Title：A simulation exercise

Author：eyangs

Overview：

Exponential distribution is simulated in R（lambda = 0.2）， and then the distribution of averages of 40 exponentials is investigated with comparison of the Central Limit Theorem

Simulations：

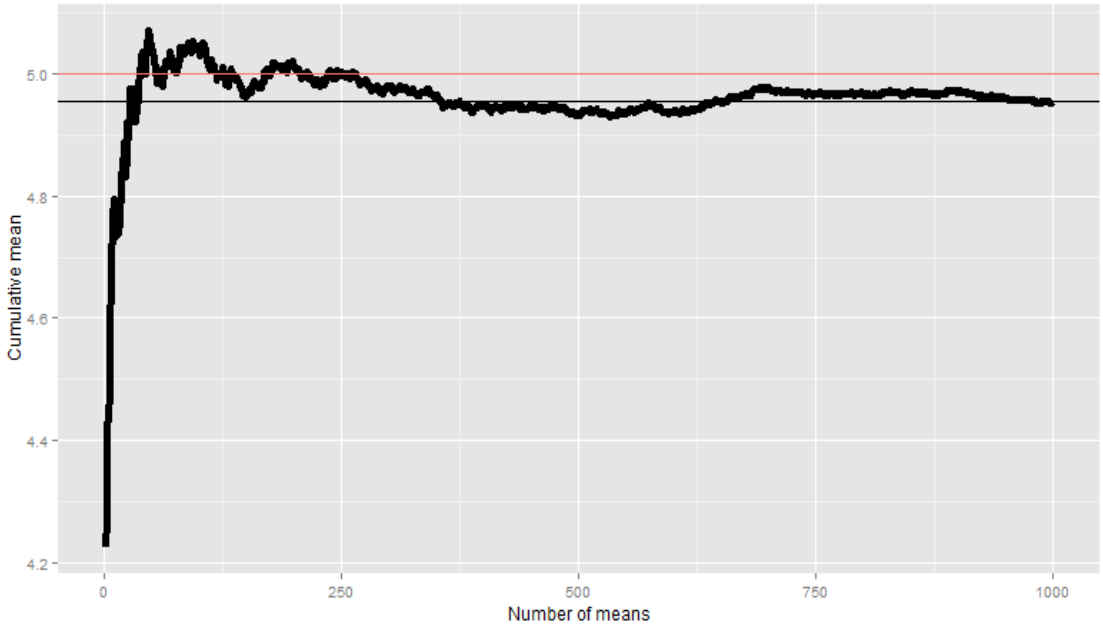
```
#using rexp function to simulate（n=40， lambda=0.2）
#using a function getmeans to obtain averages of 40 exponentials
#calculate cumsum of averages divide (1:n)
#ggplot draw figure to show the asymptopia
library(ggplot2)
n<-1000
getmeans<-function(n){
   mns = 0
   for(i in 1:n) mns=c(mns, mean(rexp(40,.2)))
   mns
}
getvars<-function(n, means){
   vars = NULL
   for(i in 1:n) vars=c(vars, var(means[1:i]))
   vars
}
temp<- getmeans(n)
means<- cumsum(temp)/(1:n)
vars<- getvars(n,temp)

g1<-ggplot(data.frame(X=1:1000, Y=means), aes(x=X, y=Y))
g1<- g1 + geom_line(size=2) + labs(x="Number of means", y="Cumulative mean")
g1<- g1 + geom_hline(aes(yintercept=mean(temp))) + geom_hline(aes(yintercept=1/0.2,colour =
"red"))
g1

g2<-ggplot(data.frame(X=1:1000, Y=vars), aes(x=X, y=Y))
g2<- g2 + geom_line(size=2) + labs(x="Number of vars", y="Cumulative vars")
g2<-              g2              +              geom_hline(aes(yintercept=var(temp)))              +
geom_hline(aes(yintercept=((1/0.2)^2)/40,colour = "red"))
g2
```

Sample Mean versus Theoretical Mean：

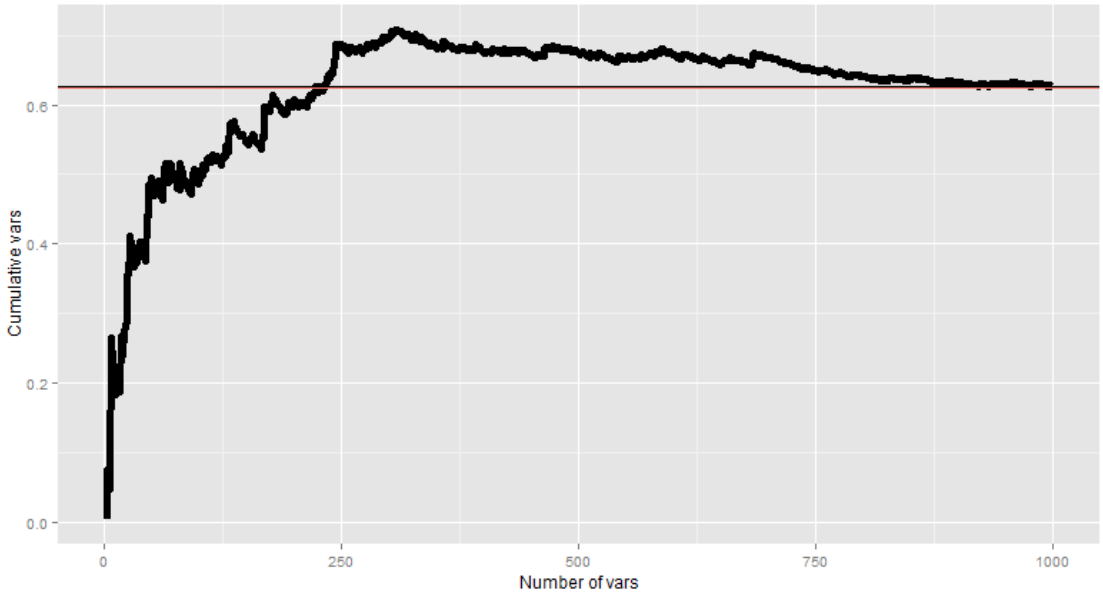
Sample mean：4.953934　　theoretical mean: 1/lambda=5



When n>500, the sample mean become close to 4.953934, the theoretical mean is 5 which is highlighted in red.

Sample Variance versus Theoretical Variance：

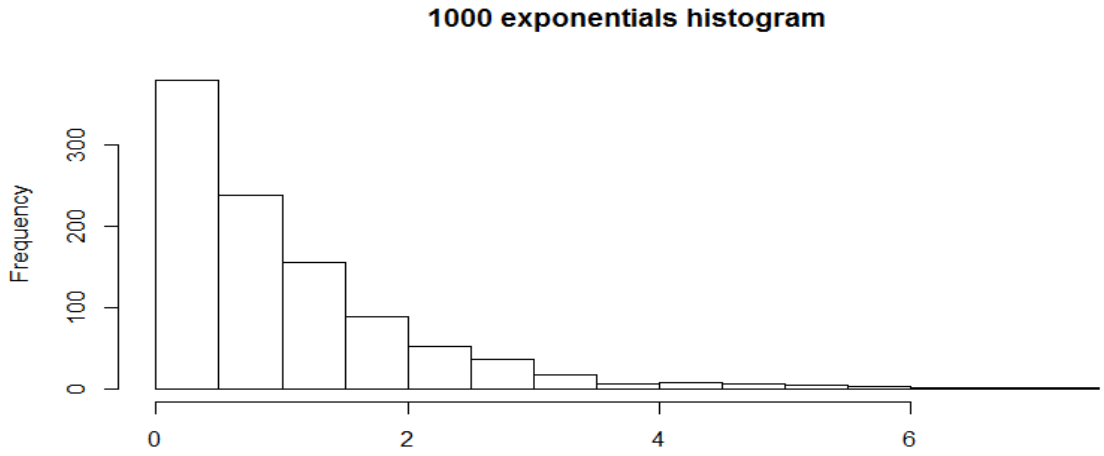
Sample variance: 0.6259166　　theoretical variance: ((1/lambda)^2)/40=25/40=0.625

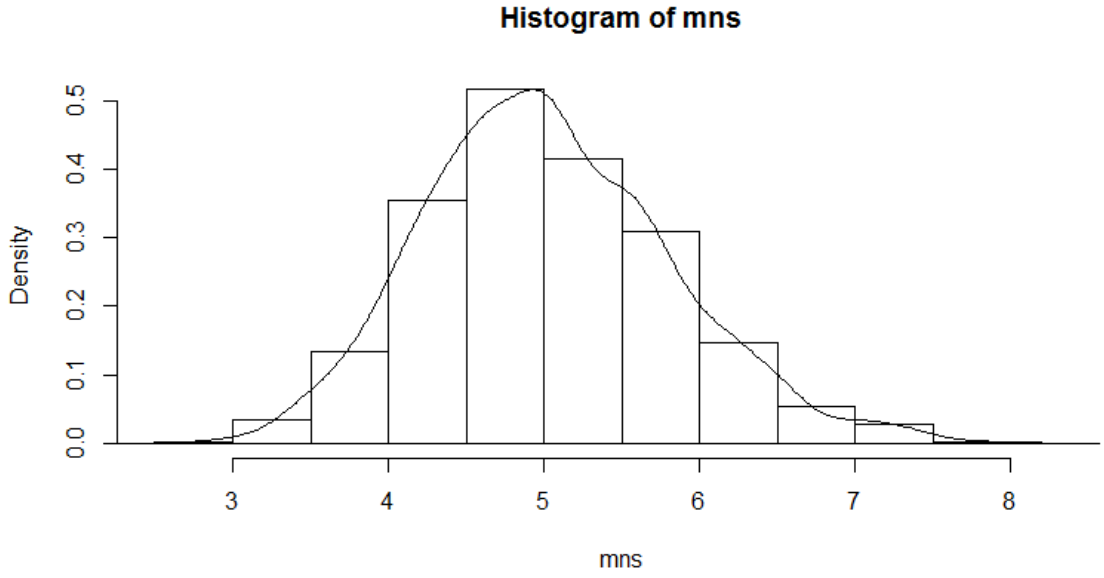


When n>750, the sample variance become close to 0.6259166, the theoretical variance is 0.625 which is highlighted in red. These two lines are Overlapping

Distribution：

#draw histogram with parameter of rexp (1000)
#using for-loop to obtain 1000 averages of 40 exponentials
#draw histogram of the 1000 averages
hist(rexp(1000), main="1000 exponentials histogram",xlab="")
mns = NULL
for (i in 1 : 1000) mns = c(mns, mean(rexp(40,0.2)))
hist(mns, freq=F, main="40 averages of exponentials")
lines(density(mns))





As the figure above shown, This distribution of averages of 40 exponentials looks far more Gaussian than the original exponentials distribution!