Si Project

Kun Qian

Thursday, February 12, 2015

Executive Summary

This is the project for the statistical inference class. In it, I will use simulation to explore inference and do some simple inferential data analysis. The project consists of two parts: 1. A simulation exercise. 2. Basic inferential data analysis.

Project Details

This project will investigate the exponential distribution in R and compare it 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. I will set lambda = 0.2 for all of the simulations. I will investigate the distribution of averages of 40 exponentials for 1000 times.

In the analysis below, the RMD file 1. Show the sample mean and compare it to the theoretical mean of the distribution. 2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution. 3. Show that the distribution is approximately normal.

1.Simulation

set seed and parameter

```
set.seed(3)
lambda = 0.2
n = 40
N = 1000
```

theoratical mean

```
tmean = 1/lambda
tmean
```

```
## [1] 5
```

theoratical standard deviation

```
tsd = 1/lambda
tsd
```

```
## [1] 5
```

Simulate 40 exponentials 1000 times

```
vexp = rexp(n*N,lambda)
vmat = matrix(vexp,n,N)
dim(vmat)
```

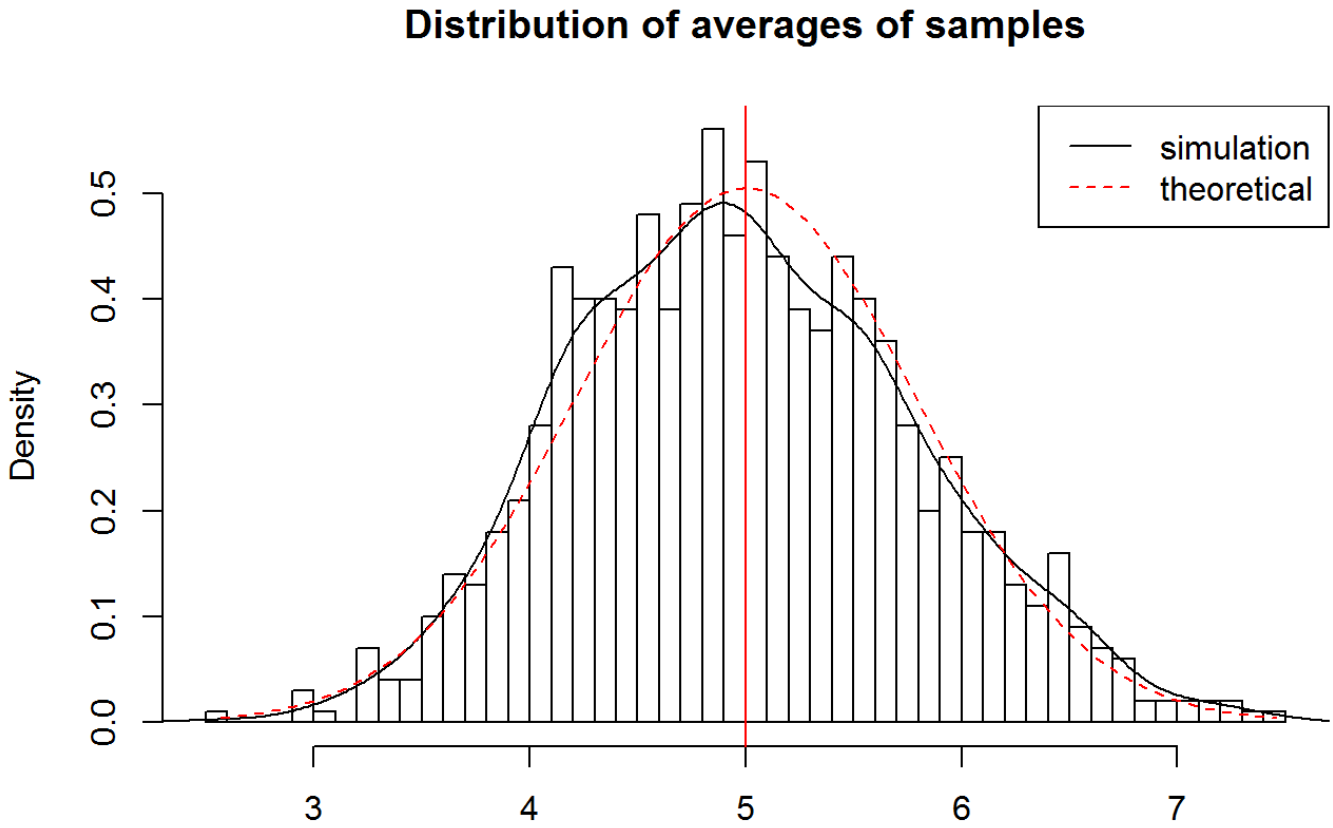
```
## [1]   40 1000
```

2.Generate samples of averages of exponentials

```
meanSample = apply(vmat,2,mean)
```

3.Histogram show the empirical distribution versus theoretical distribution

```
hist(meanSample, breaks=50, prob=TRUE,main="Distribution of averages of samples",xlab="")
lines(density(meanSample))
abline(v=tmean, col="red")
xfit <- seq(min(meanSample), max(meanSample), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(n)))
lines(xfit, yfit, pch=22, col="red", lty=2)
legend('topright', c("simulation", "theoretical"), lty=c(1,2), col=c("black", "red"))
```



Sample empirical distribution is very close to theoretical distribution from the histogram plot.

4. Compare theoretical mean and standard deviation vs simulated

theoretical mean:

```
tmean
```

```
## [1] 5
```

simulated mean:

```
mean(meanSample)
```

```
## [1] 4.98662
```

theoretical sd:

```
tsd/(n)^0.5
```

```
## [1] 0.7905694
```

simulated sd:

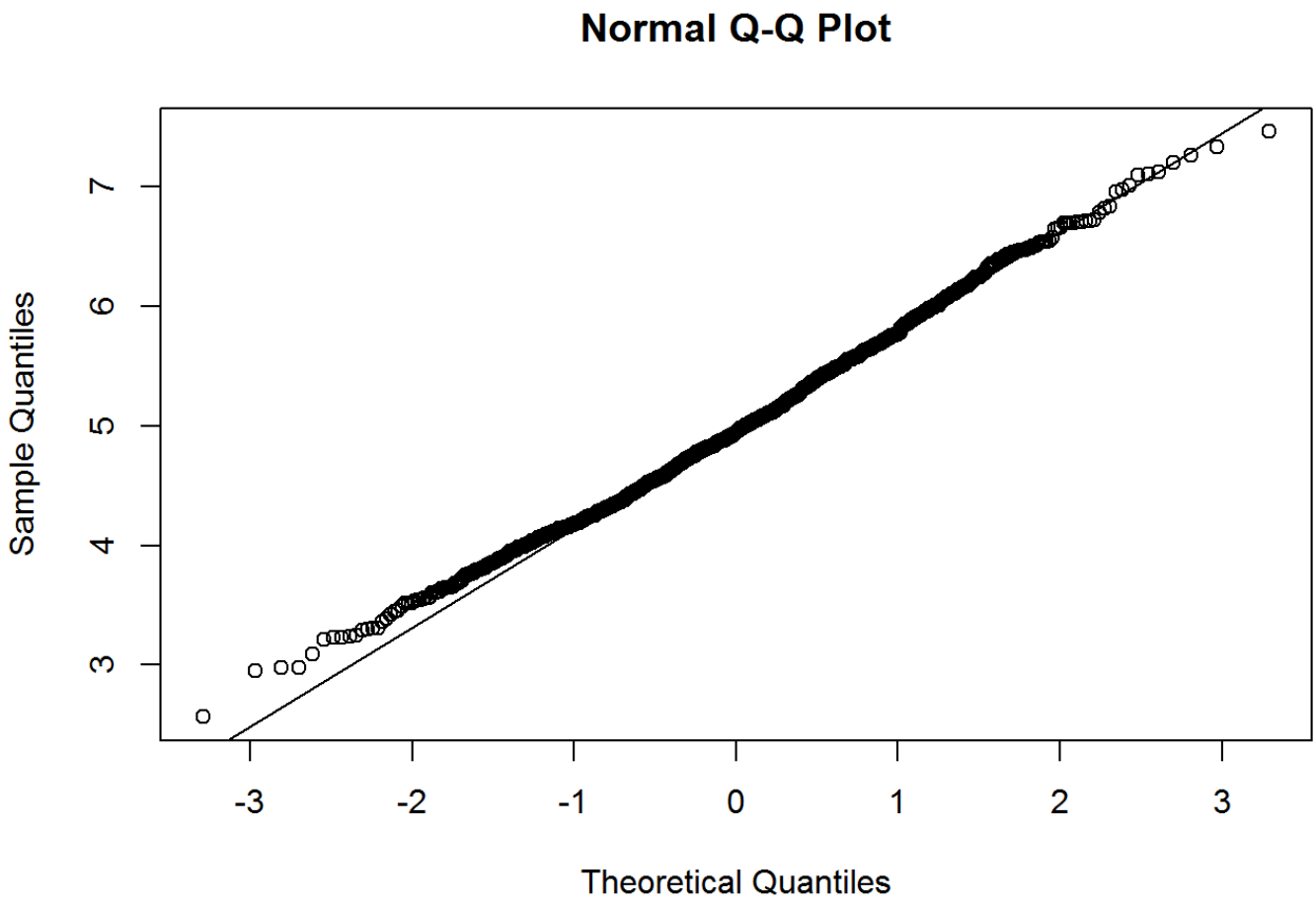
```
sd(meanSample)
```

```
## [1] 0.7947823
```

As you can see, the simulated vs theoretical are close enough given 1000 times simulation. As the simulation time increasing, the results will be even more converged.

4.Check Normality

```
qqnorm(meanSample)
qqline(meanSample)
```



From the qq plot, meanSample of 1000 is showing normality.