Statistical_inference

Author : *Fred Zhou*

Overview

In this exercise, we will try to simulate data followed exponential distribution to test the CLT.

We need to prove:

For `{X1, ..., Xn}` be a random sample of size `n` from distributions of expected values given by `miu` and finite variances given by `sigma^2`:

sqrt(`n`)*(mean(`Xn`) - `miu`)) approximates the normal distribution with mean `0` and variance `sigma^2`.

Simulations

```
## Warning: package 'lattice' was built under R version 3.1.3
```

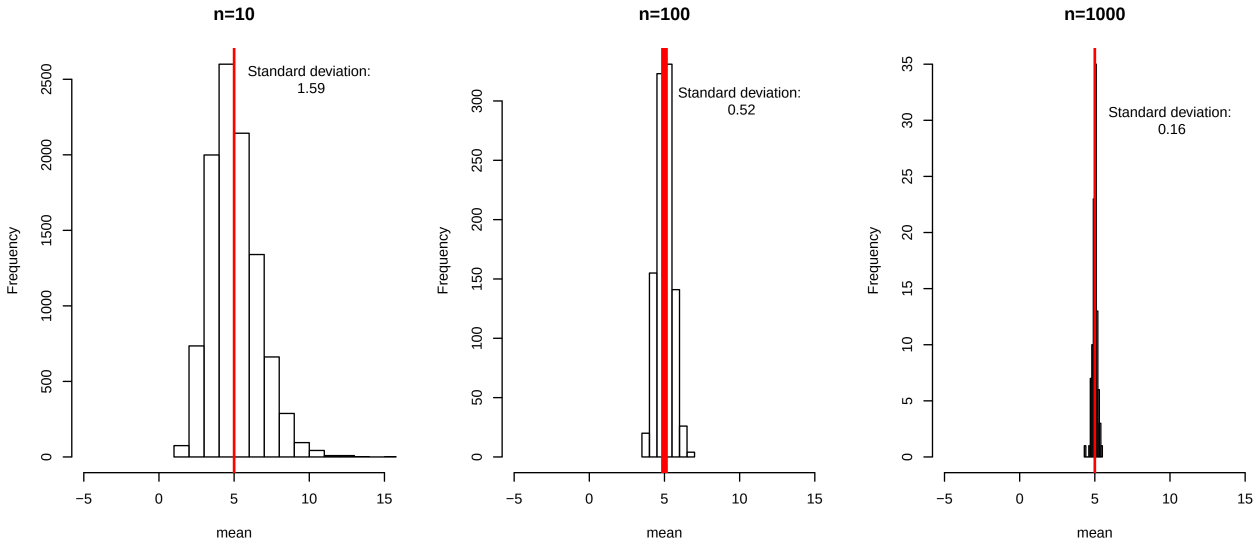
Sample Mean versus Theoretical Mean

We proceed the matrix rowwise to get the mean and sd with n = 10, 100, 1000 to check the sample mean distribution.

```
par(mfrow=c(1,3))
hist(apply(random_data_10,1,mean),xlim=c(-5,15),main='n=10',xlab='mean')
abline(v = 5, col = "red",lwd = 2)
text(10,2500,labels=paste('Standard deviation:\n',round(sd(apply(random_data_10,1,mean)),digits = 2)))

hist(apply(random_data_100,1,mean),xlim=c(-5,15),main='n=100',xlab='mean')
abline(v = 5, col = "red",lwd = 5)
text(10,300,labels=paste('Standard deviation:\n',round(sd(apply(random_data_100,1,mean)),digits = 2)))

hist(apply(random_data_1000,1,mean),xlim=c(-5,15),main='n=1000',xlab='mean')
abline(v = 5, col = "red",lwd = 2)
text(10,30,labels=paste('Standard deviation:\n',round(sd(apply(random_data_1000,1,mean)),digits = 2)))
```



The Sample Mean is approximating to the Theoretical Mean as the `n` increasing which is indicated by the decreasing of the standard deviation.

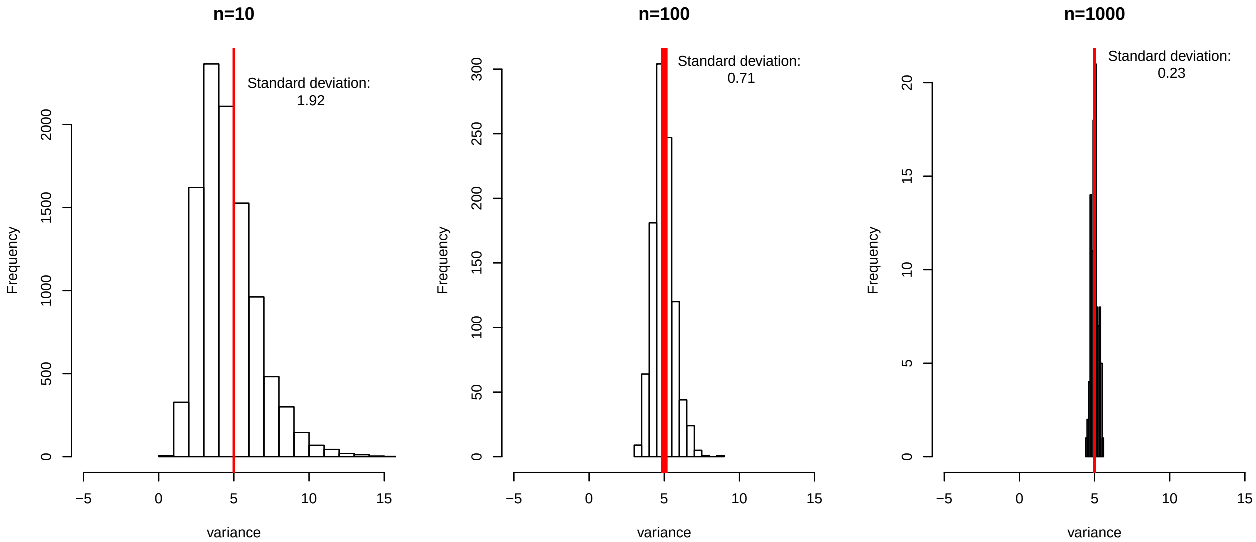
Sample Variance versus Theoretical Variance

We proceed the matrix rowwise to get the mean and sd with n = 10, 100, 1000 to check the sample variance distribution.

```
par(mfrow=c(1,3))
hist(apply(random_data_10,1,sd),xlim=c(-5,15),main='n=10',xlab='variance')
abline(v = 5, col = "red",lwd = 2)
text(10,2200,labels=paste('Standard deviation:\n',round(sd(apply(random_data_10,1,sd)),digits = 2)))

hist(apply(random_data_100,1,sd),xlim=c(-5,15),main='n=100',xlab='variance')
abline(v = 5, col = "red",lwd = 5)
text(10,300,labels=paste('Standard deviation:\n',round(sd(apply(random_data_100,1,sd)),digits = 2)))

hist(apply(random_data_1000,1,sd),xlim=c(-5,15),main='n=1000',xlab='variance')
abline(v = 5, col = "red",lwd = 2)
text(10,21,labels=paste('Standard deviation:\n',round(sd(apply(random_data_1000,1,sd)),digits = 2)))
```



The Sample Variance is approximating to the Theoretical Variance as the `n` increasing which is indicated by the decreasing of the standard deviation.

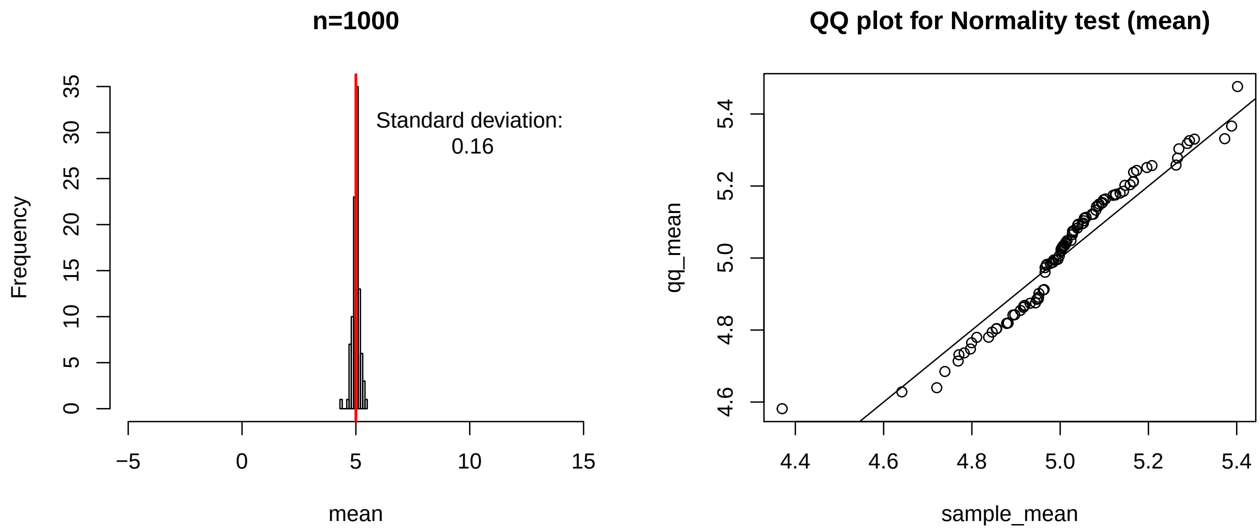
Distribution

For the previouse part, we overlap the normal distribution with sample mean and sample variance as parameter to draw the Q-Q plot.

```
#For Normality test of mean
par(mfrow=c(1,2))
sample_mean=apply(random_data_1000,1,mean)
mean_of_sample_mean=mean(sample_mean)
sd_of_sample_mean=sd(sample_mean)
qq_mean=rnorm(length(sample_mean),mean=mean_of_sample_mean,sd=sd_of_sample_mean)

hist(sample_mean,xlim=c(-5,15),main='n=1000',xlab='mean')
abline(v = 5, col = "red",lwd = 2)
text(10,30,labels=paste('Standard deviation:\n',round(sd(sample_mean),digits = 2)))

qqplot(sample_mean,qq_mean,main='QQ plot for Normality test (mean)')
abline(0,1)
```




```
#For Normality test of standard deviation
par(mfrow=c(1,2))
sample_sd=apply(random_data_1000,1,sd)
mean_of_sample_sd=mean(sample_sd)
sd_of_sample_sd=sd(sample_sd)
qq_sd=rnorm(length(sample_sd),mean=mean_of_sample_sd,sd=sd_of_sample_sd)

hist(apply(random_data_1000,1,sd),xlim=c(-5,15),main='n=1000',xlab='variance')
abline(v = 5, col = "red",lwd = 2)
text(10,21,labels=paste('Standard deviation:\n',round(sd(apply(random_data_1000,1,sd)),digits = 2)))

qqplot(sample_sd,qq_sd,main='QQ plot for Normality test (sd)')
abline(0,1)
```