Assignment 4

Munan Hou

November 4, 2016

1 1. CLT

I've searched the google and found 10+ tests. The tests that I familiar with are KolmogorovâĂŞSmirnov test, Pearson's chi-squared test, test of kurtosis and skewnessand, and shapiro-wilk test and qq plot.

So I am going to dig this on my own way. This is a little bit mess.

```
> library(ggplot2)
> library(nortest)
> library(normtest)
```

First, sample and summary that on average, for sums of how many uniform distribution could follow a normal distribution, on the significance level p=.05

```
> data1 <- c()
> number <- c()
> for (k in 1:100) {
+   for (i in 1:50) {
+     data1 <- c()
+     for (j in 1:1000) {
+       index <- runif(i)
+       sumindex <- sum(index)
+       data1[j] <- sumindex
+     }
+     if (shapiro.test(data1)["p.value"] > .05) {
+       number <- c(number,i);break
+     }
+   }
+ }
> number
```

```
  [1] 4 5 5 5 4 4 5 4 4 5 4 3 6 5 6 4 4 4 4 5 4 5 4 4 6 3 4 4 5 4 4 5 4 5 4 4 4
 [38] 4 3 3 6 3 5 3 4 4 4 4 4 3 4 4 3 4 4 4 5 3 4 4 4 4 5 5 5 4 5 5 4 4 3 4 3 4
 [75] 5 3 6 4 4 3 5 4 4 3 3 4 4 3 3 3 5 3 3 3 4 6 4 3 4 3
```

```
> mean(number)
```

```
[1] 4.1
```

Result shows, on average, the mean of "the sum of minimum amount of uniform distributions" which can construct a normal distribution is approximately 4.

Then test Poisson distribution with lambda = 10.

```
> data1 <- c()
> number <- c()
> for (k in 1:100) {
+   for (i in 1:50) {
+     for (j in 1:1000) {
+       index <- rpois(i, lambda = 5)
+       sumindex <- sum(index)
+       data1[j] <- sumindex
+     }
+     if (shapiro.test(data1)["p.value"] > .05) {
+       number <- c(number,i);break
+     }
+   }
+ }
> number
```

```
  [1] 12 13 10 11 10 13 13 10 11 10 15 14 13 16 10 14 12 10 14 15 13 10 13 10 13
 [26] 15 16 16 16 14 10 16 13 14  8 20 13 10 17 14 13 14 13 10 13 11 13 12 12 16
 [51] 17 14 11 14 12  8 15 10 18 11 14 12 13 11 12 13 11  9 10 15 13  9 10 13  8
 [76] 14 13 12 11 17 14 14 10 15 11 15 10 10  9  9 12 10 14  9  9 19  7 15 11  9
```

```
> mean(number)
```

```
[1] 12.46
```

Approximately 11 - 13 Poisson distributions.

Compare with uniform, Poisson converges more slowly.

2 2. Delta Method

Sampling Poisson(5)

```
> x <- rpois(1000, 5)
> data2 <- c()
> hist(x)
> for (j in 1:1000) {
+   index <- rpois(1000, 5)
+   data2 <- c(data2, mean(rpois(1000, 5)))
+   }
> y <- 1/data2
> qqnorm(y);qqline(y)
> shapiro.test(data2)
```

```
        Shapiro-Wilk normality test

data:  data2
W = 0.9986, p-value = 0.6235
```

Test shows it follows normal dist.

3 3. Student's t-test

```
> x <- seq(-4, 4, length = 10000)
> normdist <- dnorm(x, mean = 0, sd = 1)
> tdist1df <- dt(x, df = 1)
> tdist10df <- dt(x, df = 10)
> tdist100df <- dt(x, df = 100)
> plot(x, normdist, type = "l")
> lines(x, tdist1df, col = 'red')
> lines(x, tdist10df, col = 'green')
> lines(x, tdist100df, col = 'blue')
```

Could easily see that, with higher d.f., a t-distribution are approaching the normal distribution.

4 4. Median, Mean Square Error, Bootstrap

```
> productsales <- read.delim("C:/YUKIHO/BU/MA 615/Assignments/6/productsales.dat")
> productsales <- as.numeric(as.matrix(productsales))
> median <- median(productsales)
> Tboot <- rep(0,10000)
> for (i in 1:10000){
+   rep = sample(productsales, length(productsales), replace = TRUE)
+   Tboot[i] = median(rep)
+   }
> median <- mean(Tboot)
> median

[1] 90182.79

> MSE <- sum((median - Tboot) ^ 2) / length(Tboot)
> MSE

[1] 25032750

>
```