Peer-graded Assignment: Statistical Inference Course Project

Part 1:Simulation exercise.

In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. 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. We will investigate the distribution of averages of 40 exponentials.

Comparison of sample mean with the theoretical mean.

```
library(ggplot2)
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
```

```
mn<-vector("numeric",length = 1000)
for(i in 1:1000){
  mn[i]<-mean(rexp(40,0.2))
}
th_mean=1/0.2
sample_mean=mean(mn)
```

Hence the theoretical mean 5 and the sample mean 4.9736847 are close,which proves the point of central limit theory.

comparison of sample variance and theoretical variance.

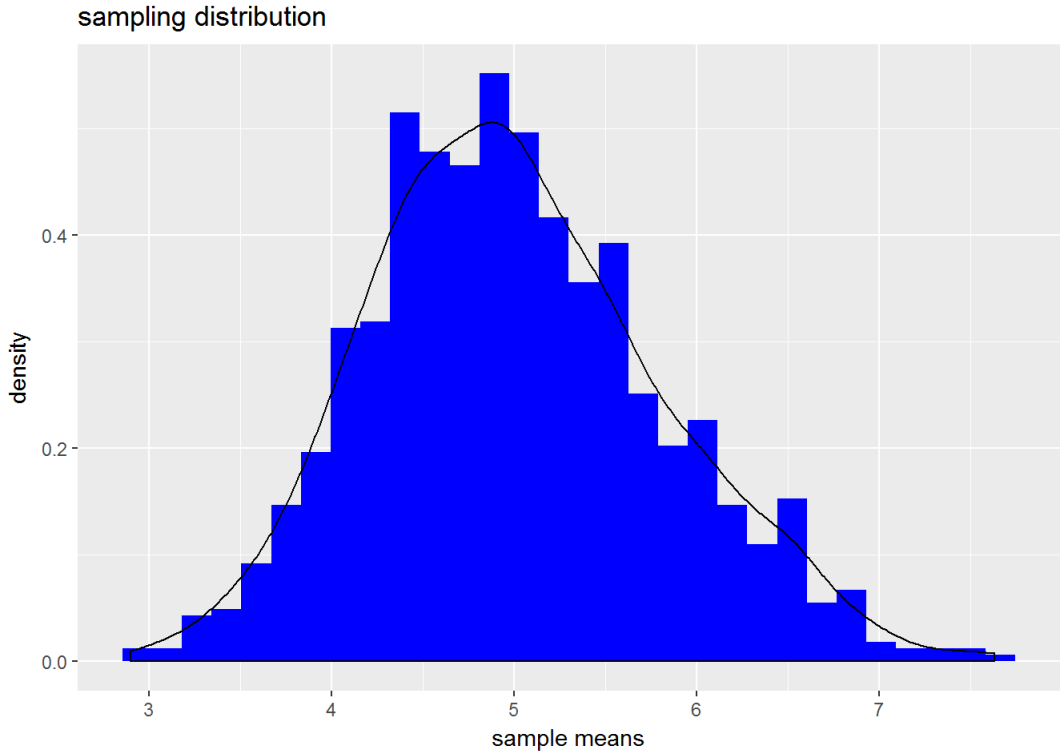
```
th_var<-((1/0.2)^2)/40
sample_var<-var(mn)
```

The observed variance of the sampling distribution is 0.6071421 and the theoretical variance of sampling distribution is 0.625.This simulates the central limit theory.

Showing the sampling distribution to be approximately normal.

```
g<-ggplot(data.frame(mn=mn),aes(mn))+geom_histogram(fill="blue",aes(y=..density..))
g<-g+geom_density(col="black")+labs(x="sample means")+labs(title="sampling distribution")
g
```

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



The sampling distribution resembles a normal distribution.