Statistical Inference Project One - Simulation

Tianxiang Liu

Sunday, September 07, 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, we will investigate the distribution of averages of 40 exponential(0.2)s. Note that we will need to do a thousand or so simulated averages of 40 exponentials.

```
# Setup and run the simulation
setwd('C:\\Users\\Ivan.Liuyanfeng\\Desktop\\Data_Mining_Work_Space\\Data-Science-Statistical
-Inference\\project')
library(ggplot2)
lambda <- 0.2
n <- 40
itr <- 1000
set.seed(1234)
sim <- replicate(itr, rexp(n,lambda))
```

Section One

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

```
sim.mean <- sapply(1:itr, function(i) mean(sim[,i]))
# Calculated mean and theoretical mean
s1 <- data.frame('mean'=c(mean(sim.mean),1/lambda), row.names=c('calculated','theoretical'))
s1
```

```
##              mean
## calculated  4.974
## theoretical 5.000
```

Section Two

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

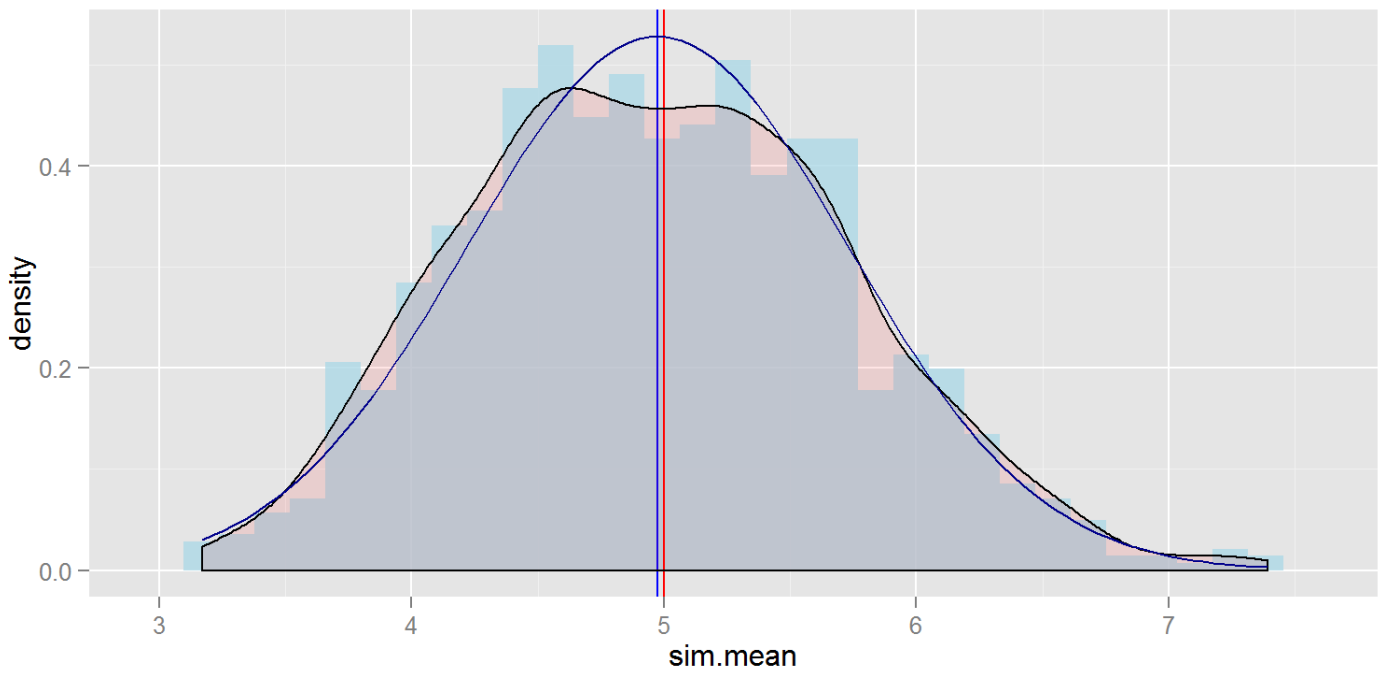
```
sim.sd <- sd(as.vector(sim.mean))
sim.var <- var(as.vector(sim.mean))
theo.sd <- (1/lambda * 1/sqrt(n))
theo.var <- theo.sd ^ 2
# Calculated sd, var and theoretical sd, var
s2 <- data.frame('sd'=c(sim.sd, theo.sd), 'var'=c(sim.var,theo.var), row.names=c('calculated
','theoretical'))
s2
```

```
##                  sd    var
## calculated  0.7554 0.5707
## theoretical 0.7906 0.6250
```

Section Three

Show that the distribution is approximately normal.

```
# 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=s1, aes(xintercept=mean), color=c('blue','red')) +
    geom_density(alpha=.1, fill='red') +
    stat_function(fun=dnorm, args=list(mean=mean(sim.mean[,1]), sd=sim.sd),color = "darkblue
")
```



We can see the shape of density plot of simulation is almost identical to the shape of normal distribution (which is dark blue line in the diagram).

Section Four

Evaluate the coverage of the confidence interval for 1/lambda: $\bar{X} \pm 1.96 Snv$.

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

```
##     Low  High
## 1 3.389 6.549
## 2 4.207 7.290
## 3 2.447 4.184
## 4 4.335 8.496
## 5 3.519 5.812
```

```
Coverage <- nrow(coverage[which(coverage$Low < 5 & coverage$High > 5),])/itr
# Coverage of the confidence interval for 1/lambda is:
Coverage
```

```
## [1] 0.932
```

Thank you for reading! Ivan Liu 07/09/2014