Distribution of exponential sample mean

Luis Talavera

July 25th 2022

Synopsis

The exponential distribution is a continuous probability distribution that often concerns the amount of time until some specific event happens. The objective of this project is to build a distribution by sampling 40 exponentials, calcualte its mean, repeat it one thousand times and compare the distribution obtained with the Central Limit Theorem.

Simulations

First, we will build the distribution of averages of 40 exponentials using one thousand simulations. Lambda will be set to 0.2 for all the simulations.

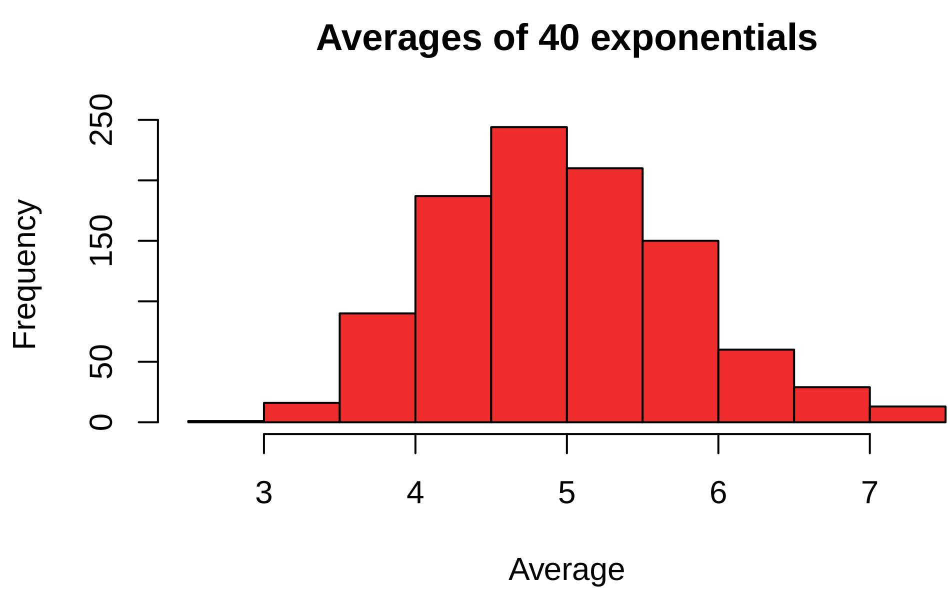

Figure 1: Distribution of exponential sample mean

The mean of exponential distribution is given by $\mu = \frac{1}{\lambda}$, which will be the same as our expected mean therefore our *theorical mean* will be

[1] 5

Now, we calculate the *sample mean* following the formula

$$\overline{x} = \frac{\Sigma x_i}{n}$$

[1] 4.976

We can see the difference between theorical mean and sample mean is too small, they are very close to each other.

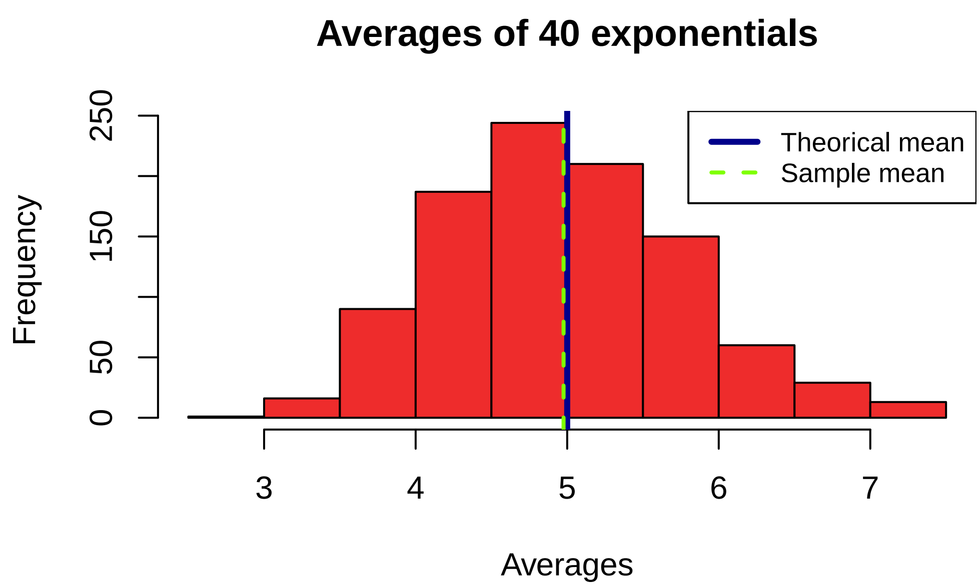

Figure 2: Figure 2. Theorical mean vs sample mean

The variance of exponential distribution is given by $\sigma^2 = \frac{1}{\lambda^2}$, and the variance of a sample mean is

$$Var(\overline{X}) = \frac{\sigma^2}{n}$$

hence our *theorical variance* will be

[1] 0.025

After that, we calculate the *sample variance* by the following formula

$$s^2 = \frac{\Sigma(x_i - \overline{x})^2}{n-1}$$

[1] 0.64

We can see the difference between theorical variance and sample variance is also too small.

Up to this point, we can see that the distribution obtained is approximately normal. And the central limit theorem proves this to be true since the theorem states that the distribution of sample means becomes a normal distribution as the sample size increases.

If we draw a normal curve and overlay it on our distribution, we can observe that they match almost perfectly, giving us visual proof that the distribution obtained is approximately normal.

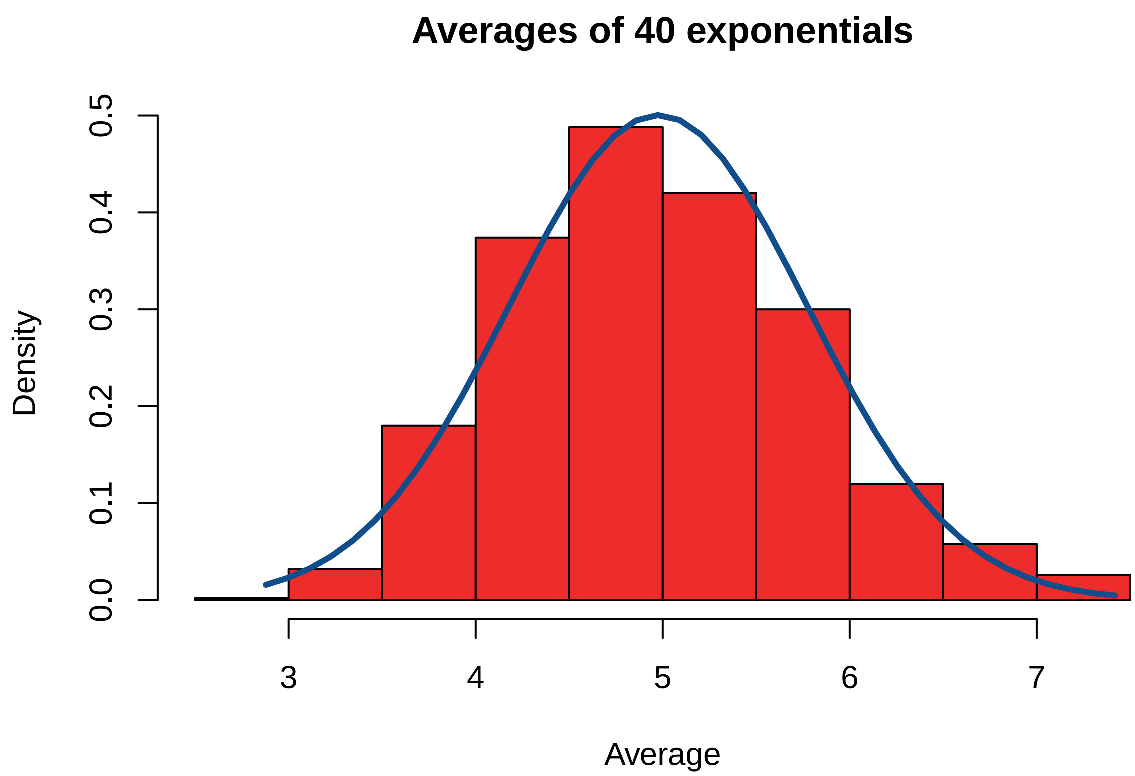

Figure 3: Figure 3. Normal curve over histogram

Finally I would like to highlight the fact that we had to use 1000 simulations because the exponential distribution is extremely skewed to the left, this implies a larger sample size so that the distribution of sample means approximates the normal distribution.

Appendix

Theorical mean

```
tmean <- 1/lambda
```

Sample mean

```
smean <- round(sum(mns)/n, 3) # 1000
```

Theorical variance

```
tvar <- round(1/(lambda^2)/n,3)
```

Sample variance

```
svar <- round(sum((mns-smean)^2)/(n-1), 3)
```

Figure 1

```
hist(mns,
     main="Averages of 40 exponentials",
     xlab = "Average",
     col="firebrick2")
```

Figure 2

```
hist(mns,
     main="Averages of 40 exponentials",
     xlab = "Averages",
     col="firebrick2")

abline(v=c(tmean,smean),
       col = c("blue4","chartreuse"),
       lty=c(1,2),
       lwd=c(3,2))

legend("topright",
       legend = c("Theorical mean", "Sample mean"),
       col = c("blue4","chartreuse"),
       lty = c(1,2),
       lwd=c(3,2),
       cex = 0.8)
```

Figure 3

```
h <- hist(mns,
       main="Averages of 40 exponentials",
       xlab = "Average",
       col="firebrick2")
```

```r
xfit <- seq(min(mns), max(mns), length = 40)
yfit <- dnorm(xfit, mean = mean(mns), sd = sd(mns))
yfit <- yfit * diff(h$mids[1:2]) * length(mns)

lines(xfit, yfit, col = "dodgerblue4", lwd = 3)
```