Analysis Exponential Function

Mathieu Wauters

11 februari 2016

In this report, I will show using a simulation approach what the relation is between the exponential distribution and the Central Limit Theorem (CLT) for a sufficiently large sample size `n`. First, I will cover some general parameters and settings for the simulations, after which a comparison of the sample mean and sample variances to the theoretical mean and variance will be shown. Finally, the relation to the normal distribution will be covered. I wish you, the reader, an interesting and enjoyable reading experience.

General simulation settings and data generation

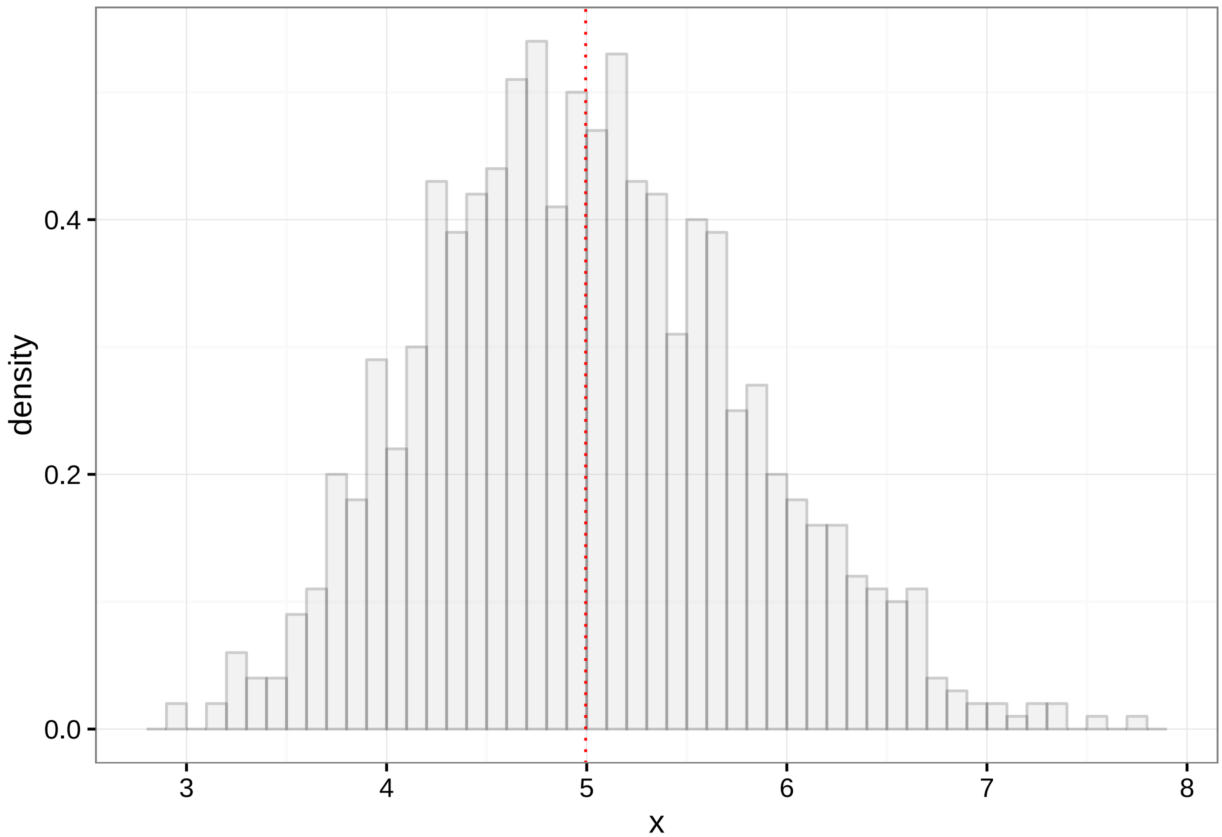
First, we need to specify some general settings. We will load `{r,eval=F} library(ggplot2)` for making the plots. Additionally, 1000 simulations will be run of 40 exponentials characterized by the parameter $\lambda = 0.2$. In order to enhance reproducibility, a random seed value is set.

```
library(ggplot2)
nbr_simulations<-1000
nbr_exponentials<-40
lambda<-0.2
set.seed(2534)
```

Next, we generate some data. I like to have access to all data, which is why I store the generated data in a (40×1000) matrix instead of retaining merely the sample means.

```
value_matrix<-matrix(ncol=nbr_exponentials,nrow=nbr_simulations)
for(i in 1:nbr_simulations){
  value_matrix[i,]<-rexp(nbr_exponentials,lambda)
}
sample.means<-data.frame(x=apply(value_matrix,1,mean))
```

Before diving into the Mean and Variance, let us take a look at the histogram. Before running the numbers, it is good practice to explore the data in a visual way. The graph below shows the histogram, as well as a vertical red line which is centered at the sample mean. This is the subject of the following section.



Mean

Theoretically, we expect the mean of the distribution to be equal to $\frac{1}{\lambda}$ which in our case would be equal to 5 ($\frac{1}{0.2}$). The *sample* mean however is equal to NA, which is pretty close to $\mu = 5$.

```
mean(sample.means$x)
```

```
## [1] 4.994606
```

Variance

The theoretical variance is equal to $\frac{\sigma^2}{n}$. With $n = 40$ this yields `0.625`. Based on the simulations, the variance of the sample means is equal to:

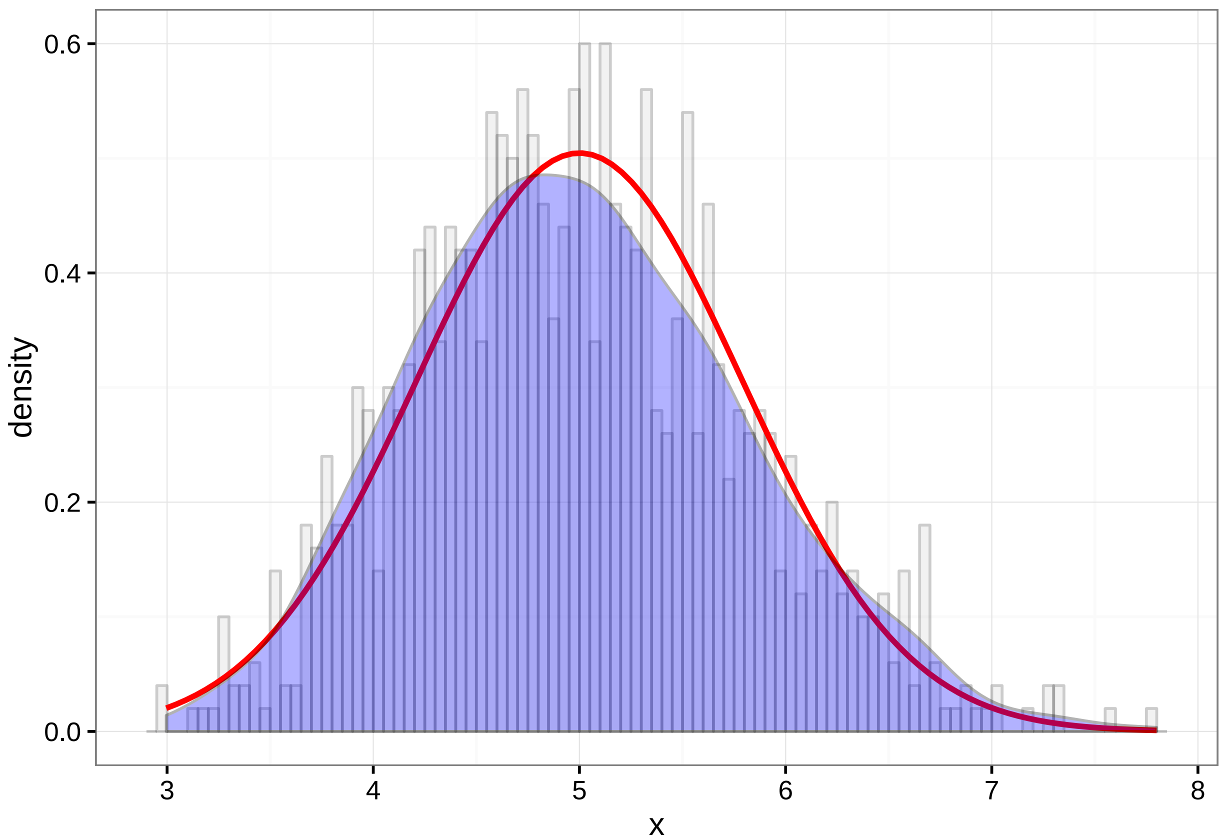
```
var(sample.means$x)
```

```
## [1] 0.6328144
```

CLT comparison

The CLT tells us that the distribution of averages of iid variables becomes that of a normal distribution as the sample size increases. From this statement follows that the sample means are normally distributed with a mean μ and variance $\frac{\sigma^2}{n}$. Note that the variance is the squared standard error of the estimate. In the plot below, three types of information are provided:

- In the background, the histogram (which we saw earlier in this report) is provided. Note that the y-axis represents the density and not the frequency of the counts of the observations.
- The curve in blue depicts the density of the sample means.
- The red line shows the normal distribution $N(\mu, \frac{\sigma^2}{n})$ with $\mu = \frac{1}{\lambda}$ and $\sigma = \frac{1}{\lambda}$.



We can also construct a confidence interval to check how many of the 1000 sample means fall within (or without this interval). A 95% confidence interval is given by $[\bar{X} - 1.96\frac{\sigma}{\sqrt{(n)}}, \bar{X} + 1.96\frac{\sigma}{\sqrt{(n)}}]$. Let us count the number of times our sample means do *not* fall within this interval.

```
confint<-mean(sample.means$x)+c(-1,1)*1.96*sigma/sqrt(nbr_exponentials)
(length(which(sample.means$x<confint[1]))+length(which(sample.means$x>confint[2])))/nbr_simulations
```

```
## [1] 0.052
```

This value is very close to the expected value of 0.05 and reinforces our belief in the CLT for sample means of an exponential distribution that are iid.