Simulation Excercise: Exponential Distribution

Andrés Hidalgo Vargas

14/7/2020

Overview

Simulations

First we set a seed so we can have a constant generation of pseudorandom numbers, making the plots replicable.

```
set.seed(3081997)
```

After the seed is set, we define our variables. The exponential distribution depends on the rate `lambda`, the sample size will be set to 40.

```
lambda <- 0.2
n <- 40
means = NULL
```

The simulation is run to generate 1000 runs of 40 random exponential numbers with lambda 0.2. The `mean` is taken from each run and stored.

```
for (i in 1:1000){
  means = c(means, mean(rexp(n,lambda)))
}
```

A one sample with 1000 observations is generated to later illustrate the distribution behaviour.

```
set.seed(3081997)
onesample <- rexp(1000,0.2)
```

Sample vs theoretical mean

The theoretical mean is defined by 1/lambda. Therefore, the theoretical mean for the simulated data is

```
1/lambda
```

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

The sample mean is

```r
mean(means)
```

```
## [1] 4.998339
```

4.99 is very close to 5.

Sample vs theoretical variance

The theoretical variance is defined by (1/lambda^2)/n. Therefore, the theoretical variance for the simulated data is

```r
(1/lambda^2)/n
```

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

The sample variance is

```r
var(means)
```

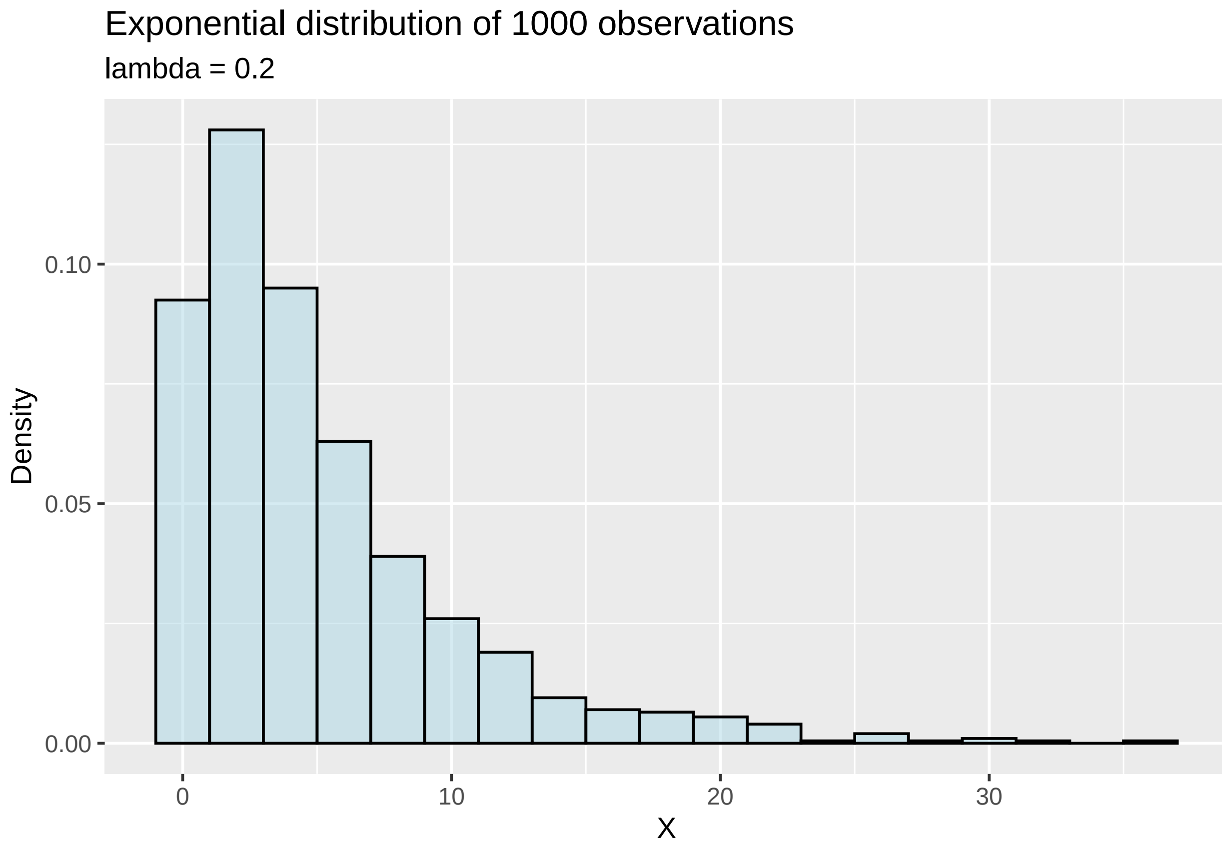
```
## [1] 0.6704198
```

0.67 is close to the theoretical variance 0.625

Distribution

To explain how the distibution appears normal, first we need to see how does the original exponential distribution looks like

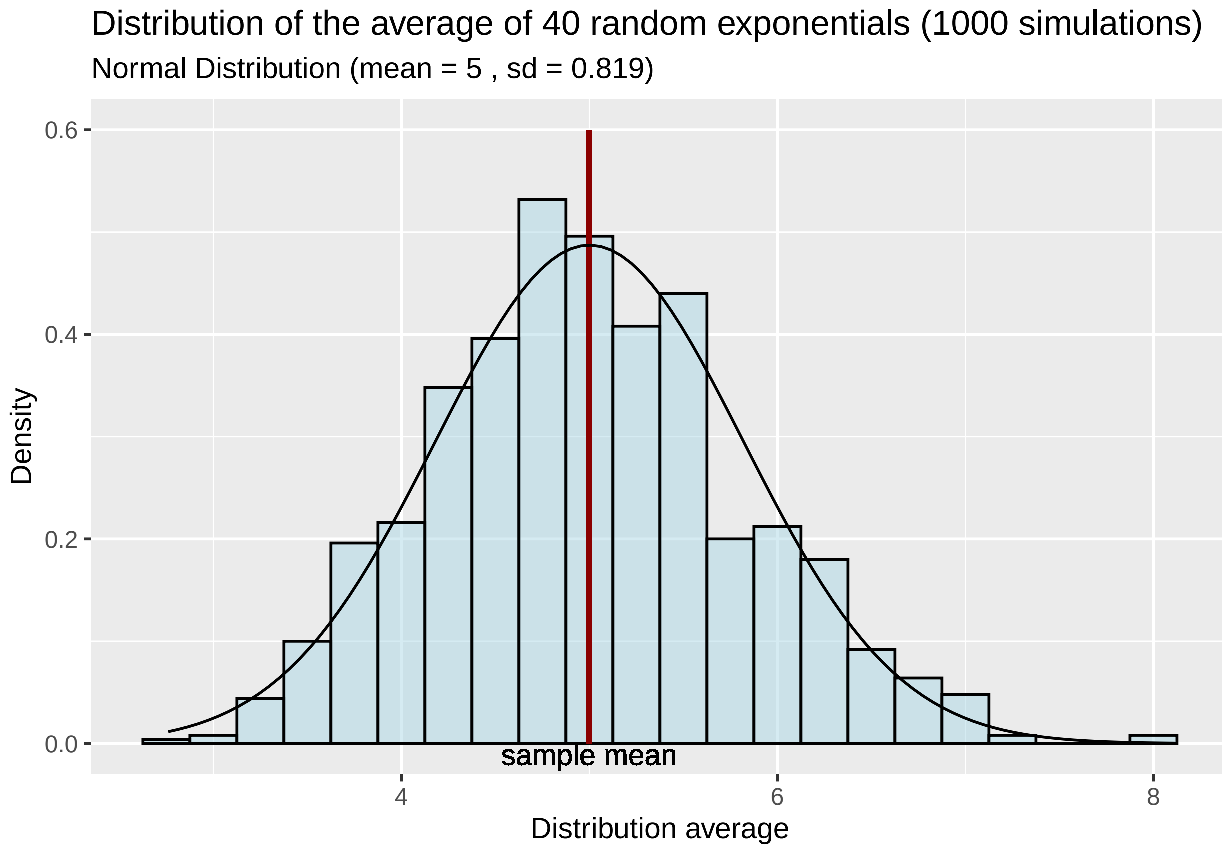
```r
library(ggplot2)
ggplot(data.frame(onesample), aes(onesample)) +
  geom_histogram(aes(y=..density..),
                 binwidth=2,col="black",
                 fill = "lightblue" ,
                 alpha = 0.5) +
  xlab("X") +
  ylab("Density") +
  labs(title = "Exponential distribution of 1000 observations", subtitle = "lambda = 0.2")
```



This distribution looks similar to the probability density function graph

Now, the distribution of the averages of 40 random exponentials and 1000 simulations will be plotted.

```
ggplot(data.frame(means), aes(x = means)) +
  geom_histogram(aes(y = ..density..),
                 binwidth=.25,
                 col="black",fill = "lightblue",
                 alpha = 0.5) +
  geom_segment(aes(x = mean(means), y = 0, xend = mean(means), yend = 0.6),
               color = "darkred", size=1) +
  stat_function(fun=dnorm,
                args = list(mean=5,sd=sd(means))) +
  geom_text(x = mean(means), label = "\nsample mean", y= 0.01) +
  xlab("Distribution average") +
  ylab("Density") +
  labs(title = "Distribution of the average of 40 random exponentials (1000 simulations)",
       subtitle = "Normal Distribution (mean = 5 , sd = 0.819)")
```



As the CLT says, with sufficient random samples, the distribution of the sample means will be approximately normally distributed.