# Week4 Project Part1

Melody Zhang

2022-06-27

## Part 1: Simulation Exercise Instructions

Preparation for analysis.

```
library(stats)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

Use set.seed() to reproduce results. Set lambda = 0.2, exponentials = 40 and simulations = 1000.

```
set.seed(624)
lambda = 0.2
n = 40
nosim = 1000
```

### 1. Show the sample mean and compare it to the theoretical mean of the distribution.

```
data1 <- replicate(nosim,rexp(n, lambda))
exp_mean <- apply(data1,2,mean)

samp_mean <- mean(exp_mean)
samp_mean
```

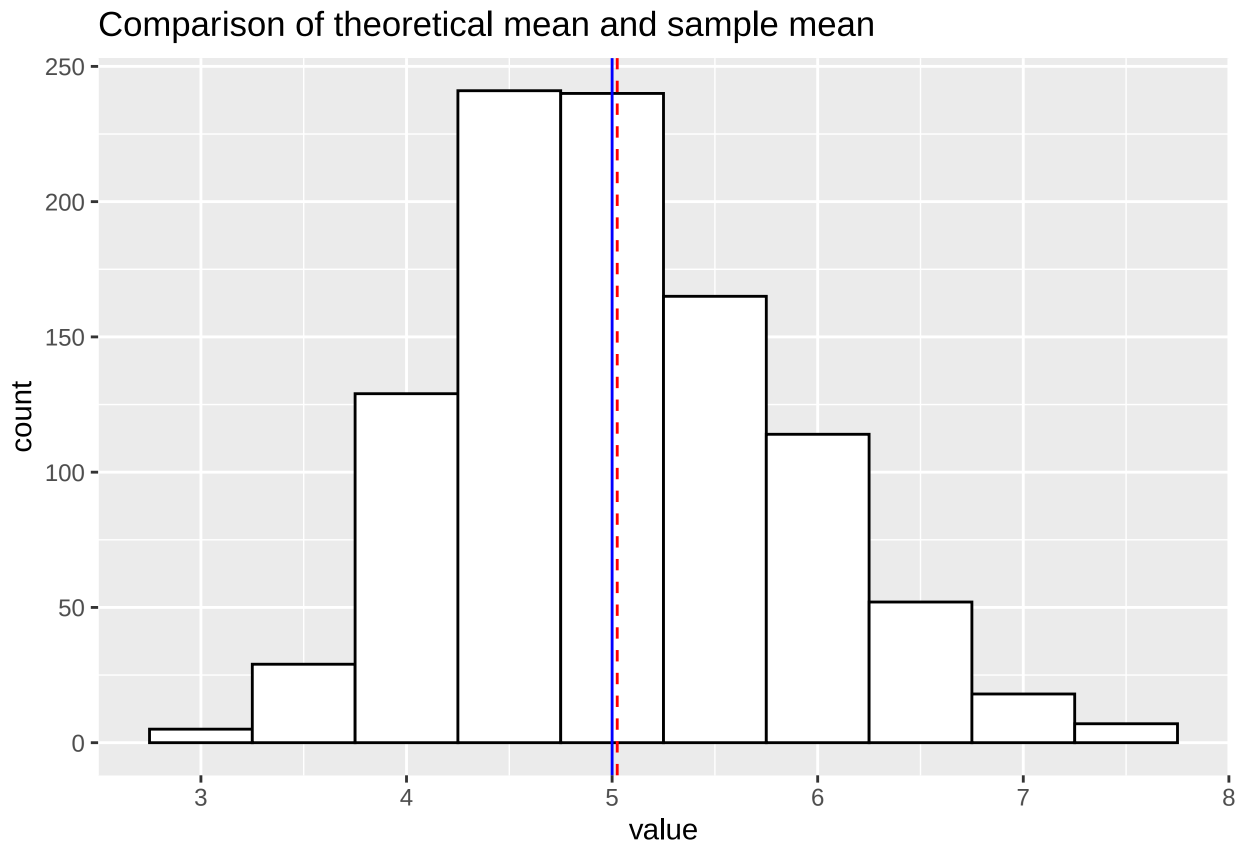
```
## [1] 5.024981
```

```
theo_mean <- 1/lambda
theo_mean
```

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

```
data2 <- data.frame(Mean_of_exp='exp_mean',value=exp_mean)
ggplot(data2,aes(x=value)) +
  geom_histogram(binwidth = 0.5,colour = "black",fill = "white") +
  geom_vline(aes(xintercept=samp_mean), color="red", linetype="dashed", size=0.5)+
```

```
  geom_vline(aes(xintercept=theo_mean), color="blue", size=0.5)+
  ggtitle("Comparison of theoretical mean and sample mean")
```




According to our results, the sample mean is about 5.02(red dashed line) and the theoretical mean is 5(blue line).

**2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.**

```
samp_sd <- sd(exp_mean)
samp_sd
```

```
## [1] 0.8007116
```

```
samp_var <- samp_sd^2
samp_var
```

```
## [1] 0.6411391
```

```
theo_sd <- 1/(lambda*sqrt(n))
theo_sd
```

```
## [1] 0.7905694
```

```
theo_var <- theo_sd^2
theo_var
```

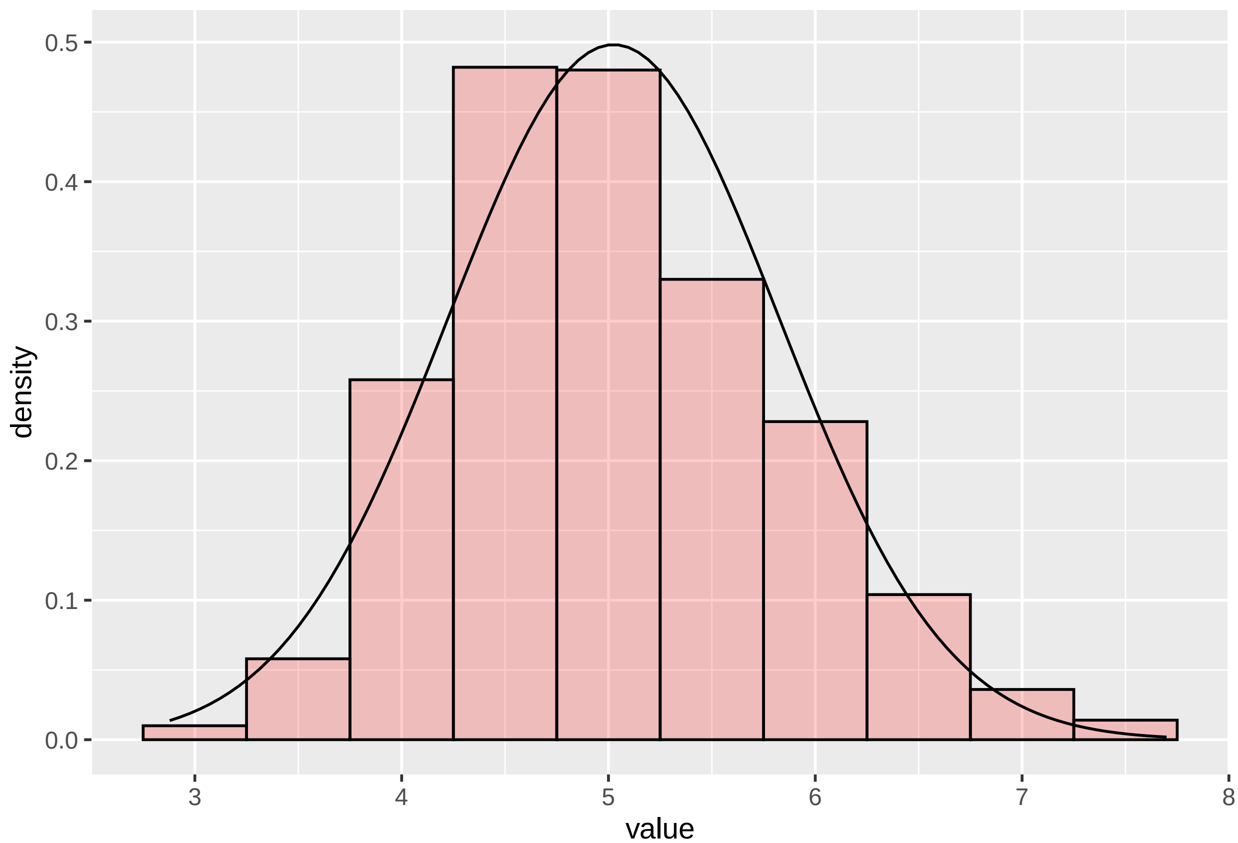
```
## [1] 0.625
```

The standard deviation of exponential is (1/lambda)/sqrt(n), and variation is calculated by ((1/lambda)/sqrt(n))ˆ2. The sample variance is about 0.64 which is really approximate to the theoretical variance is around 0.625.

**3. Show that the distribution is approximately normal.**

```
x <- seq(min(exp_mean), max(exp_mean), length=100)
y <- dnorm(x, mean=1/lambda, sd=(1/lambda/sqrt(n)))

ggplot(data2, aes(x=value)) +
  geom_histogram(aes(y=..density..),
                 binwidth=.5,
                 colour="black", fill="red",alpha=0.2) +
  stat_function(fun = dnorm, args = list(mean = mean(data2$value), sd = sd(data2$value)))
```




Due to the Central Limit Theorem, we roughly get a normal distribution. The curve would even be more approximate to normal distribution as the sample number increase.