Simulation Exercise on Exponential Distribution

Zhang,Xiaoyu

Thursday, May 21, 2015

Overview

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. 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.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should:

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.

Simulations

The simulation can be built in R using the rexp(n,lambda) function, where n is the number of exponentials per set, and lambda is the rate parameter(Set lambda = 0.2 of the simulations). We need to build a single column data frame with the average of each of the sets. We run 1000 simulations and use a matrix to collect the simulation data with 1000 rows and 40 columns.

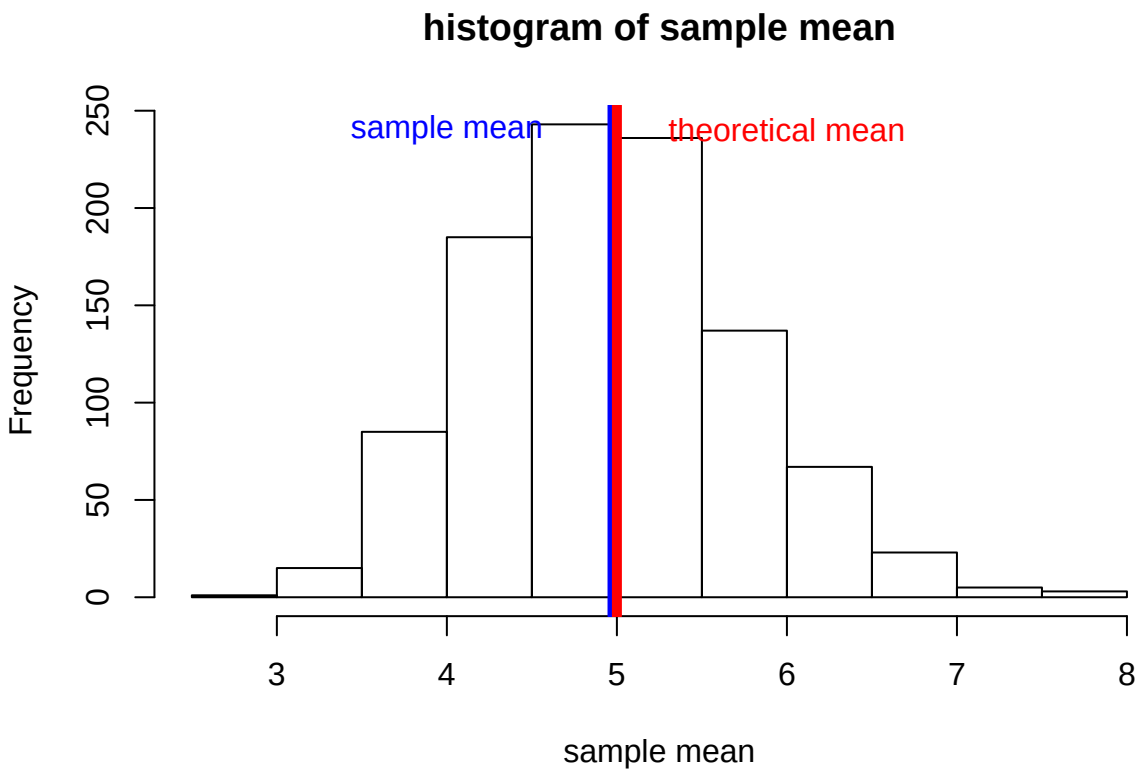
```
nosim <- 1000
lambda <- 0.2
n <- 40

set.seed(1234)
dat <- matrix(rexp(n*nosim,lambda),nrow=nosim)
```

Sample Mean versus Theoretical Mean

We calculate the row mean of the simulation data to get the sample of each simulation and calculate the mean of 1000 simulations(1000 rows) to get the sample mean.

```
sim_mean <- mean(rowMeans(dat))
theory_mean <- 1/lambda
hist(rowMeans(dat),xlab='sample mean',main='histogram of sample mean')
abline(v=sim_mean,col='blue',lwd=5)
text(x=4.0,y=240,labels='sample mean',col='blue')
abline(v=theory_mean,col='red',lwd=5)
text(x=6.0,y=240,labels='theoretical mean',col='red')
```



The figure above shows the distribution of the 1000 samples, the sample mean is 4.9742388(blue line in the figure), the theretical mean is 1/lambda = 5. As is shown, the two mean values are very close.

Sample Variance versus Theoretical Variance

We then calculate the row mean of the simulation data to get the sample of each simulation and calculate the variance of 1000 simulations(1000 rows) to get the sample variance.

```
sim_var <- var(rowMeans(dat))
theory_var <- (1/lambda)^2/n
```

The sample variance of the simulation data is `sim_var`,while the theoretical variance is `theory_var`

Distribution

Finally we plot the distribution of the sample. The green line gives the density curve of the sample and the red line gives the normal curve with parameters of sample mean and sample standard deviation. We can see clearly that they are very similar,which states the central limit theorem.

```
library(ggplot2)
dat <- data.frame(x=rowMeans(dat))
g <- ggplot(dat,aes(x=x)) +
        geom_histogram(aes(y=..density..),binwidth=0.2,col='black',fill='blue')
```

```
g <- g + geom_density(size=2,color='green')
g + stat_function(fun=dnorm,
                  args=list(mean=sim_mean,sd = sqrt(sim_var)),
                  color='red',size=2)
```