Analysis of the Exponential Distribution in R

19/06/2019

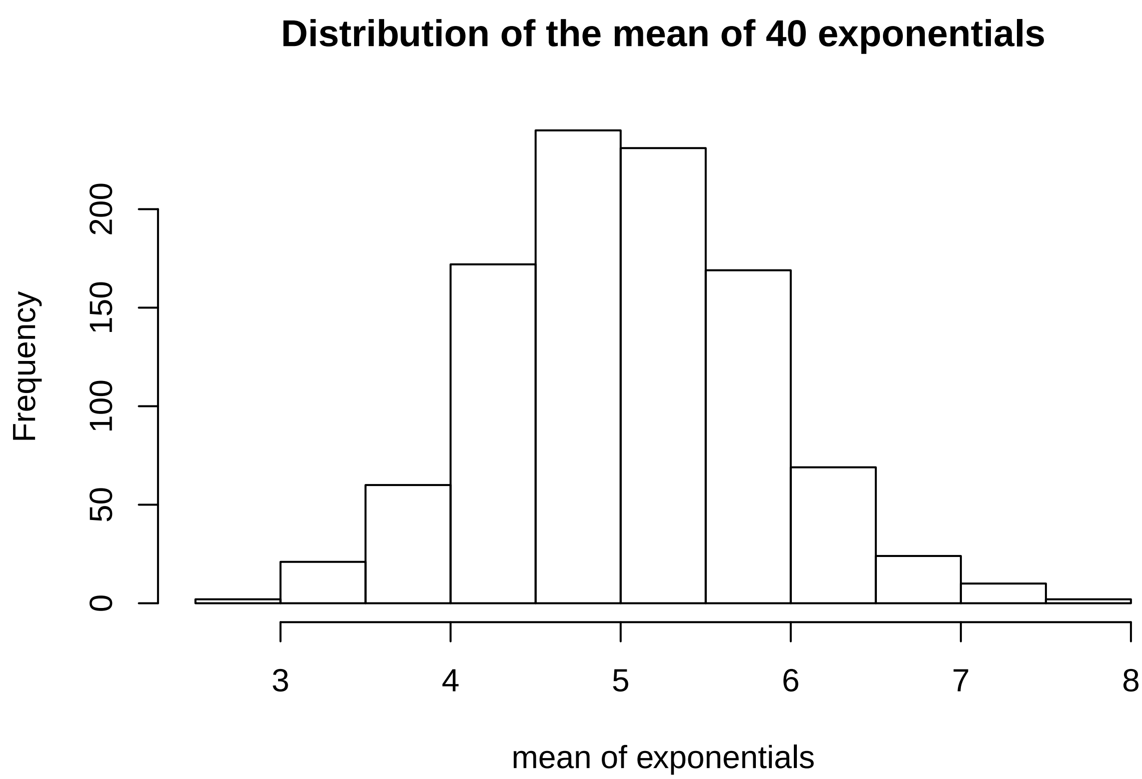
Synopsis

In this document we explore the proprieties of the exponential distribution in R and compare it with the Central Limit Theorem. Our goal is to show how the mean of 40 exponentials can be approximated by a normal distribution.

Part 1: Simulation Exercise Instructions

The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. We set lambda = 0.2 for all of the simulations.

```
lambda <- 0.2
pop.mean <- 1/lambda
pop.sd <- 1/lambda
n <- 40
set.seed(2019)
sim <- NULL
for (i in 1:1000) sim = c(sim, mean(rexp(n, lambda)))
hist(sim, main = "Distribution of the mean of 40 exponentials", xlab = "mean of exponentials")
```



At first glance, the shape of our histogram looks like a bell curve.

Sample mean

Let's calculate the mean of our sample.

```
msim <- mean(sim)
print(msim)
```

```
## [1] 5.038755
```

The mean of our sample is equal to 5.04, compared with 5 for the population. The expected value of the sample mean is quite the same as the expected value of the population it estimates. We can say it is unbiased.

Variability of the sample

We will now take a look at the variance of the mean distribution. Theoretically, we know the formula of the variance of the mean distribution to be sigma/sqrt(n).
We can also compute the variance directly from our data using the apply function.
Let's calculate the variance both ways and check if they match.

```
vsim.th <- (1/lambda)^2/n
vsim.pr <- (sd(apply(matrix(rexp(n*1000,lambda),1000),1,mean)))^2
cat(" Theoretical Value :", vsim.th,"\n","Actual value :", vsim.pr)
```

```
##  Theoretical Value : 0.625
##  Actual value : 0.6530026
```

The variance of our sample is equal to 0.653, compared with 0.625 in theory. It's close, but it's not equal. Maybe we don't have enough trials. Let's calculate with 1 000 000 trials instead of 1 000.

```
vsim.pr2 <- (sd(apply(matrix(rexp(n*1000000,lambda),1000000),1,mean)))^2
cat(" Theoretical Value :", vsim.th,"\n","Actual value :", vsim.pr2)
```

```
##  Theoretical Value : 0.625
##  Actual value : 0.625586
```

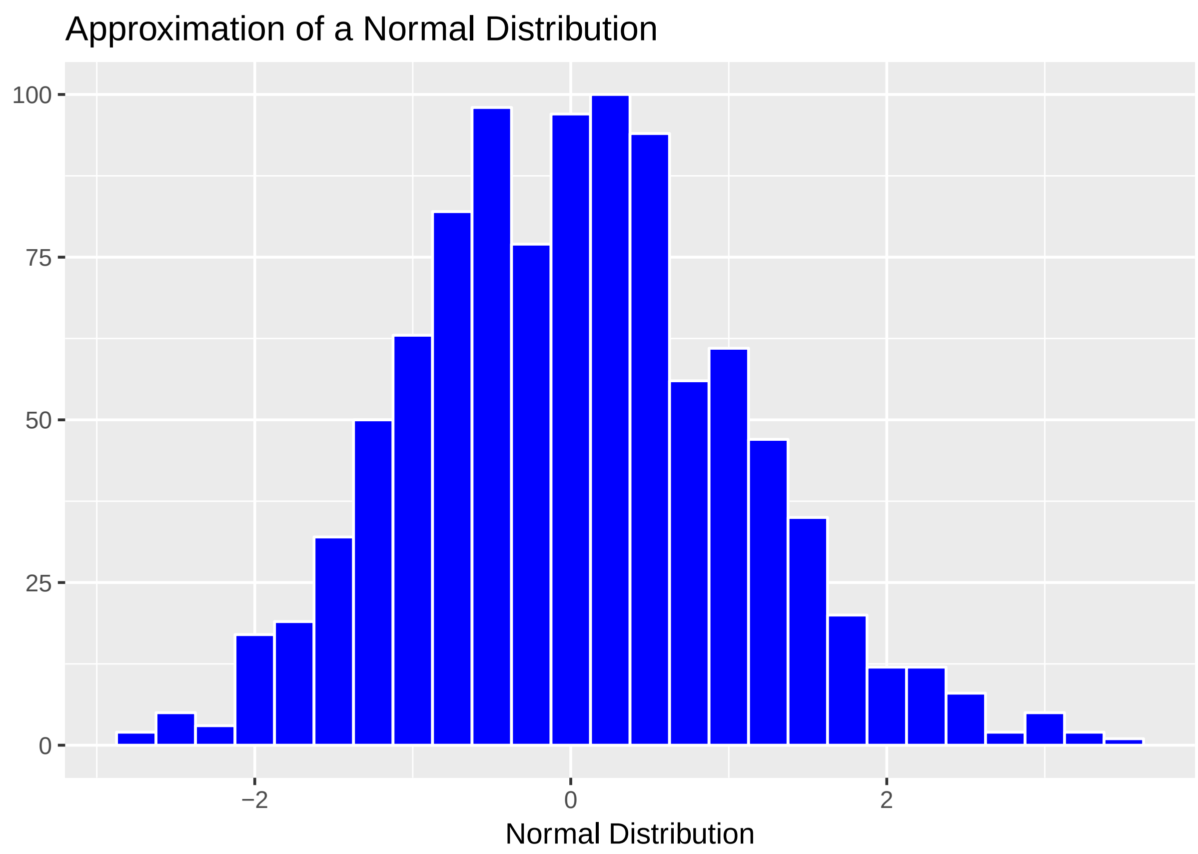
This time we obtain 0.626, which is much closer to the theoretical value.
That last calculation took a while, but at least we can see that the more trials we launch, the closer we get to the theoretical value.

Distribution

Now, let's use the Central Limit Theorem to check if we can approximate the values with a Normal distribution. According to this theorem, the distribution of averages of iid random variables becomes that of a standard normal as the sample size increases.
Since we have iid variables and our sample size is large enough, we can apply the Central Limit Theorem : let's subtract the mean (1/lambda) and divide by the square root of the variance/n.

```r
mns = NULL
for (i in 1:1000) mns = c(mns, mean(rexp(40, lambda)))
clt <- (mns-1/lambda)/((1/lambda)/sqrt(40))

qplot(clt, binwidth = 0.25, fill = I("blue"), colour = I("white"), main = "Approximation of a Normal Di
```



In the figure above, we can observe that our approximation follows a standard bell curve centered on 0 with an approximate standard deviation of 1.