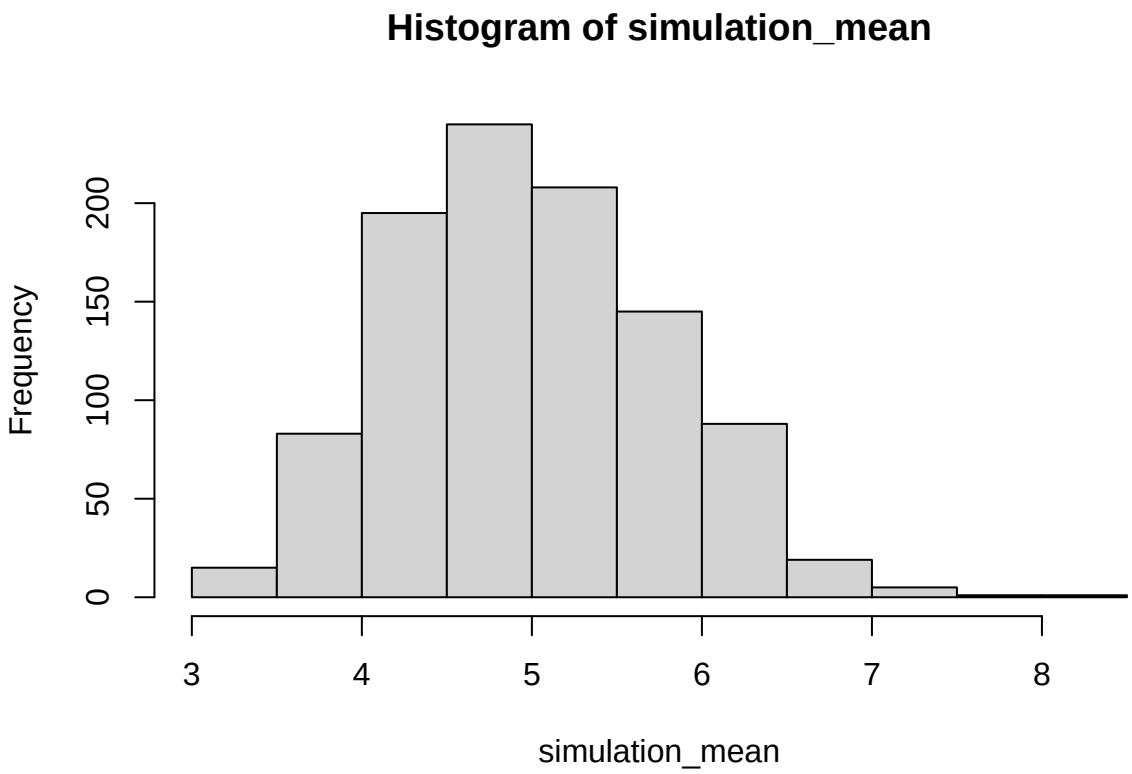
Statistical-Inference-Course-Project Part 1

Yigang

Part 1: Simulation

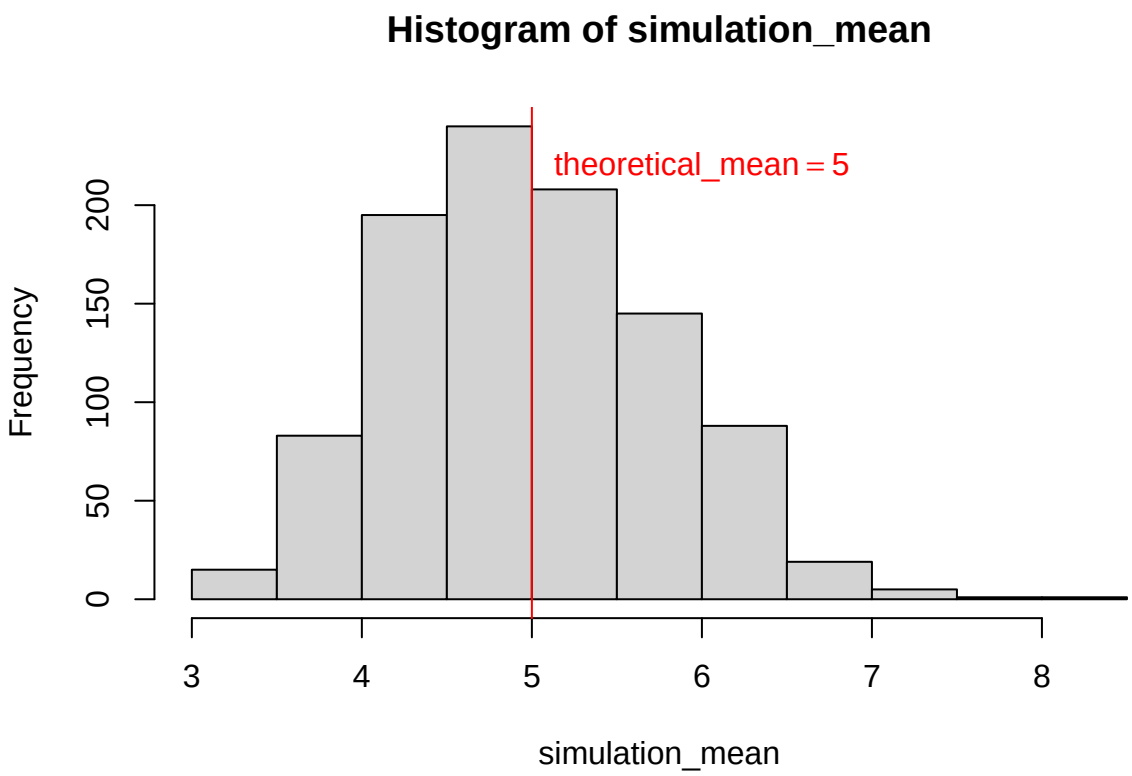
Run 1000 simulations for exponential distribution, where each simulations has lambda = 0.2 and n = 40

```
set.seed(1)

lambda <- 0.2
n <- 40
times <- 1000

simulations <- matrix(rexp(n*times, rate = lambda), times)
simulation_mean <- apply(simulations, 1, mean)

hist(simulation_mean)
```



Compare mean from simulation with theoretical mean

```
theoretical_mean <- 1/lambda

hist(simulation_mean)
abline(v = theoretical_mean, col = "red")
text(6, 220, expression("theoretical_mean"  == 5), col = "red")
```



Compare variation from simulation with theoretical variation

```
theoretical_var <- (1/lambda)^2/n;
theoretical_sd <- 1/lambda/sqrt(n);

simulation_sd <- sd(simulation_mean)

print(paste("Theoretical variance = ", theoretical_var))
```

```
## [1] "Theoretical variance =  0.625"
```

```
print (paste("Simulation Variance = ",round(var(simulation_mean), 3)))
```

```
## [1] "Simulation Variance =  0.618"
```

```
print (paste("Theoretical standard deviation = ", round(theoretical_sd, 3)))
```

```
## [1] "Theoretical standard deviation =  0.791"
```

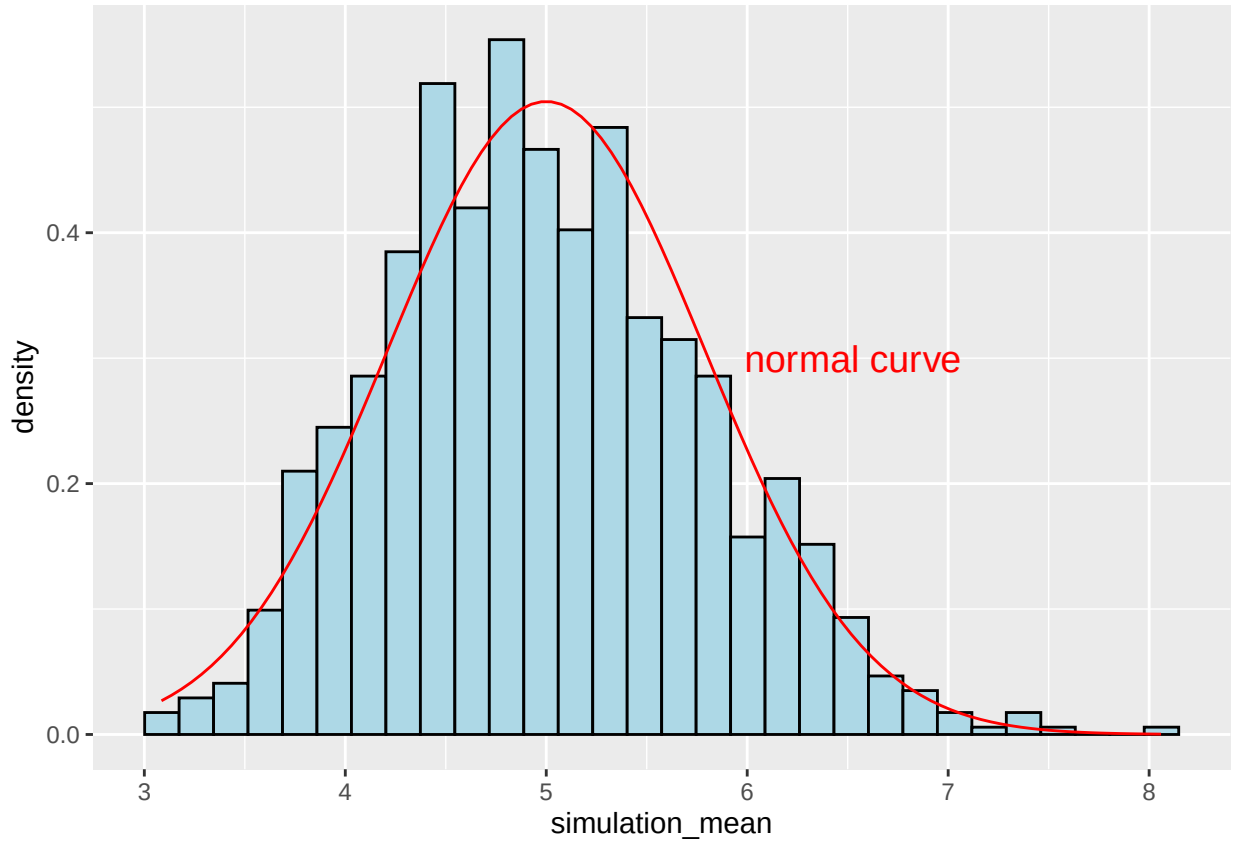
```
print (paste("Simulation standard deviation = ",round(simulation_sd, 3)))
```

```
## [1] "Simulation standard deviation =  0.786"
```

Show the distribution is approximately normal

```
library(ggplot2)

df <- data.frame(simulation_mean)

ggplot(df, aes(x = simulation_mean))+
geom_histogram(aes(y=..density..), colour="black", fill = "lightblue")+
stat_function(fun = dnorm, args = list(mean = theoretical_mean, sd = theoretical_sd),
              colour = "red")+
annotate("text", label = " normal curve", x = 6.5, y = 0.3,
         size = 5, colour = "red")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Compare Confidence Interval from simulation with theoretical Confidence Interval

```
simulation_CI <- round (mean(simulation_mean) + c(-1,1)*1.96*simulation_sd/sqrt(times),3)
theoretical_CI <- theoretical_mean + c(-1,1) * 1.96 * sqrt(theoretical_var)/sqrt(n)
```

Conclusion

The simulations provide an evidence that support the Central Limit Theorem and the distribution of sample means are approximately normal.