Statistical Inference: Simulation

by YK Zhao

0 Introduction

In this project you 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 λ. 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. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

1 Overview

In this final project, I will illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. I will do the following steps:
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.

2 Simulations

In this section, I will first simulate exponential distribution for 1000 times as required in guidelines and I will set parameter $\lambda = 0.2$ and sample size $n = 40$. Then I will calculate average of 40 random values for each simulation and calculate mean and standard deviation for this average.

```
##Set environment
##setwd("C:/Users/User/Desktop/Statistical Inference/")
library(knitr)
library(plotrix)
opts_chunk$set(echo=TRUE)
```

```
##Set parameter, sample size and simulation size
lambda <- 0.2
n <- 40
N <- 1000
```

```
##Set seed for random draw
set.seed(999)
```

```
##Simulate N times and calculate mean and standard deviation
meanValue <- NULL ##Initialize
meanStd <-NULL

for (i in 1:N){

  ##Calculate mean and std of sample means for each simulation
  expo <- rexp(n, lambda)
  mean <- mean(expo)
```

```
  std <- sd(expo)
  meanValue <- c(meanValue, mean)
  meanStd <- c(meanStd, std)

} ##End for
```

3 Results

In this section, I will compare sample mean and std of simulation with theoretical mean and std via figures and textual explanations.

First I will show histogram of simulated samples with frequency and with density. The density curve is Gaussian type, centering at 5 and concentrated at the center of histgram plot.
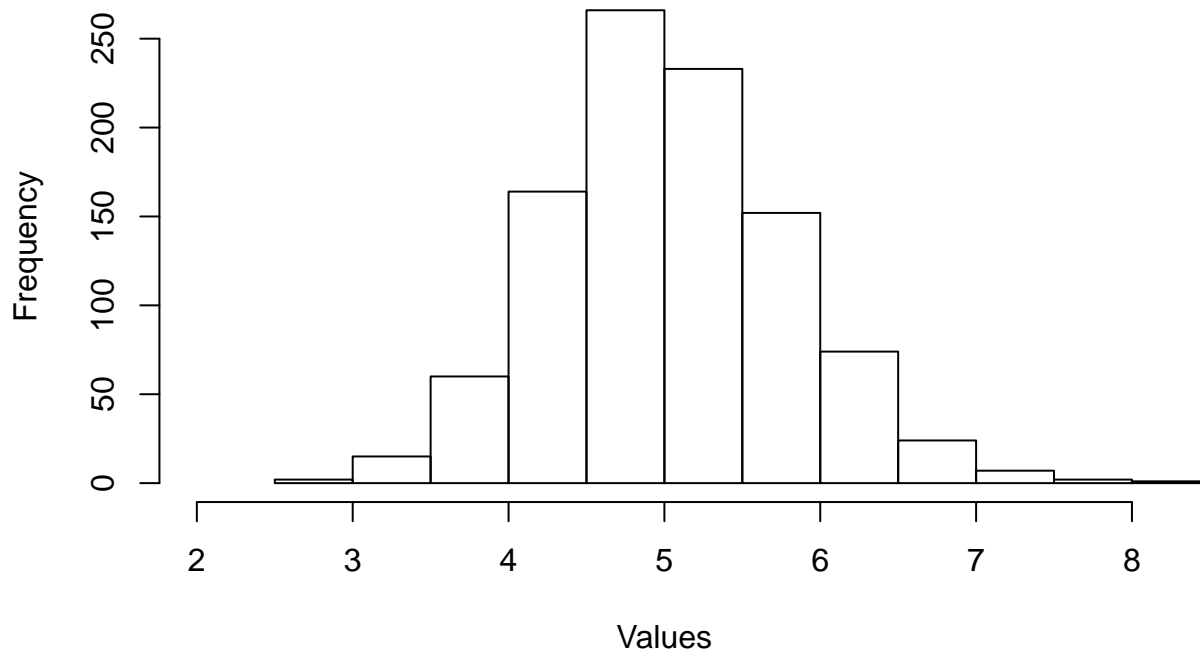
```
# All plotting is clipped to the 2*1 device region
par(mfrow = c(1,1), xpd=NA)

##Plot histogram of frequency and probability density
expoHistFreq <- hist(meanValue, freq=TRUE, xlim=c(2,8),
                     main=paste("Histogram of",N,"simulations"), xlab="Values")
```


Histogram of 1000 simulations
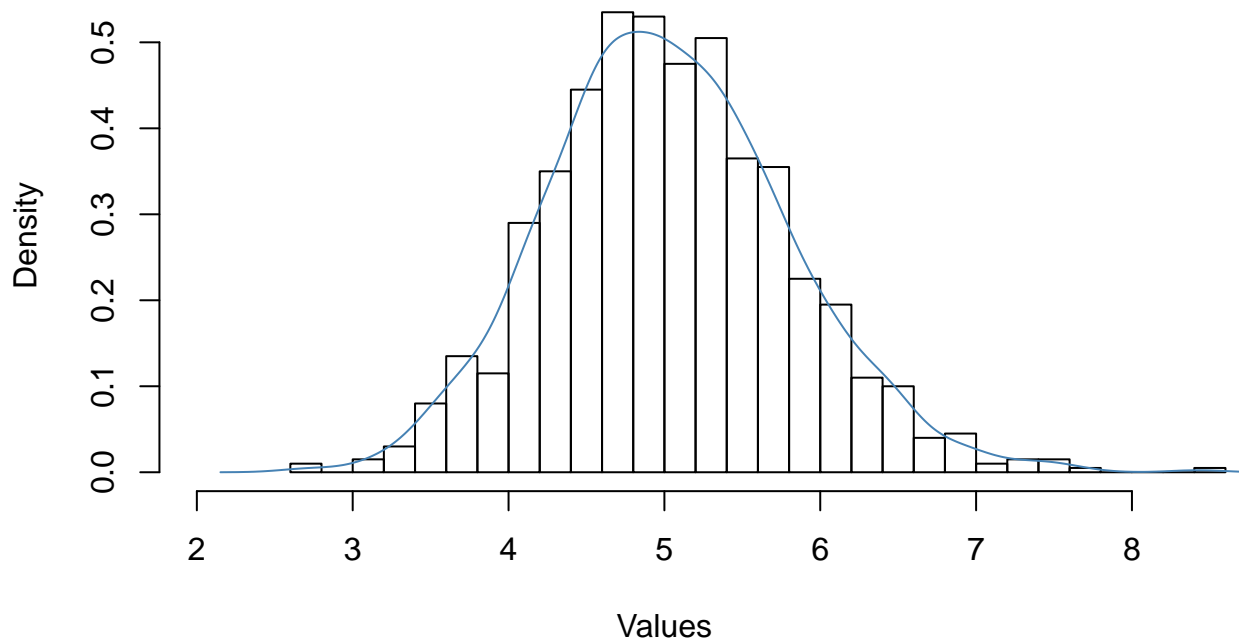
```
expoHistProb <- hist(meanValue, freq=FALSE, xlim=c(2,8), ylim=c(0,0.55), breaks=25,
                     main=paste("Probability density of",N,"simulations"), xlab="Values")
lines(density(meanValue), col="steelblue") ##Simulated Gaussian probability density curve
```

Probability density of 1000 simulations

Next I will calculate mean and std of simulated samples and compare them with theoretical mean and std. And then I will plot them in figures.

```
##Sample mean and sample variance
meanSample <- mean(meanValue)
stdSample <- sd(meanValue)

##Theoretical mean and theoretical variance
meanThy <- 1/lambda
stdThy <- 1/lambda

##Compare sample mean and theoretical mean
meanSample ##meanSample=meanThy
```

```
## [1] 5.029028
```

```
meanThy
```

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

```
##Compare sample variance and theoretical variance
stdSample ##stdSample=stdThy/sqrt(n)
```

```
## [1] 0.7782133
```

```
stdThy
```

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

```
##Plot and compare sample mean with theoretical mean
par(mfrow = c(1,1))

hist(meanValue, freq=FALSE, xlim=c(2,8), ylim=c(0,0.55), breaks=25,
     main=paste("Probability density of",N,"simulations"), xlab="Values")

abline(v=meanSample, col="red", lwd=3, lty=2) ##Plot sample mean
abline(v=meanThy, col="blue", lwd=3, lty=9) ##Plot theoretical mean

##Plot theoretical normal distribution density curve
x <- seq(min(meanValue), max(meanValue))

curve(dnorm(x, mean=meanSample, sd=stdSample), col="black", lwd=3, lty=3, add=TRUE)

legend('topright', c(paste("Sample mean", meanSample), paste("Theoretical mean", meanThy),
                    "Normal distrubution curve"), lty=1, col=c('red', 'blue', "black"),
       bty='n', cex=.75) ##Add indicators
```
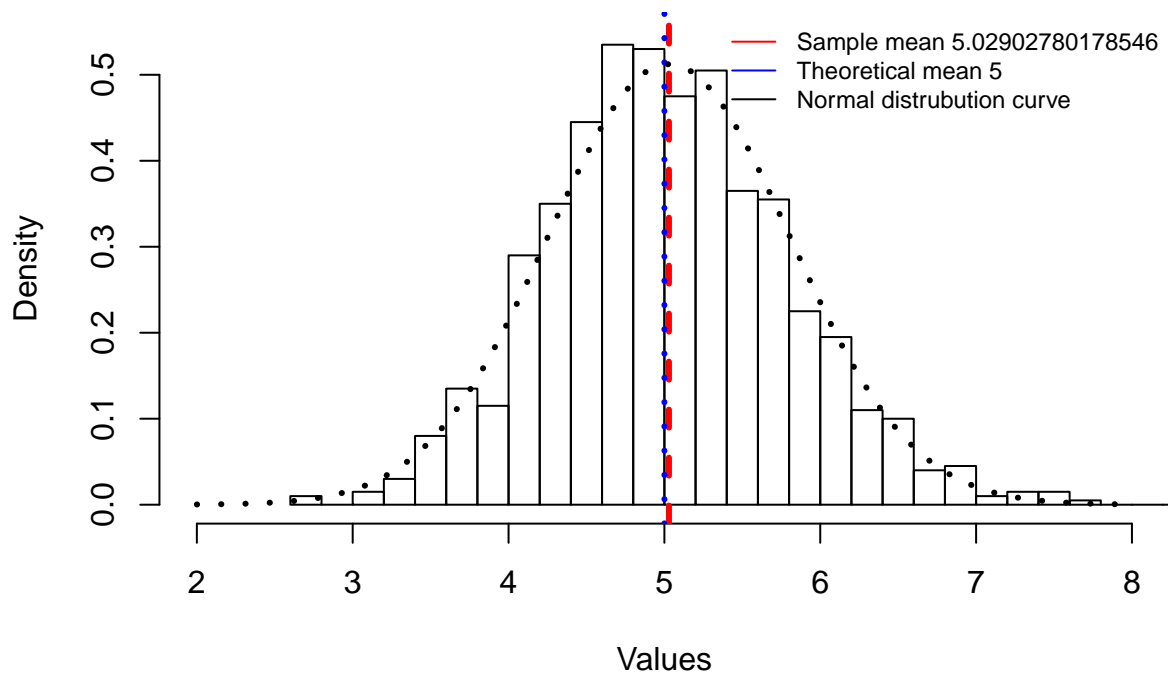
## Probability density of 1000 simulations

The theoretical mean and standard deviation of exponential distribution are both $1/\lambda$, which are 5 and 5 respectively. The sample mean and standard deviation of exponential distribution in this project which sets simulation number 1000 are 5.0290278 and 0.7782133 respectively. It is obvious that sample mean is almost

4

equal to theoretical mean, while sample standard deviation nearly equals to theoretical standard deviation divided by $\sqrt{40}$. These are consistent with **Central Limit Theorem**.

4 Distribution

In this section, I will use Q-Q plot to check normality of these simulated exponential distributed samples.

```
##Q-Q plot of mean values to show normality
qqnorm(meanValue)
qqline(meanValue)
```

Normal Q–Q Plot
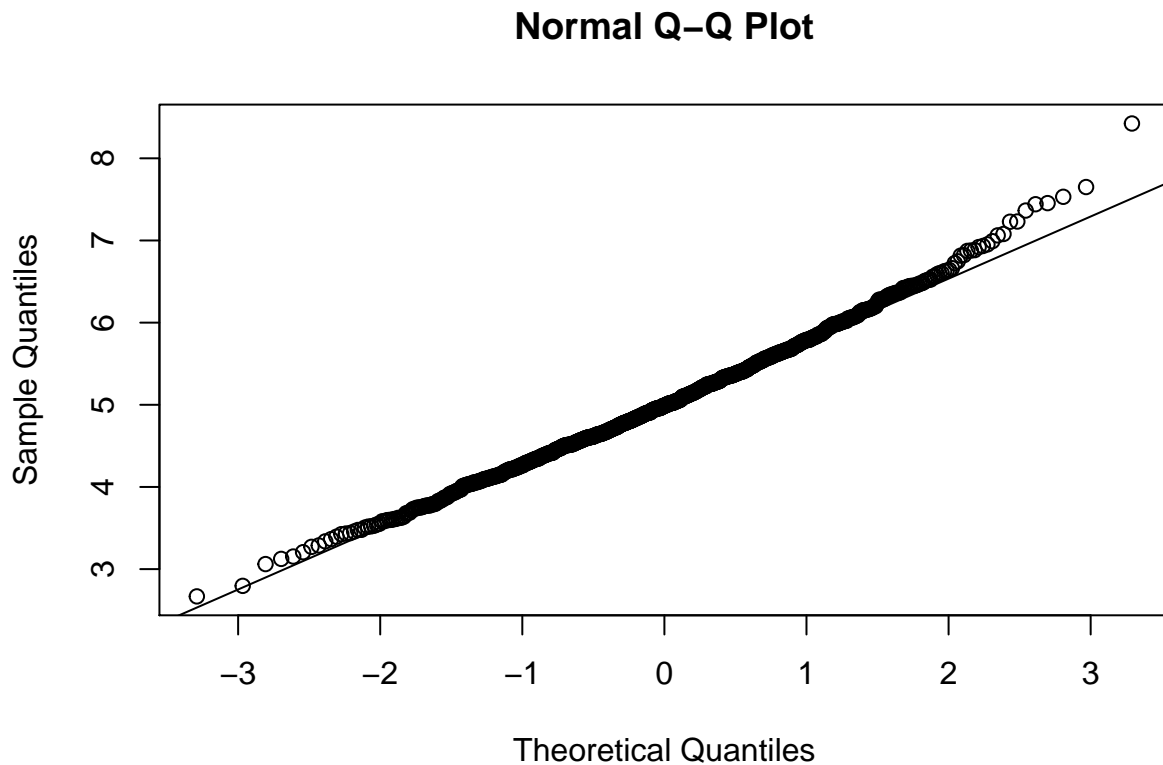
It is obvious from plots that simulated quantile line is very close to theoretical quantile line, indicating that these simulated samples nearly follow normal distribution.