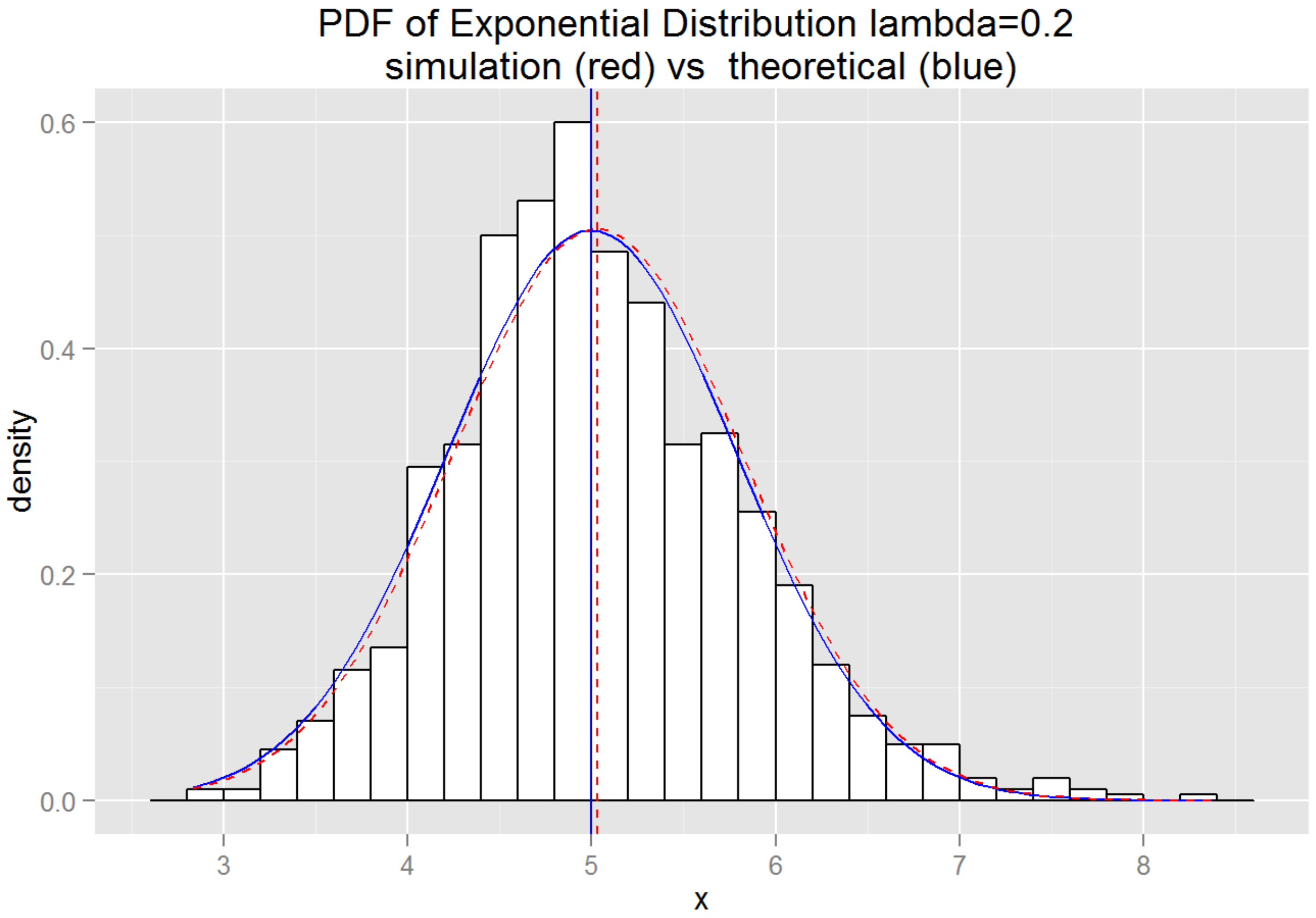
```
simsd # Simulated Standard deviation
```

```
## [1] 0.7888
```

The standard deviation of simulated data is 0.79. This corresponds with the standard error of the mean(i.e. SE=sigma/sqrt(n) =(1/.2)/sqrt(40) which is equal to 0.791 for the 40 observations.

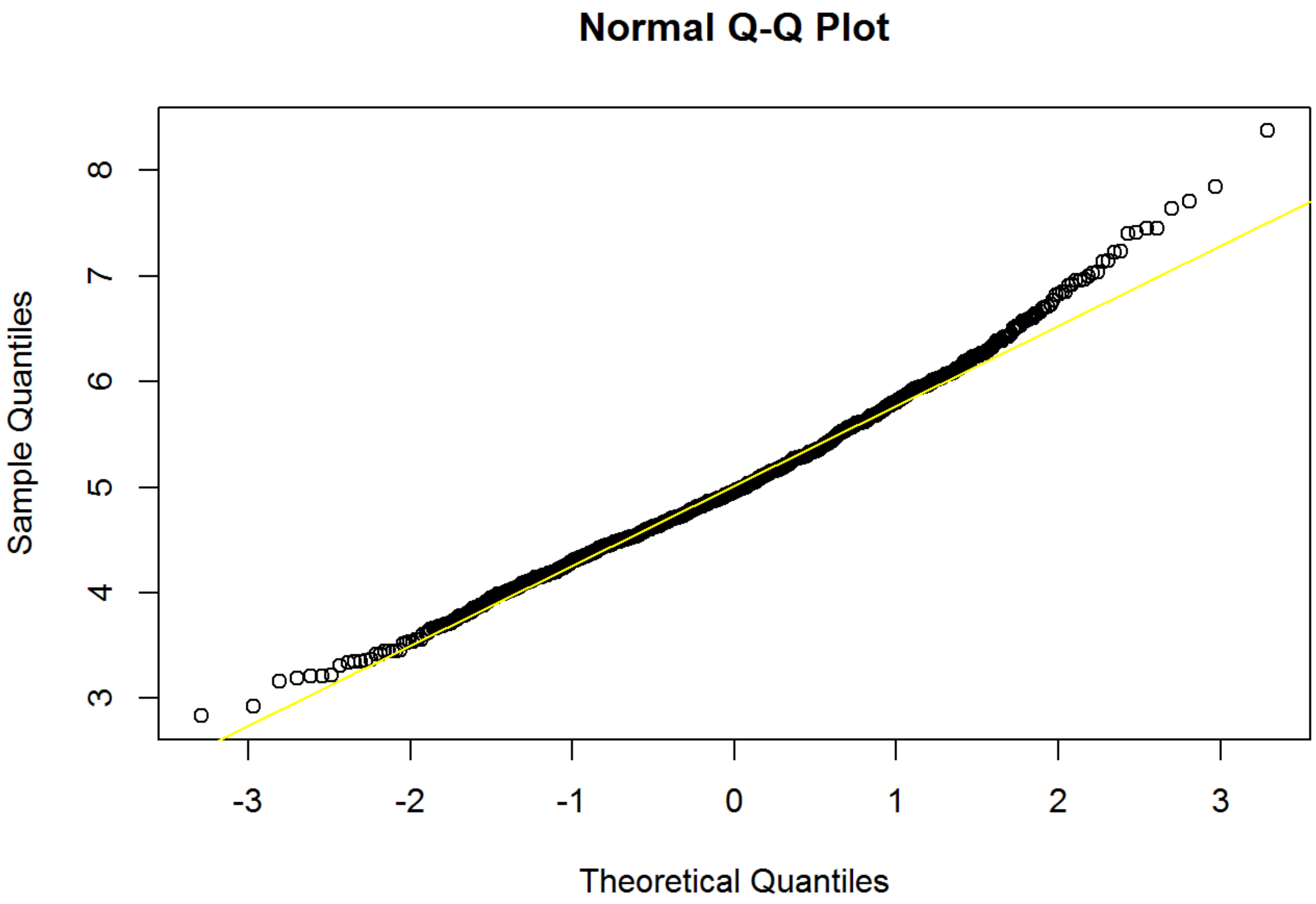
3. Show that the distribution is approximately normal.

```
library(ggplot2)
ggplot(data = simdata, aes(x = x)) +
geom_histogram(aes(y=..density..), fill = 'white',
                 binwidth = 0.20, color = 'black') +
stat_function(fun = dnorm, color="blue", pct=20,arg = list(mean = 5, sd = (1/lambda/sqrt(n)))) +
stat_function(fun = dnorm, color="red",pct=20,lty=2, arg = list(mean = simmean, sd = simsd)) +
geom_vline(xintercept = 5,color='blue') +
geom_vline(xintercept = simmean,color='red',lty=2) +
scale_x_continuous(breaks = seq(from = 0, to = 12, by = 1))+
ggtitle(label="PDF of Exponential Distribution lambda=0.2 \n simulation (red) vs  theoretical (blue)
")
```



Due to the central limit theorem, the averages of samples follow normal distribution,as seen above in the histogram plot that shows the density computed using the histogram and the normal density plotted with theoretical mean and sd values. It is overlaid with a normal distribution with mean 5 and standard deviation 0.791, and Yes, the distribution of the simulations appears normal.Also, the q-q plot below suggests the normality.

```
qqnorm(simdata$x)
qqline(simdata,col=7)
```



4. Evaluate the coverage of the confidence interval for 1/lambda: X+-1.96S/sqrt(n).

(This only needs to be done for the specific value of lambda).

```
library(data.table)
#Calculate the confidence interval
ci <- simmean + c(-1,1)*1.96*simsd
ci
```

```
## [1] 3.485 6.577
```

```
#Calculate the coverage
sum(between(simdata, ci[1], ci[2])) / length(simdata)
```

```
## [1] 944
```

Since, we consider the distribution of averages of exponentials, the standard deviation of this distribution already incorporates the (sqrt(n)) term i.e. it is the standard error. The confidence interval is given by [3.485 6.577].

The Confidence interval coverage 94.4% is given by [3.485 6.577]