Simulation of the exponential distribution and the CLT

TehTY

Jul 20, 2015

Overview

In this project we 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. We will investigate the distribution of averages of 40 exponentials with a thousand simulations.

In this exercise, we will

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.

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

First, we will draw 1000 samples of size 40 from an Exp(1/0.2, 1/0.2) distribution. For each of these 1000 samples we will calculate the mean same as drawing a single sample of size 1000 from the corresponding sampling distribution = N(1/0.2, (1/0.2)/sqrt(40).

According to the CLT, we would expect that each single mean of those 1000 means approximately 1/lambda = 1/0.2 = 5. Then we caluculate the mean of 1000 sampled means. The output is to be expected very close to 5.

Let check with below codes.

```r
set.seed(1)

sample_means <- NULL
for(i in 1:1000) {
  sample_means <- c(sample_means, mean(rexp(40, 0.2)))
}
mean(sample_means)
```

```
## [1] 4.990025
```

The output is 4.99 which is very close to the mean of the theoretical distribution i.e 1/0.2 = 5.

2. How variable of the sample variance with the theoretical distribution

According to the CLT we would expect that the variance of the sample of the 1000 means approximately (1 / (0.2^2)) / 40 = 0.625.

```
var(sample_means)
```

```
## [1] 0.6111165
```

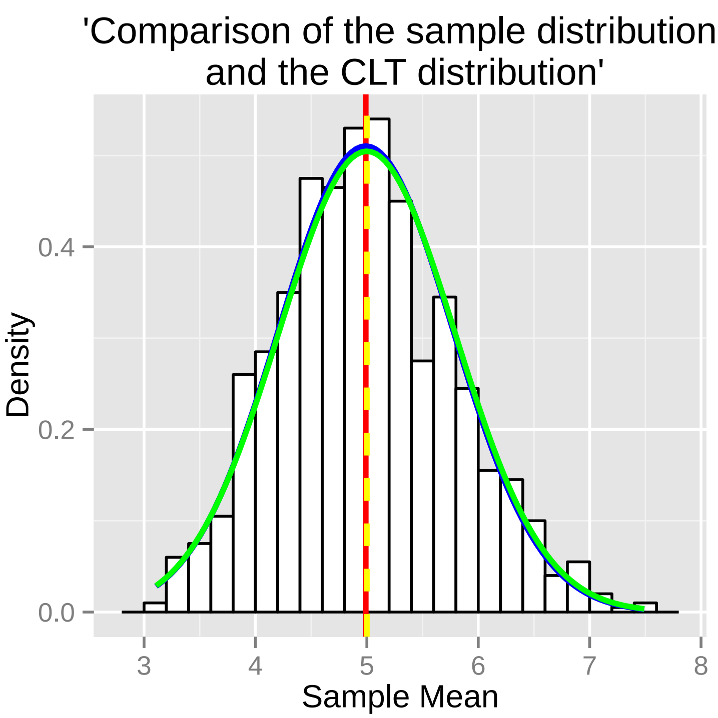
The output is 0.61 which is close to the variance of the theoretical distribution above.

3. Showing that the sample distribution is approximately normal

Let demonstrate that the sample distribution of the 1000 sampled means into histogram and overlay it with the density function from the theoretical sampling distribution which is N(1/0.2, (1/0.2)/sqrt(40)) distributed.

```
ActualMean<-mean(sample_means)
ActualSD<-sd(sample_means)
ActualVar<-var(sample_means)
TheoryMean<-1/0.2
TheorySD<-((1/0.2) * (1/sqrt(40)))
TheoryVar<-TheorySD^2

data<-data.frame(sample_means)
ggplot(data,aes(x=sample_means)) +
geom_histogram(binwidth = 0.2,fill="white",color="black",aes(y = ..density..))+
labs(title="'Comparison of the sample distribution\n and the CLT distribution'", x="Sample Mean", y="Der
geom_vline(xintercept=ActualMean,size=1.0, color="red") +
geom_vline(xintercept=TheoryMean,size=1.0, color="yellow",linetype = "dashed")  +
stat_function(fun=dnorm,args=list(mean=ActualMean, sd=ActualSD),size = 1.0, color = "blue")+
stat_function(fun=dnorm,args=list(mean=TheoryMean, sd=TheorySD),size = 1.0, color = "green")
```



Summary

In this analysis, it showed that the sampling distribution of the mean of an exponential distribution with n = 40 observations and lambda = 0.2 is very **close** to the theoretical normal curve N (1/0.2, (1/0.2)/sqrt(40)).