Exponential Distribution Investigation

liwenlong

Overview

In this project I will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution is simulated in R with rexp(n, lambda) where lambda = 0.2. I will investigate the distribution of averages of 40 exponentials with a thousand simulations.

```
lamada <- .2
sample_count <- 40
sim_count <- 1000
```

Simulations

Simulate random exponential distribution data,set seed to make the data reproducable.

```
set.seed(521) #521 means "I love you! :) "
expdata <- rexp(sample_count*sim_count,rate = lamada)
#Store the 1000 simulations into matrix.
exp_matrix <- matrix(expdata,sim_count,sample_count,byrow = T)
sampleMean <- apply(exp_matrix,1,mean)
```

Mean and Variance of the sample (Compare with the Theoretical metrics)

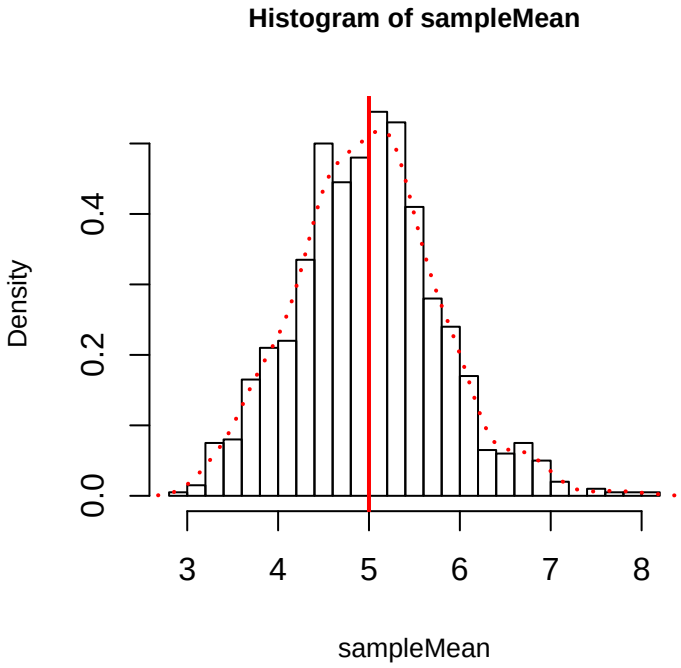
```
c(mean(sampleMean),var(sampleMean)) #Mean and Variance of Sample
```

```
## [1] 4.9843762 0.6217492
```

```
c(1/lamada,(1/lamada)^2/sample_count) #Theoretical Mean and Variance
```

```
## [1] 5.000 0.625
```

```
hist(sampleMean,20,prob = T, cex.main = .8,cex.lab=.8)
lines(density(sampleMean),col = "red", lwd = 2,lty = "dotted")
abline(v = 1/lamada,col = "red", lwd = 2)
```

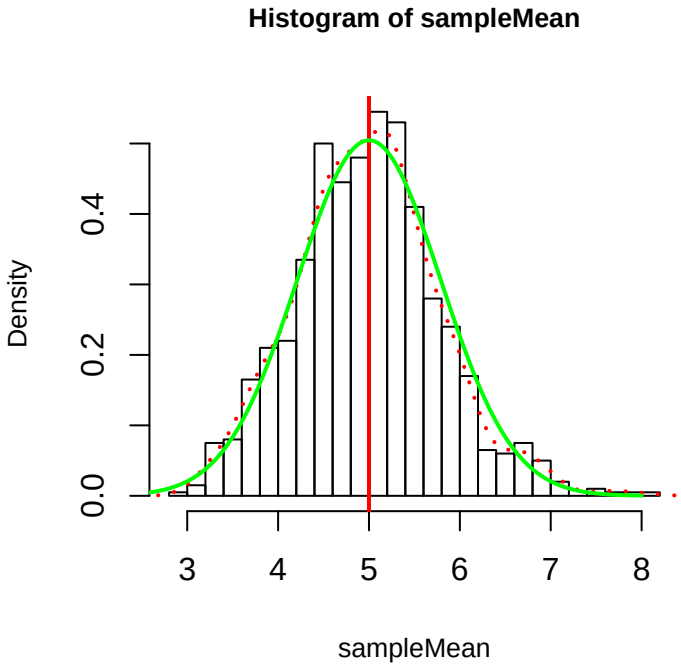


- The mean of the Sample [**4.9843762**] is close to the Theoretical Mean 1/lamada[**5**].
- The variance of the Sample [**0.6217492**] is also close to the Theoretical variance sˆ2/n [**0.625**]

Distribution

- Plot the distribution of the sample

```
hist(sampleMean,20,prob = T, cex.main = .8,cex.lab=.8)
lines(density(sampleMean),col = "red", lwd = 2,lty = "dotted")
abline(v = 1/lamada,col = "red", lwd = 2) #theoretical mean
#add the normal distribution line.
x<-seq(0,8,length=200)
y<-dnorm(x,mean=1/lamada, sd=1/lamada/sqrt(sample_count))
lines(x = x,y = y, col="green",lwd = 2)  #normal distribution line
```



- From the previous section , we've already known that the sample mean and variance is quite close to the theoretical mean and variance.
- And from the above figure we can found that the distribution of the sample is quite similar with the normal distirubtion with mean =1/lamada and sd = 1/lamada/sqrt(n) (the theoretical mean and sd)
- So, wen can say that the distribution is approximately normal.

SessionInfo

```
sessionInfo()
```

```
## R version 3.1.2 (2014-10-31)
## Platform: x86_64-apple-darwin13.4.0 (64-bit)
##
## locale:
## [1] zh_CN.UTF-8/zh_CN.UTF-8/zh_CN.UTF-8/C/zh_CN.UTF-8/zh_CN.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## loaded via a namespace (and not attached):
## [1] digest_0.6.4    evaluate_0.5.5  formatR_1.0     htmltools_0.2.6
## [5] knitr_1.8       rmarkdown_0.3.3 stringr_0.6.2   tools_3.1.2
## [9] yaml_2.1.13
```