Simulations In Statistics

Jerome Cordjotse

5/18/2020

Overview

The mean and standard deviation for an exponential distribution with lambda, λ, 0.2 is 5 since mean $= \frac{1}{\lambda}$ and standard deviation $= \sqrt(var) = \sqrt(\frac{1}{\lambda^2}) = \frac{1}{\lambda}$. In this experiment, the Central Limiting Theory, CTL would be used to figure out this mean through experiments and simulations. Detailed code and simulation explanations can be found here in the link to my repo.

Simulations

Here a thousand simulations where made, each retrieving the mean of 40 exponentials. This distribution is stored in a 42×1000 matrix with columns storing the sample of size 40 and the last rows, 41 & 42, sotres the mean of the samples from row 1 to 40 and their respective variance respectively. For a thousand simulations the column number becomes 1000.

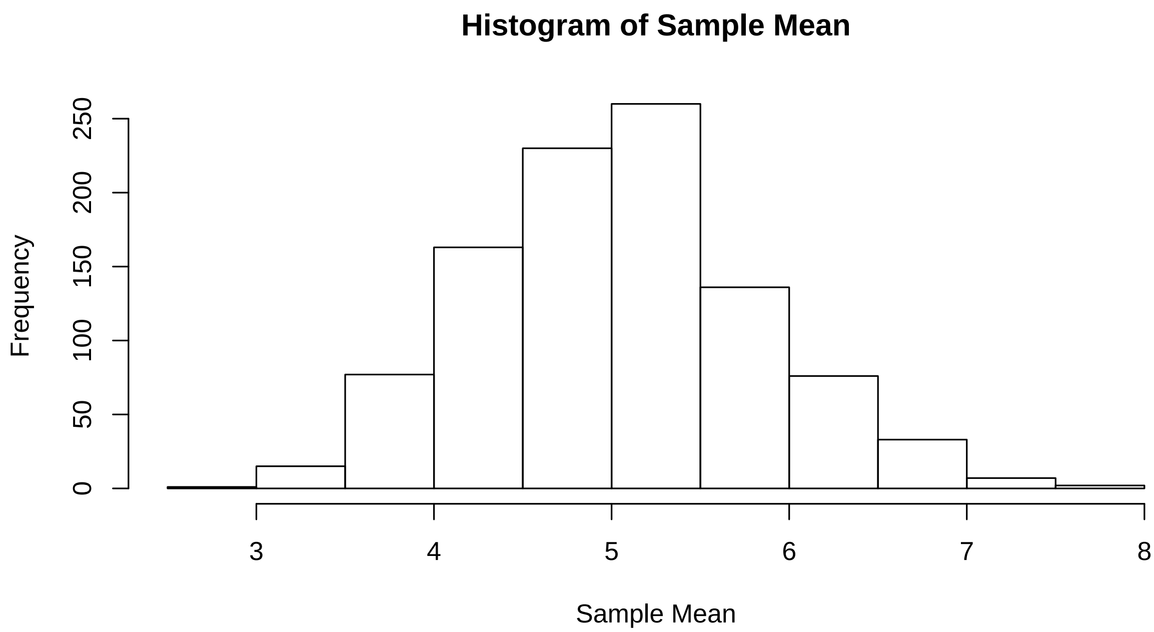
```
lambda = 0.2
mean <- 1/lambda           # Theoritical Mean
sd <- 1/lambda             # Theoritical Standard Deviation
var <- sd^2                # and variance
## To make a 1000 simulations with a for loop
sim_mat <- NULL
for (i in 1 : 1000){
  expObs <- rexp(40,0.2)
  sim_mat <- cbind(          # Adding simulations by
      sim_mat,               # attaching new columns
        c(
          expObs,            # 40 exponentials
          mean(expObs),      # adding their mean to at row 41
          var(expObs)        # adding their variance at row 42
          )
        )
}

sampleMean <- sim_mat[41,]
sample_Variance_Distribution <- sim_mat[42,]
summary(sampleMean)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.796   4.487   5.028   5.034   5.509   7.821
```

Sample Mean vs Theoretical Mean

Knowing that the sample mean is a good estimator of the population mean. So ploting the histogram of means shows this as it balances around 5.



This can be confirmed as the average of sample mean is below which is close to 5.

```r
mean(sampleMean)
```

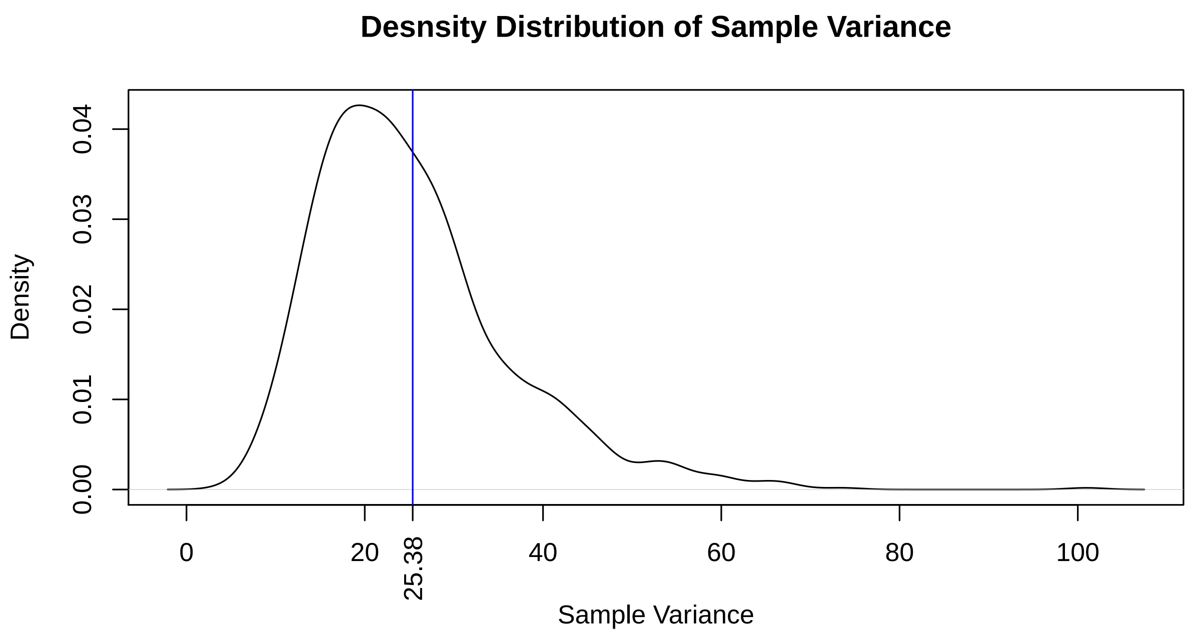
```
## [1] 5.033948
```

Sample Variance vs Theoretical Variance

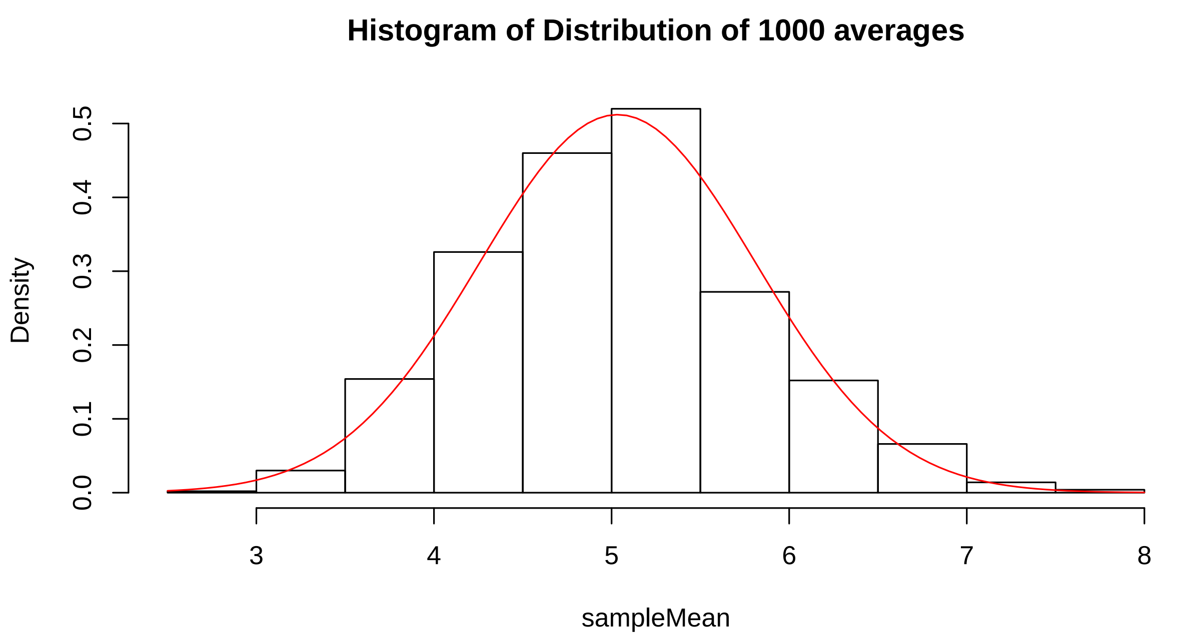
As seen with the distribution of sample means, the sample variance estimates the population variance. Taking average yields

```r
mean(sample_Variance_Distribution)
```

```
## [1] 25.37907
```



Distribution



The CLT states that the distribution of averages of iid variables becomes that of a standard normal as the sample size increases. It can be seen than the Figure above shows an approximation to a normal as the bell-curve shape overlays it.