Statistical Inference Course Project Part1

Yao Bin

June 12, 2015

Overview

- In this study, we investigate the exponential distribution in R and compared it with the Central Limit Theorem. The distribution of averages of 40 exponentials are investigated.

- Code chunks of the data analysis and plots are listed in appendix section in the end of this document.

Simulations

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.

In this study, we set n=40 and lambda=0.2. Generate 1000 simulations of averages of 40 random exponentials: (see Chunk1)

Sample Mean versus Theoretical Mean

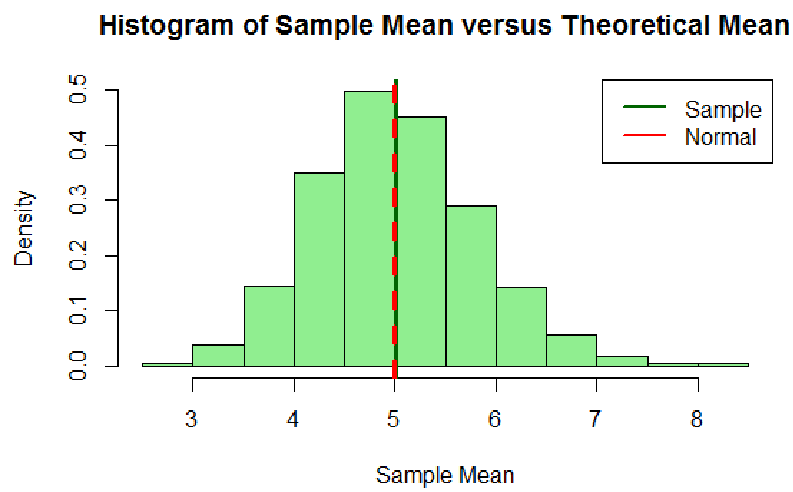
As mentioned above, the theoretical mean mean_t=1/lambda

```
[1] "Theoretical mean mean_t= 5"

[1] "Sample mean mean_s= 5.0096"
```

The sample mean **`5.0096`** is close to theoretical mean **`5`**.

Here, we draw a plot to show that the sample mean compared with the theoretical mean.



Sample Variance versus Theoretical Variance

The theoretical variance for the exponential distribution is sigma^2 =(1/lambda)^2/n. The sample variance is calculated by var() function in R.

```
[1] "Theoretical variance var_t = 0.625"

[1] "Sample variance var_s= 0.6589"
```

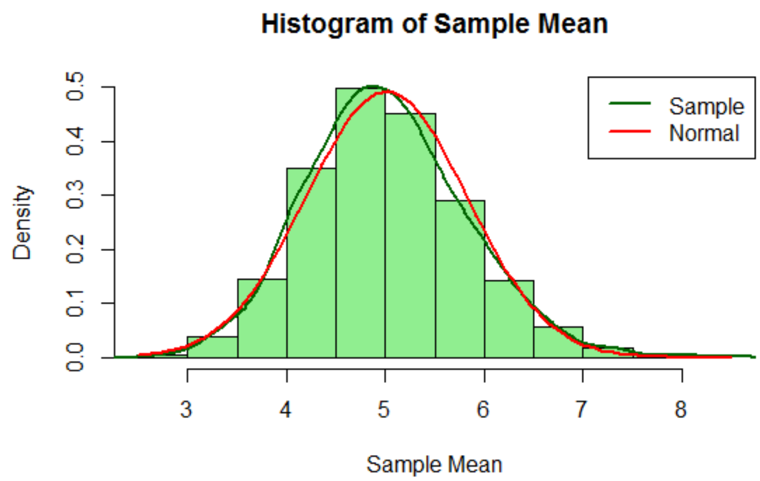
Also, the sample variance **`0.6589`** is close to theoretical variance **`0.625`**.

Distribution

Here, 2 methods were performed to assess if the sample means were roughly normally distributed.(see Chunk5)

1. Standard normal density curve

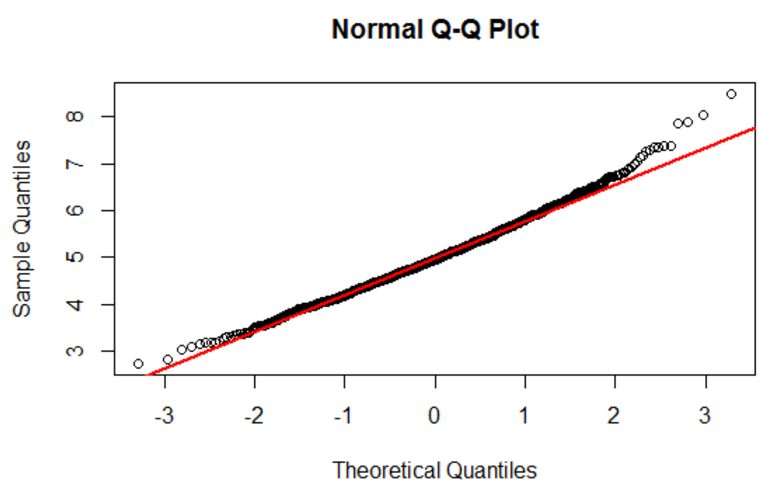
Draw a normal density curve over the distribution of the sample means, and compare it with the distribution density curve.



The normal density curve mostly matched the distribution density curve.

2. Q-Q plot

Quantile-quantile plot (Q-Q plot) is also a widely used method to check the normality of a distribution.



The Q-Q plot shows the points scatter along the line, which indicates that this distibution is close to normal distribution.

In conculsion, the distribution of the sample means were approximately a normal distribution.

Appendix: Code Chunks

Chunk1. Parameter setting and simulations

```
# set the parameters
n=40
simu=1000
lambda=0.2
# simulation
exp<-NULL
for(i in 1:simu) exp<-c(exp, mean(rexp(n=n, rate=lambda)))
```

Chunk2. Calculating theoretical mean and sample mean

```
# theoretical mean
mean_t<-1/lambda
# sample mean
mean_s<-mean(exp)
mean_t;mean_s
```

Chunk3. Plotting theoretical mean and sample mean

```
hist(exp, prob=T, col="lightgreen",xlab="Sample Mean",
     main="Histogram of Sample Mean versus Theoretical Mean")
# plot the sample mean
abline(v=mean_s, col="darkgreen", lwd=3)
# plot the theoretical mean
abline(v=mean_t, col="red", lty=2, lwd=3)
legend("topright", c("Sample", "Normal"), col=c("darkgreen", "red"), lw
d=2)
```

Chunk4. Calculating theoretical variance and sample variance

```
# theoretical variance
var_t<-(1/lambda)^2/n
# sample variance
var_s<-var(exp)
var_t;var_s
```

Chunk5. Distribution test

```
# plot histogram of the sample means
hist(exp, prob=T, col="lightgreen",xlab="Sample Mean", main="Histogram
of Sample Mean")
# plot sample distribution
lines(density(exp), lwd=2, col="darkgreen")
# plot normal distribution
curve(dnorm(x, mean=mean(exp), sd=sd(exp)), add=T, lwd=2, col="red")
legend("topright",c("Sample", "Normal"), col=c("darkgreen", "red"), lwd
```

```r
=2)

# Q-Q plot
qqnorm(exp)
qqline(exp, col="red", lwd=2)
```