JHU Statistical Inference - Simulations

LiJun

24 October 2018

Introduction

This analysis aims to 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.

Simulations

```
#Set simulation variables lambda, exponentials an seed
set.seed(1234)
lambda = 0.2
exponentials = 40

#Run Simulations with variables
simulations = NULL
for (i in 1 : 1000) simulations = c(simulations, mean(rexp(exponentials,
lambda)))
```

1.Show the sample mean and compare it to the theoretical mean of the distribution.

```
#Sample mean - mean of the simulations
mean(simulations)

## [1] 4.974239

#Theoretical mean - lambda^-1
lambda^-1

## [1] 5

#Comparison between sample mean and theoretical mean
compare_mean = abs(mean(simulations)-lambda^-1)
compare_mean

## [1] 0.02576123
```

Conclusion: There is only a slight difference between the sample mean and the inference mean.

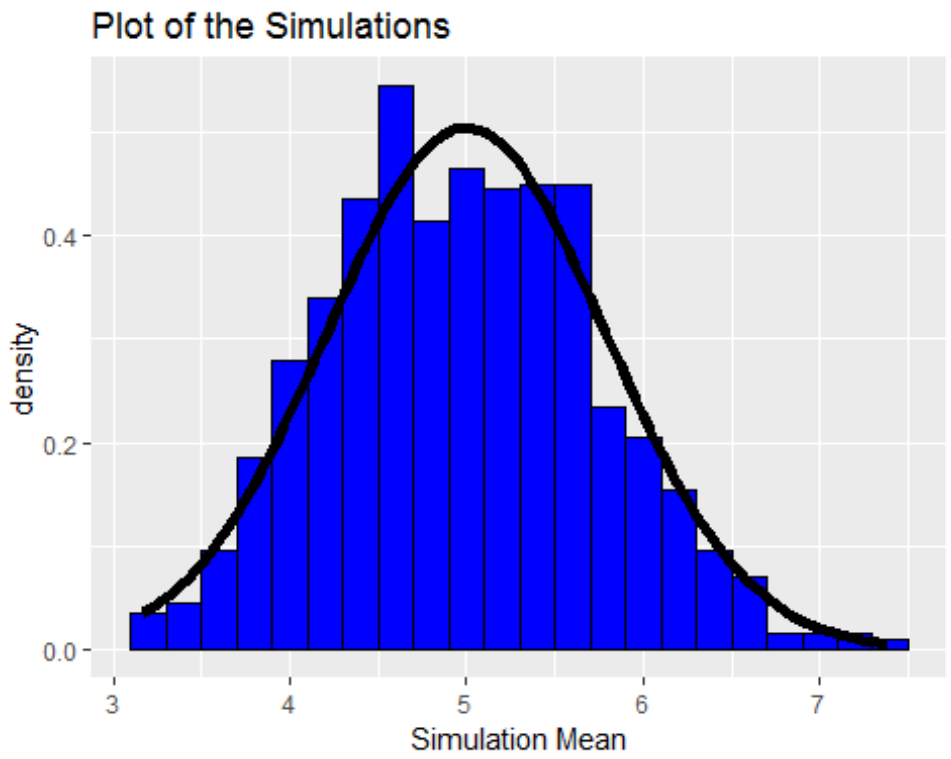
2.Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
#Sample variance - variance of the simulations
var(simulations)

## [1] 0.5706551

#Theoretical mean - (lambda * sqrt(n))^-2
(lambda * sqrt(exponentials))^-2

## [1] 0.625

#Comparison between sample variance and theoretical variance
compare_variance = abs(var(simulations)-(lambda * sqrt(exponentials))^-2)
compare_variance

## [1] 0.05434495
```

Conclusion: There is only a slight difference between the sample variance and the inference variance.

3.Show that the distribution is approximately normal.

```
#Distribution - based on density histogram of 1000 simulations, overlay with
theoretical normal distribution
library(ggplot2)
ggplot(data.frame(y=simulations), aes(x=y)) +
  geom_histogram(aes(y=..density..), binwidth=0.2, fill="blue",
color="black") +
  stat_function(fun=dnorm, args=list(mean=lambda^-1,
                                     sd=(lambda*sqrt(exponentials))^-1),
                size=2) +
  labs(title="Plot of the Simulations", x="Simulation Mean")
```


Plot of the Simulations
density
0.0
0.2
0.4
3
4
5
6
7
Simulation Mean