Statistical Inference Project1

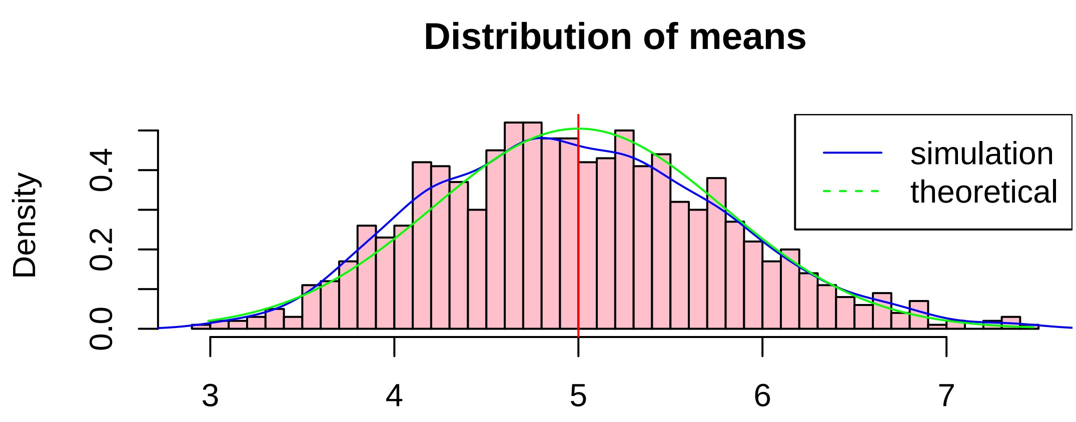
Yujing Cao

Sunday, September 21, 2014

```
#simulate data
#make a 1000 by 40 empty matrix, expmat
set.seed(3)
expmat=matrix(0,1000,40)
#Use for loop to sillumate data from exponential distribution with rate is 0.2
for(i in 1:40){
  expmat[,i]=rexp(1000,0.2)
  expmat
}
lambda=0.2
#calculte the averages of each column
#store averages of 40 exponential distribution in a vector,expmean
expmean=rowMeans(expmat)
```

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

```
#Histogram of the averages
hist(expmean,breaks=50,prob=T,main="Distribution of means",xlab="",col="pink")
#Density of the averages
lines(density(expmean),col="blue")
#The theoretical mean of the averages
abline(v=1/lambda,col="red")
theo.x=seq(min(expmean),max(expmean),length=1000)
theo.y=dnorm(theo.x,mean=1/lambda,sd=1/(lambda*sqrt(40)))
lines(theo.x,theo.y,pch=30,col="green")
legend("topright", c("simulation", "theoretical"), lty=c(1,2), col=c("blue", "green"))
```



Theoretical density function,by the center limit theorem, the density of averages approaches normal distribution

PART 2 Show how variable it is and compare it to the theoretical variance of the distribution. The theoretical variance is 1/lambda^2/n, and the sample variance we got is var(expmean). And the theoretical mean is 1/lambda, and the sample mean we got from simulation is rowMeans(expmean).

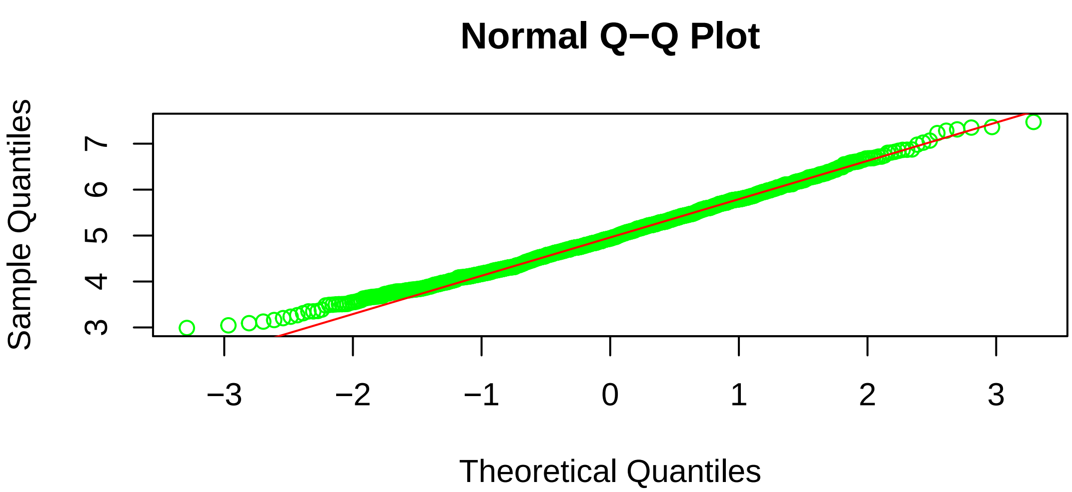
PART3 Show that the distribution is approximately normal Normality Test with Shapiro-Wilk Normality test

```
shapiro.test(expmean)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  expmean
## W = 0.9943, p-value = 0.0007348
```

p-value is 0.8902, which is much smaller than 0.05, thus we wouldn't reject null hypothesis of the ditribution is approximately normal

```
qqnorm(expmean,col="green")
qqline(expmean,col="red")
```



we could find most of variables fit the qqline very well, so the normal QQ plot tells us these variables are appoximately normal,which is consistent with the conclusion we got from shapiro.test

PART4 Evaluate the coverage of the confidence interval for 1/lambda

```
n=40
sd=apply(expmat,1,sd)
lower.bound=expmean-1.96*sd/sqrt(n)
upper.bound=expmean+1.96*sd/sqrt(n)
mean((1/lambda)>lower.bound &(1/lambda)<upper.bound)
```

```
## [1] 0.924
```

The 95% confidenc interval is [lower.bound,upper.bound]. And the coverage of confidence interval for 1/lambda is 92.4%, that is, 92.4% of confidence intervals can cover the sample mean.