Simulation Exercise

HPSUN

2017-12-3

Overview

This project is about to test Central Limit Theorem with exponential distribution using R simulation.
I will set like a thousand simulations and compare the mean and variance from sample to the theoretical mean and variance. ## Simulations In the simulation, n is set to 40, lambda set to 0.2.

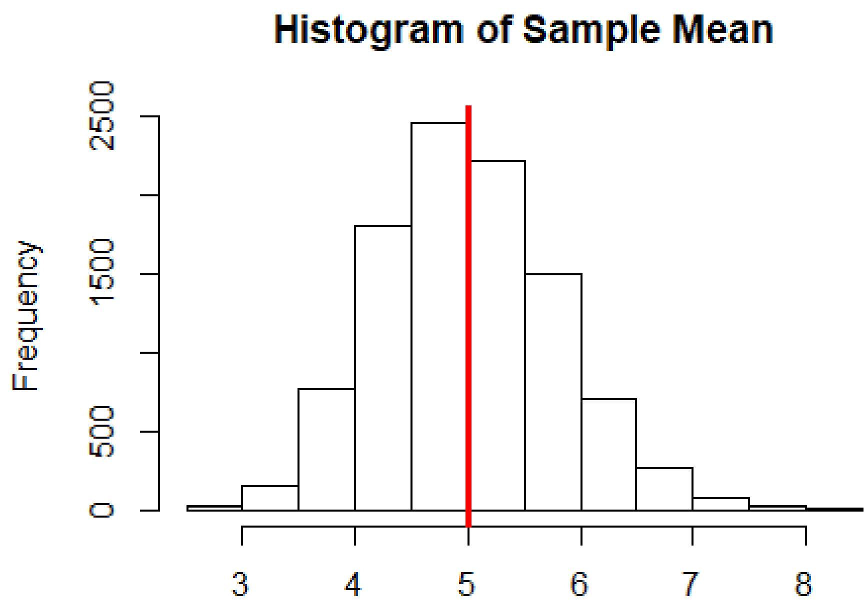
```
n <- 40
lambda <- 0.2
```

let's run 10000 simulations and store the means and variances of samples.

```
sample_mean <- NULL
for (i in 1:10000){
    sample_mean <- c(sample_mean, mean(rexp(n,lambda)))
}
```

Sample Mean versus Theoretical Mean

```
hist(sample_mean,main="Histogram of Sample Mean",xlab="")
abline(v=(lambda^-1),col="red",lwd=3)
```



The means of sample are showed as a histograph, the red line indicates the the theoretical mean. And the means of sample mean is very close to the theoretical mean, see below.

```
#theoretical mean
lambda^-1

## [1] 5

#mean of means
mean(sample_mean)

## [1] 4.99874

#difference
diff(c(lambda^-1,mean(sample_mean)))

## [1] -0.00126003
```

Sample Variance versus Theoretical Variance

```
#variance of sample mean
var(sample_mean)

## [1] 0.6219731

#theoretical variance
lambda^-2/n
```

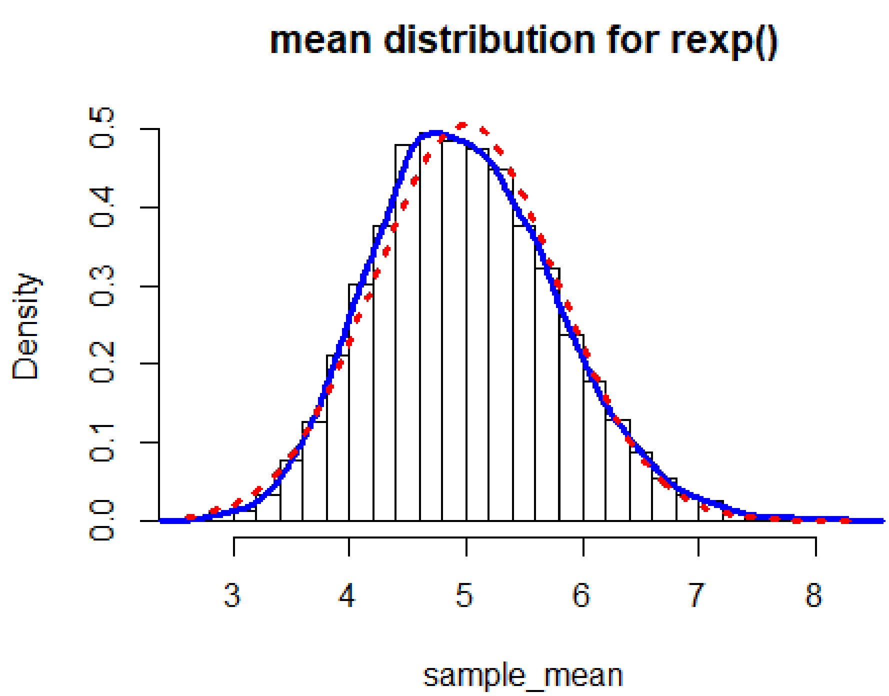
```
## [1] 0.625

#difference
diff(c(var(sample_mean),lambda^-2/n))

## [1] 0.003026949
```

The variace of sample mean is very close to the theoretical variance too. ## Distribution Let's compare the distribution of sample means and a normal distribution of sample mean and sample sd.

```
hist(sample_mean, prob=TRUE, main="mean distribution for rexp()", break
s=20)
lines(density(sample_mean), lwd=3, col="blue")
curve(dnorm(x,mean=mean(sample_mean), sd=sd(sample_mean)),add = T,col="
red", lwd=3, lty = "dotted", yaxt="n")
```



From the graph, we found the distribution of sample means is very similar to a normal distribution.That is what Central Limit Theorem told us: The averages are approximately normal, with distributions centered at the population mean and standard deviation equal to the standard error of the mean.So the 95% confidence interval of sampel means can be calcualate as follow:

```
mean(sample_mean) + c(-1,1)*qnorm(.975)*sqrt(var(sample_mean)/n)

## [1] 4.754338 5.243141
```