Statistical Inference Assignment 1 Part 1

MHiero

Wednesday, September 17, 2014

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 also 1/lambda. Set lambda = 0.2 for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials.

```
# Prerequisites and simulation run with 1000 repeats
library(knitr)
library(plotrix)
library(ggplot2)


lambda <- .2
n <- 40
repeats <- 1000
set.seed(4711)
sim <- replicate(repeats, rexp(n,lambda))
```

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.

```
sim.mean <- sapply(1:repeats, function(i) mean(sim[,i]))
# Calculated mean and theoretical mean
res1 <- data.frame('mean'=c(mean(sim.mean),1/lambda),
row.names=c('calculated:','theoretical:'))
res1

##              mean
## calculated:  5.021
## theoretical: 5.000
```

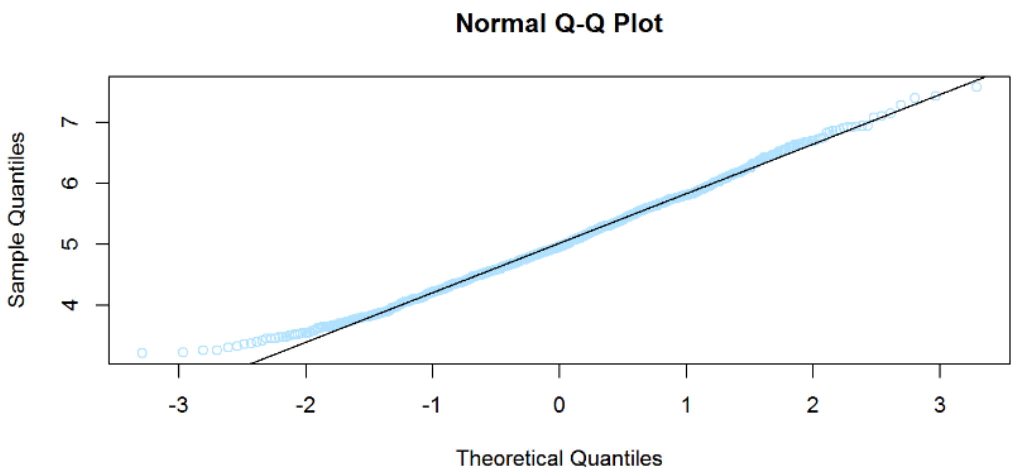
There is a tiny difference the calculated and the theoretical center of the distribution.

2. Show how variable it is and compare it to the theoretical variance of the distribution.

```
sim.sd <- sd(as.vector(sim.mean))
sim.var <- var(as.vector(sim.mean))
theo.sd <- (1/lambda * 1/sqrt(n))
theo.var <- theo.sd ^ 2
res2 <- data.frame('sd'=c(sim.sd, theo.sd), 'var'=c(sim.var,theo.var),
row.names=c('calculated:','theoretical:'))
res2
```

```
##                  sd    var
## calculated:  0.7968 0.6349
## theoretical: 0.7906 0.6250
```

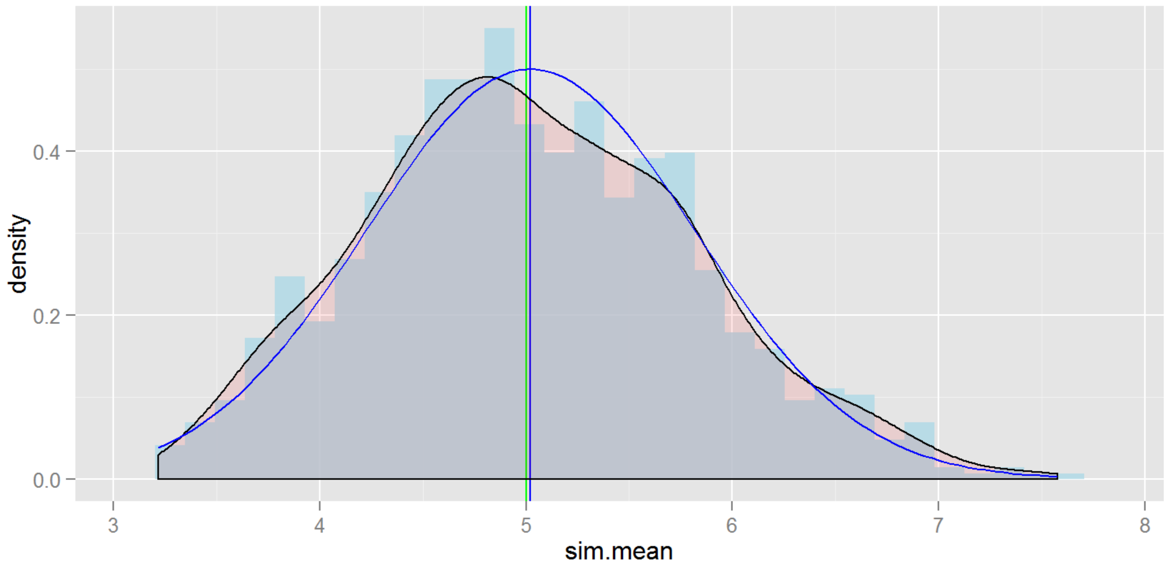
```
qqnorm(sim.mean, col = "lightskyblue1")
qqline(sim.mean)
```



Again a very tiny difference beetween theoretical and calculated values.

3. Show that the distribution is approximately normal.

```
# Distribution density plot
sim.mean <- as.data.frame(sim.mean)
g <- ggplot(sim.mean,aes(x=sim.mean))
g + geom_histogram(aes(y = ..density..),fill='lightblue',alpha=.8) +
    geom_vline(data=res1, aes(xintercept=mean), color=c('blue','green')) +
    geom_density(alpha=.1, fill='red') +
    stat_function(fun=dnorm, args=list(mean=mean(sim.mean[,1]), sd=sim.sd),color =
"blue")
```



The density plot of simulation is nearly identical to the shape of normal distribution (which is blue curve in the diagram).

4. Evaluate the coverage of the confidence interval for 1/lambda

Evaluate the coverage of the confidence interval for 1/lambda: X +/- 1.96 S/sqrt(n).

```
# Calculating confidence interval for each simulation

coverage <- data.frame('Low'=NA,'High'=NA)
for(i in 1:repeats){
    coverage[i,]<- mean(sim[,i])+c(-1.96,1.96)*sd(sim[,i])/sqrt(n)
    }
head(coverage,n=5)

##     Low  High
## 1 3.128 5.426
## 2 3.768 7.450
## 3 3.192 6.191
## 4 2.921 6.075
## 5 3.487 5.874

Coverage <- nrow(coverage[which(coverage$Low < 5 & coverage$High > 5),])/repeats
Coverage

## [1] 0.923
```

The confidence interval containing the theoretical mean is 0.92 or 92%.