Inference Statistics Course Project 3

Mathias Barat

25/09/2020

PART 1 - SIMULATION EXERCISE

Overview

Investigate the exponential distribution in R and compare it with the Central limit theorem

Simulation

```
# Set-up my inputs:
lambda <- 0.2
mu <- 1/lambda
sd <- 1/lambda
n <- 40
n_simu <- 1000


# data simulation
set.seed(12345)
matrix_sample <- matrix(rexp(n*n_simu,lambda), n_simu, n)
mean_sample <- rowMeans(matrix_sample)


# summary
summary(mean_sample)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.032   4.424   4.938   4.972   5.492   8.380
```

Sample Mean

```
mua <- mean(mean_sample)

deltamu <- mua - mu
deltamu
```

```
## [1] -0.02802804
```

```r
round (deltamu/mu*100,1)
```

```
## [1] -0.6
```

The sample mean (4.971972) is 0.6% lower than the theoretical mean (5). The difference is really minimal.

Sample Variance

```r
Var_theo <- sd^2 / n
Var_sample <- var(mean_sample)

Var_theo
```

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

```r
Var_sample
```

```
## [1] 0.6157926
```

```r
deltavar = Var_sample - Var_theo
var_perc <- round(deltavar/Var_theo*100,2)
```

The Theoretical Variance (0.625) and is -1.47% lower than the Sample Variance is (0.6157926).

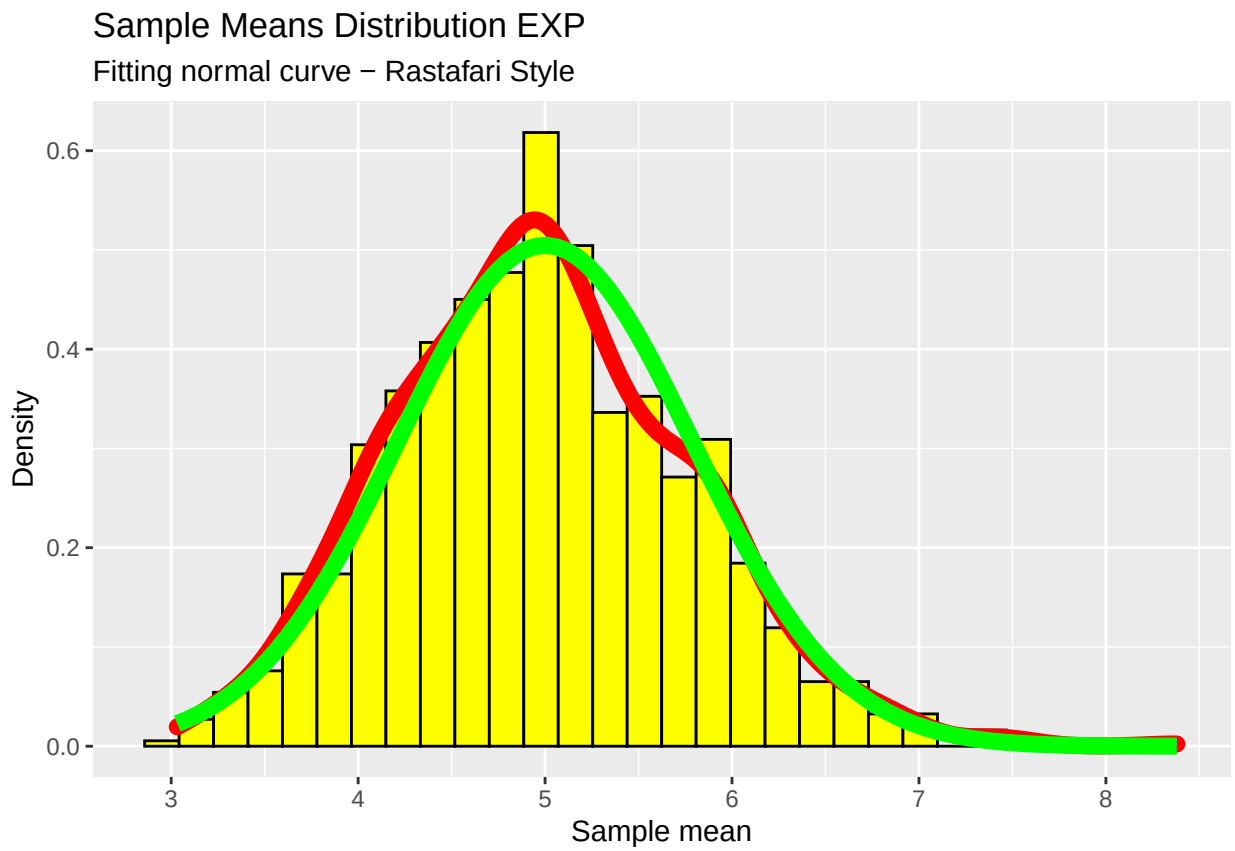
Distribution

```r
library(ggplot2)

sd <- sqrt(Var_theo)

ggplot(as.data.frame(mean_sample), aes(as.data.frame(mean_sample)[,1]))+
        geom_histogram(aes(y=..density..), position="identity", fill="yellow", col="black")+
        geom_density(colour="red", size=3)+
        stat_function(fun = dnorm, colour = "green", args = list(mean = mu, sd = sd), size=3)+
        ggtitle ("Sample Means Distribution EXP")+
        labs(subtitle = "Fitting normal curve - Rastafari Style")+
        xlab("Sample mean")+
        ylab("Density")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



As highlighted by the plot, we can see a small discrepancy between the sample distribution and the theoretical one.

Conclusion

We can conclude that the simulation performed with the rexp R function has produced a random dataset globally normal.