statiscal inference : assignement 1

vincent trouillet

6 mars 2018

0verview

The exponential distribution in R is investigating, and compare 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. In this study, the distribution of averages of 40 exponentials is used. A thousand simulations has been made.

1) Initialisation of simulation

```
lambda <- 0.2
n<- 40
nRandomUnif<-1000
```

set seed for reproducible calculation

```
set.seed(123)
```

run simulations

```
sim <- NULL
for (i in 1 : 1000) sim <- c(sim, mean(rexp(n,lambda)))
```

2) Sample Mean versus Theoritical mean

```
theoriticalMean = 1/lambda
sampleMean = mean(sim)
diff = abs(theoriticalMean - sampleMean)

tab <- matrix(c(theoriticalMean, sampleMean, diff),
              ncol = 3, byrow=TRUE)
colnames(tab) <- c("theoritical mean","sample mean","diff")
rownames(tab) <- c("mean")
tab <- as.table(tab)
tab
```

```
##      theoritical mean sample mean       diff
## mean       5.00000000  5.01191128 0.01191128
```

The two means are very close.

3) Sample variance versus Theoritical variance

```
theoriticalVar = (lambda*sqrt(n))^-2
sampleVar = var(sim)
diff = abs(theoriticalVar - sampleVar)

tab <- matrix(c(theoriticalVar, sampleVar, diff),
              ncol = 3, byrow=TRUE)
colnames(tab) <- c("theoritical var","sample mean var","diff")
rownames(tab) <- c("variance")
tab <- as.table(tab)
tab
```

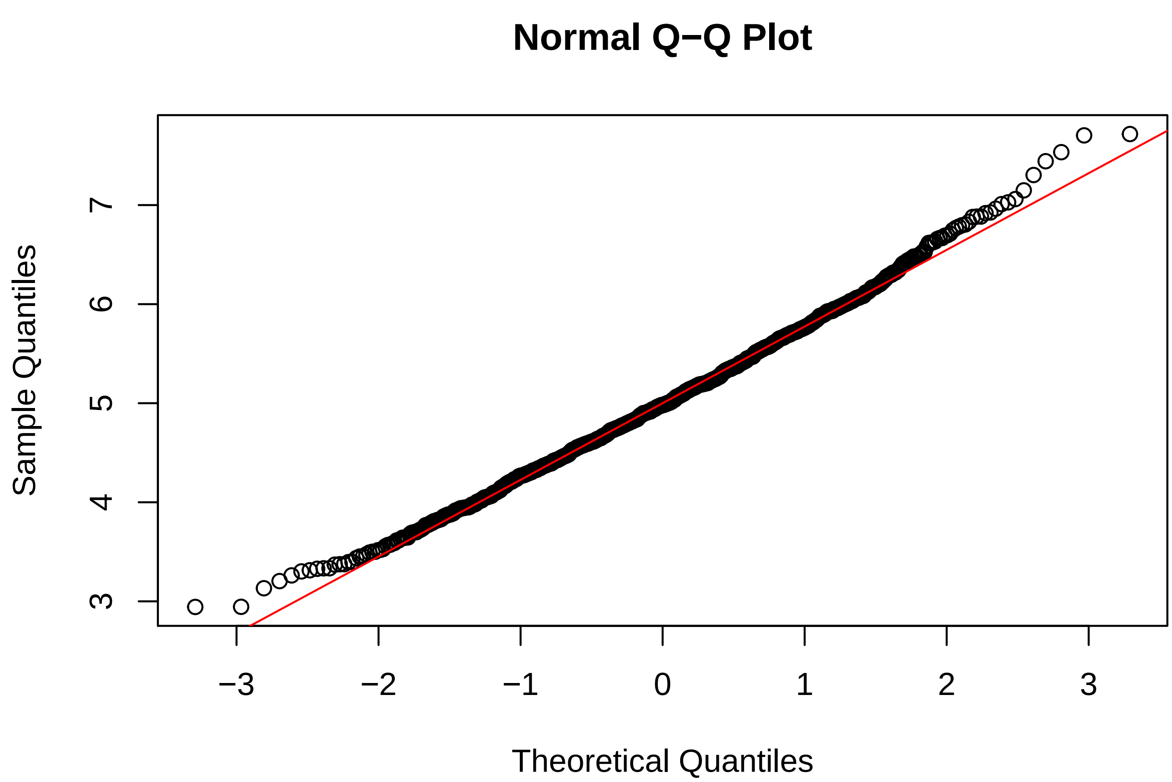
```
##          theoritical var sample mean var       diff
## variance      0.62500000      0.60049284 0.02450716
```

The two variances are very close.

4) Is the distribution approximately normal?

Q-Q Normal plot can be used to indicate the normal distribution of data.

```
qqnorm(sim)
qqline(sim, col = "red")
```



If the data is normally distributed, the points in the QQ-normal plot lie on a straight diagonal line (red line, illustrated with R code qqline). The deviations from the straight line are minimal. This indicates the sample mean is normal distribution, although the initial sample are not normal distributed.

The normal distribution can also be found from calculating the confidence intervals:

```
round((sampleMean + c(-1,1)*qnorm(.95)*sd(sim)/sqrt(nRandomUnif)),3)
```

```
## [1] 4.972 5.052
```

The 95% confidence interval of the sample mean is 4.972 to 5.052, which is very close to the theoretical mean of 5.000. It means there is 95% probability that the population mean lies between these two bounds of 4.972 and 5.052.