Central Limit Theorem simulation with exponential distribution

X. Zeng

August 18, 2015

Overview

In this report, we will investigate the distribution of averages of 40 exponentials with lamda=0.2 for all the simulation, to verify the Central Limit Theory. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda.

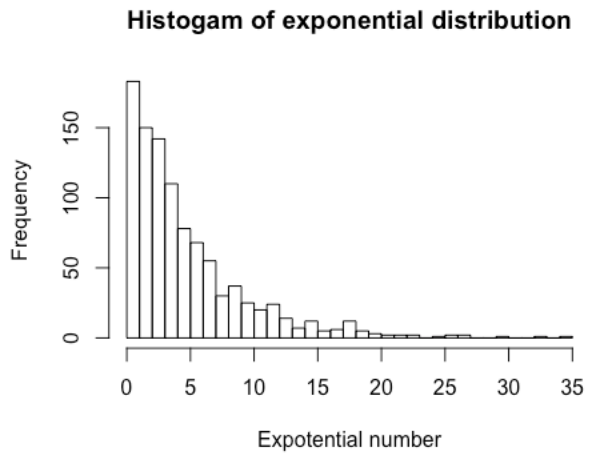
Simulations

First, we need to set up the parameters for lamda, the total number of simulations `n=1000` and average number `m=40`

```
set.seed(12345)
lamda <- 0.2
n <- 1000
m <- 40
```

Then, we show the distribution of exponentail numbers:

```
hist(rexp(n, lamda), breaks=40, xlab="Expotential number",
ylab="Frequency",main="Histogam of exponential distribution")
```



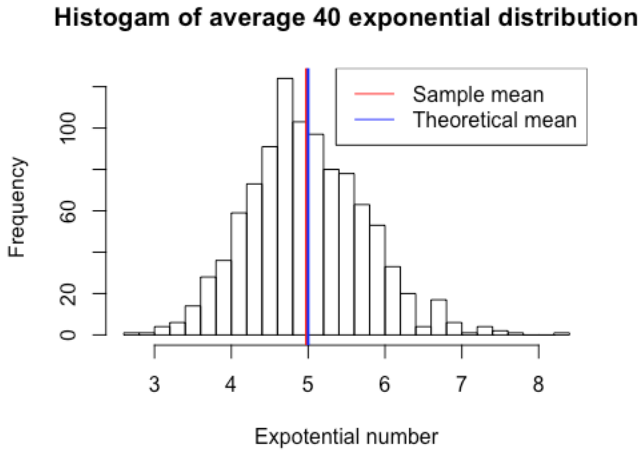
Here, we will generate the average 40 exponentials with 1000 simulations.

```r
mns = NULL
for (i in 1:n) {
  mns = c(mns, mean(rexp(m, lamda)))
}
```

Sample Mean versus Theoretical Mean

Here shows the histogram of the simulated numbers `mns` with sample mean 4.9770396 (denoted by red line) and theoretical mean 5 (denoted by blue line). As we can see, they are quite close.

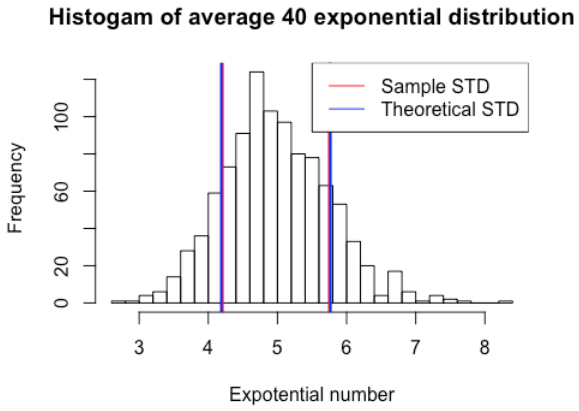
```r
hist(mns, breaks=40, xlab="Expotential number", ylab="Frequency",
     main="Histogam of average 40 exponential distribution")
abline(v=mean(mns),col="red", lwd=2)
abline(v=1/lamda, col="blue", lwd=2)
legend("topright",c("Sample mean", "Theoretical mean"),
col=c("red","blue"),lty=c(1,1))
```



Sample Variance versus Theoretical Variance

The section deals with the sample and theoretical variances. The red lines in the histogram indicate 1 sample standard deviation from the sample mean (red lines), and blue lines indicate 1 theoretical standard deviation from the sample mean (blue lines). The variances are simply sqaures of the standard deviations. That is, sample variance is 0.5941418, and theoretical variance is 0.7905694. They are qutie similar.

```r
hist(mns, breaks=40, xlab="Expotential number", ylab="Frequency",
     main="Histogam of average 40 exponential distribution")
abline(v=mean(mns)-sd(mns),col="red", lwd=2)
abline(v=mean(mns)+sd(mns),col="red", lwd=2)
abline(v=mean(mns)-1/lamda/sqrt(m), col="blue", lwd=2)
abline(v=mean(mns)+1/lamda/sqrt(m), col="blue", lwd=2)
legend("topright",c("Sample STD", "Theoretical STD"),
col=c("red","blue"), lty=c(1,1))
```

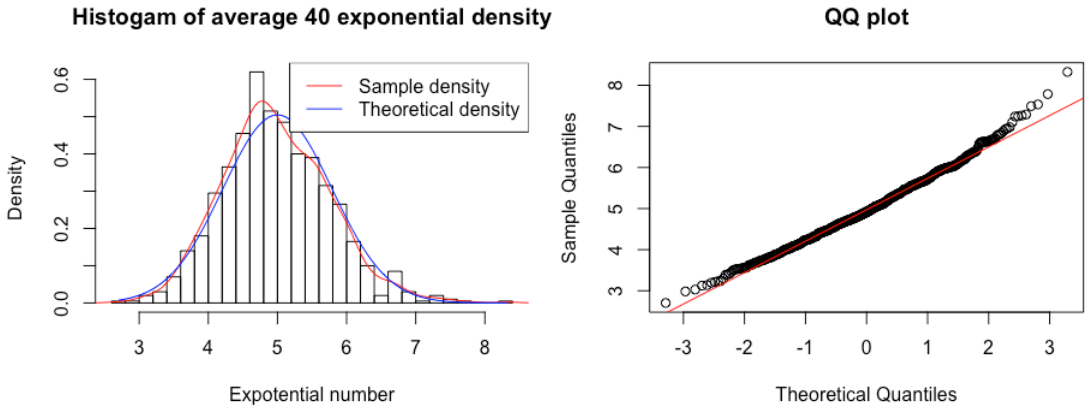


Distribution

From the above histogram, we can see gengenraly the distribution is quite similar as the norm distribution. So we calcualte the density and add the theoretical (blue) and simple density lines (red). The theoretical norm distribution is simply N(1/ lamda, $1/(\text{lamda}^2 * \text{m})$).

```
hist(mns, breaks=40, prob=TRUE, xlab="Expotential number",
ylab="Density",main="Histogam of average 40 exponential density")
lines(density(mns),col="red")
lines(seq(min(mns),max(mns),length=n),
dnorm(seq(min(mns),max(mns),length=n),mean=1/lamda,sd=1/lamda/sqrt(m)),
col="blue")
legend("topright",c("Sample density", "Theoretical density"),
col=c("red","blue"),lty=c(1,1))
```

```
qqnorm(mns, main="QQ plot")
qqline(mns, col="red")
```



From the quantile plot comparison, we can see the distribution generally is a norm distribution.

Appendix. R code

```
# This is for Statistical Inference Projcet 1
# Investigate the exponential distribution in R and compare it with the
Central Limit Theorem.

set.seed(12345)

# set lamda for the exponential distribution
lamda <- 0.2

# set random sampel number n and avearge number m
n <- 1000
m <- 40

# distribution of n random exponential distribution
hist(rexp(n, lamda), breaks=40,
      xlab="Expotential number", ylab="Frequency",
      main="Histogam of exponential distribution")

# distribution of n averages of m random exponentials & compare the
mean
mns = NULL
for (i in 1:n) {
  mns = c(mns, mean(rexp(m, lamda)))
}
hist(mns, breaks=40, xlab="Expotential number", ylab="Frequency",
     main="Histogam of average 40 exponential distribution")
abline(v=mean(mns),col="red", lwd=2)
abline(v=1/lamda, col="blue", lwd=2)
legend("topright",c("Sample mean", "Theoretical mean"),
col=c("red","blue"),
       lty=c(1,1))

# compare the variance
hist(mns, breaks=40, xlab="Expotential number", ylab="Frequency",
     main="Histogam of average 40 exponential distribution")
abline(v=mean(mns)-sd(mns),col="red", lwd=2)
abline(v=mean(mns)+sd(mns),col="red", lwd=2)
abline(v=mean(mns)-1/lamda/sqrt(m), col="blue", lwd=2)
abline(v=mean(mns)+1/lamda/sqrt(m), col="blue", lwd=2)
legend("topright",c("Sample STD", "Theoretical STD"),
col=c("red","blue"),
       lty=c(1,1))


# density distribution of sample and theoretical results
hist(mns, breaks=40, prob=TRUE, xlab="Expotential number",
ylab="Density",
```

```r
     main="Histogam of average 40 exponential density")
lines(density(mns),col="red")
lines(seq(min(mns),max(mns),length=n),

dnorm(seq(min(mns),max(mns),length=n),mean=1/lamda,sd=1/lamda/sqrt(m)),
col="blue")
legend("topright",c("Sample density", "Theoretical density"),
col=c("red","blue"),
       lty=c(1,1))

# qqplot
qqnorm(mns, main="QQ plot")
qqline(mns, col="red")
```