Statstic Inference Week4 Part 1

Jie Yang

Part 1:

Overview

We 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.

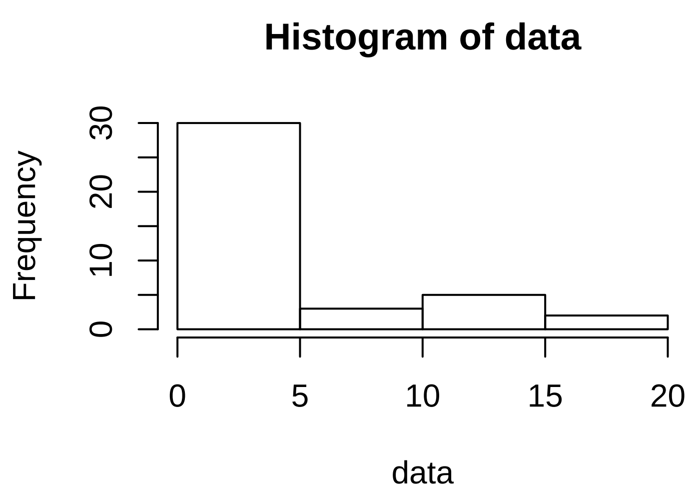
To set lambda to be 0.2

```
lambda<-0.2
```

Simulation

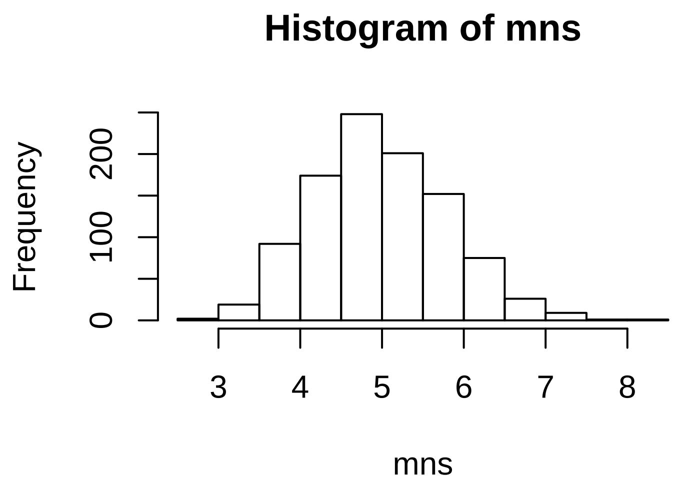
1. Create the random data set from exponetial distributions

```
n<-40
m<-1000
set.seed(1000)
data<-rexp(n, lambda)
#plot the histgram of the simulated dataset
hist(data)
```



2. Sample Mean versus Theoretical Mean:

```r
#Generating the sample mean
mns = NULL
for (i in 1 : m) mns = c(mns, mean(data<-rexp(n, lambda) ))
hist(mns)
```




```r
print("The sample means")
```

```
## [1] "The sample means"
```

```r
sample_mean<-mean(mns)
print(round(sample_mean),4)
```

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

```r
print("theoretical means should be")
```

```
## [1] "theoretical means should be"
```

```r
print(round(1/lambda),4)
```

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

```r
print("the difference is ")
```

```
## [1] "the difference is "
```

```r
print(round(abs(sample_mean-1/lambda)),5)
```

```
## [1] 0
```

Observations: The difference is very small.

3. Sample Variance versus Theoretical Variance

```r
sample_var<-var(mns)
print("this is the sample variance")
```

```
## [1] "this is the sample variance"
```

```r
print(sample_var)
```

```
## [1] 0.6541206
```

Given the theoretical standard deviation for geometric distribution is `1/lambda` the theoretical variance should be `std of pop/sqrt(n)` squared–> 1/lambda^2/m, as the code shows here

```
print("theoretical variance should be")
```

```
## [1] "theoretical variance should be"
```

```
print(round((1/lambda)^2/n,4))
```

```
## [1] 0.625
```

```
print("the difference is ")
```

```
## [1] "the difference is "
```

```
print(round(abs(sample_var-1/lambda^2/n),5))
```

```
## [1] 0.02912
```

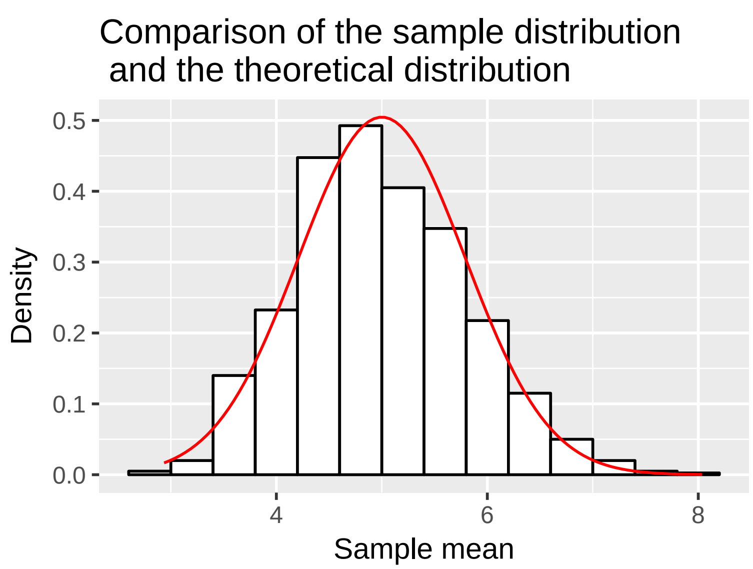
Observations: The difference is small.

4 Distribution:

According to the CLT, the averge should follow the normal distribution. As the shown before, the histgram is following the BELLshape.

As suggested, we'll focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials. Here we overlay the normal distribution from CTL

```
library(ggplot2)
data <- as.data.frame(mns)
ggplot(data, aes(x = mns)) +
  geom_histogram(binwidth = 0.4, color = 'black', fill = 'white', aes(y = ..density..)) +
  stat_function(fun = dnorm, color = 'red', args = list(mean = 5, sd = sqrt(0.625))) +  xlab('Sample mea
  ylab('Density') +  ggtitle('Comparison of the sample distribution\n and the theoretical distribution')
```



Conclusion

As one can see, the difference is small, sample follows the theoretical curve. This proves the CTL holds correct.