Exponential Distribution Simulation Exercise

Statistical Inference

Linda Angulo Lopez

01/01/2021

`key words:`Asymptotic, The Central Limit Theorem, Sample Mean vs Theoretical Mean, Sample Variance vs Theoretical Variance, Distributions, Quantiles, Statistical Inference.

Introduction

The project consists of two parts (i) a simulation exercise and a (ii) basic inferential data analysis, the former is presented here. Asymptotics form the basis for frequency interpretation of probabilities, where the behavior of statistics depends on the sample size or some other relevant quantity of limits to infinity or to zero, the Swiss Army Knives of Statistics.

An exponential distribution is the probability distribution that describes the time between events in a Poisson process, i.e. a process in which events occur continuously and independently at a constant average rate. A thousand simulations, our samples, each of $n = 40$ were run to investigate if `the sampling distribution of exponential distribution has a normal distribution with a mean that matches the population mean and a variance that matches the theoretical result`. It was found that for the exponential distribution generated the true mean that matches the population mean and a variance that matches the theoretical result in addition the distributions have similar means at the quantiles: 5%, 25%, 50% , 75% and 95%.

```
#R4.0 Environmental Set for:
library(knitr) # creating a pdf document ;
library(ggplot2) # making plots
library(dplyr) #exploring data
library(DataExplorer) # creating reports
```

Exponential Distribution Simulation:

A discreet case of the CLT, 1000 trials were simulated in with the R function `rexp(n, lambda)` where lambda, λ , was set to 0.2, this is the rate parameter. The mean & the standard deviation of an exponential distribution is $1/\lambda$, for $n = 40$. Data was then explored with glimpse() and summary().

create, explore & plot data A pseudo-random number generator was used so that the exercise can be replicated,data was explored with the dplyr and DataExplorer libraries from CRAN.

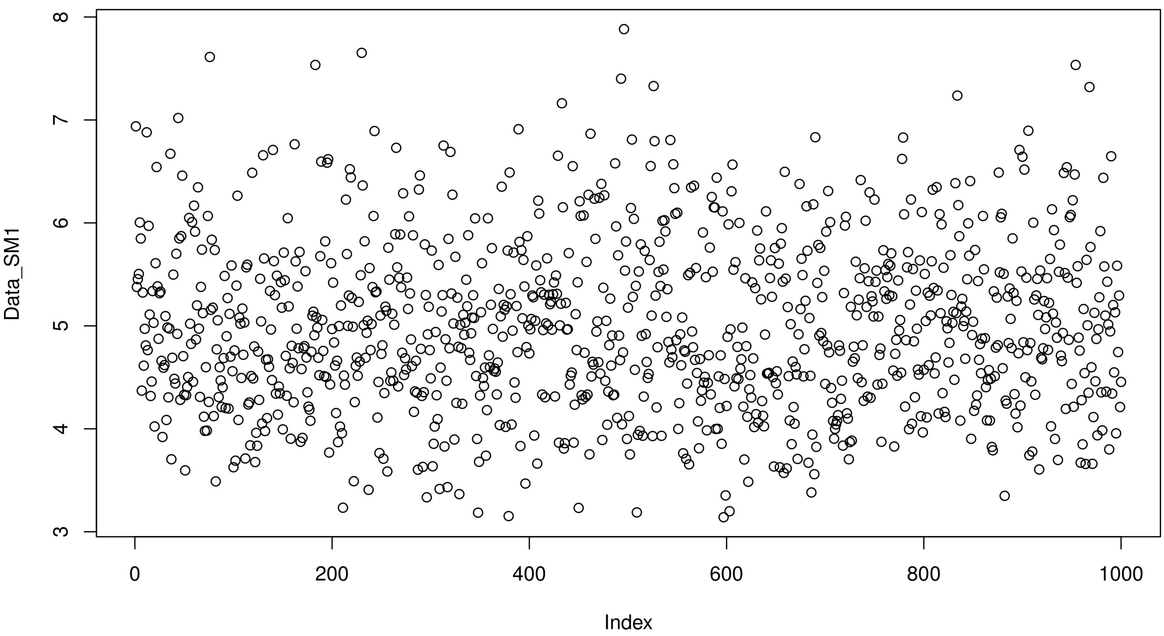
Here is a glimpse of the data:Data_SM1, and a basic data summary:

```
set.seed(42); lambda <- 0.2; nosim <- 1000; n <- 40 ;
SimulationMatrix_1 <- matrix(rexp(nosim * n, rate = lambda), nosim);
Data_SM1 <- apply(SimulationMatrix_1, 1, mean); glimpse(Data_SM1); summary(Data_SM1)
```

```
##  num [1:1000] 6.94 5.38 5.45 5.5 6 ...
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.141   4.406   4.919   4.987   5.504   7.882
```

Here is a plot of the data: Data_SM1

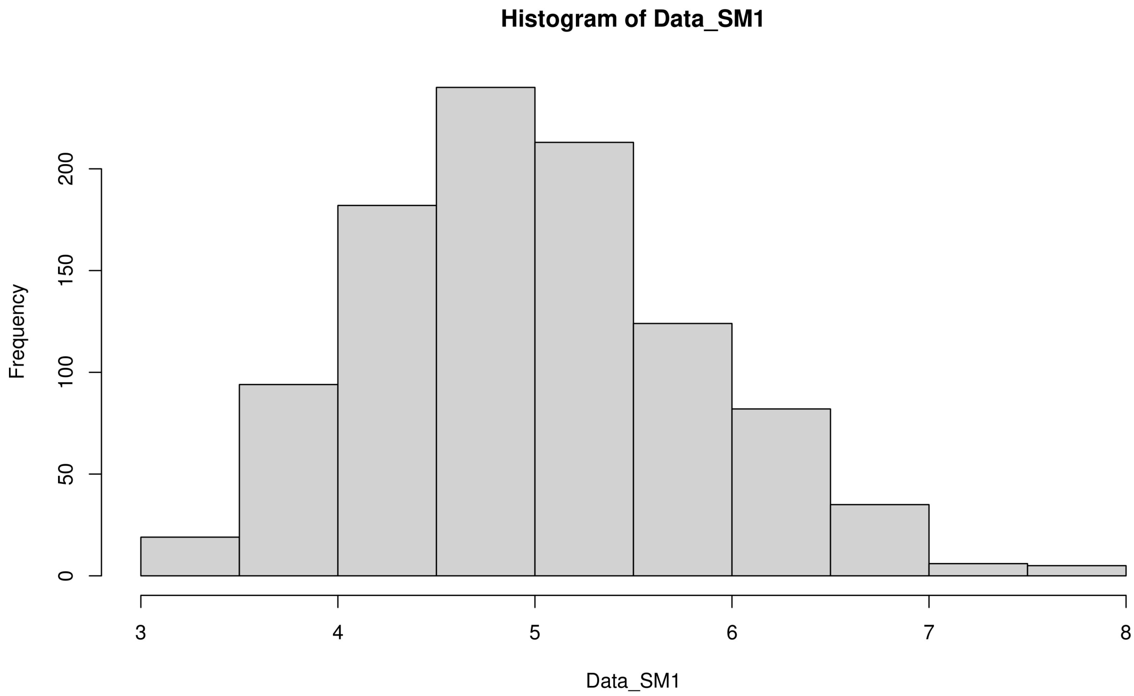


Mean vs True Mean We can get the mean of the data with mean() and the theoretical mean is $1/\lambda$. We see that The mean of the data, the true mean is close to the theoretical mean:

```
## [1] 4.986508
```

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

After taking a peak at the report generated with create_report(), we see that the histogram has a bell shaped curve it looks like a normal distribution.



True variance vs Theoretical variance We can get the variance of the data by using the result of the sd() function squared and the theoretical value is $\sigma 2n = 1/\lambda^2 n$. We see that The variance of the data is close to the theoretical variance:

```
Variance_Data <- sd(Data_SM1)^2 ; Variance_Data; TheoVar <- 1/((lambda^2)*n); TheoVar
```

```
## [1] 0.6793521
```

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

Comparing Distributions The easiest way to compare the data vs theoretical distributions would be to use the qnorn() function to get the normal distributions and quantile() to get the quantiles of the data then compare the true vs theoretical values. We see that the distribution of quantiles, for the data is close to the theoretical quantiles:

```
Quantiles_DatavsTheo <-c(0.05, 0.25, 0.5, 0.75, 0.95); quantile(Data_SM1,  Quantiles_DatavsTheo);
```

```
##       5%      25%      50%      75%      95%
## 3.750774 4.406170 4.919282 5.504413 6.485176
```

```
qnorm(Quantiles_DatavsTheo, mean = mean(Data_SM1), sd = sd(Data_SM1))
```

```
## [1] 3.630774 4.430575 4.986508 5.542442 6.342243
```

After-note:

- This is an R Markdown document,created as a submission for the Statistical Inference Course by Johns Hopkins University on Coursera.
- The the exercise can be replicated Hidden code is available here.