Statistical Inference CP1

Dan K. Hansen

26 May 2016

Overview:

We will investigate the exponential distribution in R and compare it with the Central Limit Theorem (CLT).

Assumptions:

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 *(theoMean)*
- The standard deviation is also 1/lambda. *(stdDev)*
- Set lambda = 0.2 for all of the simulations.
- Investigate the distribution of averages of 40 exponentials. *(n)*
- We will need to do a thousand simulations. *(iterates)*

Tasks to do

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

Data Processing

First we set the constants and variables

```
lambda <- 0.2
iterates <- 1000
n <- 40
theoMean <- 1/lambda
stdDev <- 1/lambda
```

Create an exponential distribution: *rexp(n,lambda)*

take the mean of this distribution: *mean(rexp(n,lambda))*

Repeat 1000 times, and put it in a vector of means *(vecMeans)*

Initialise vecMeans:

```
vecMeans = NULL
```

As we are generating random numbers, we set the seed so we can replicate the analysis later on

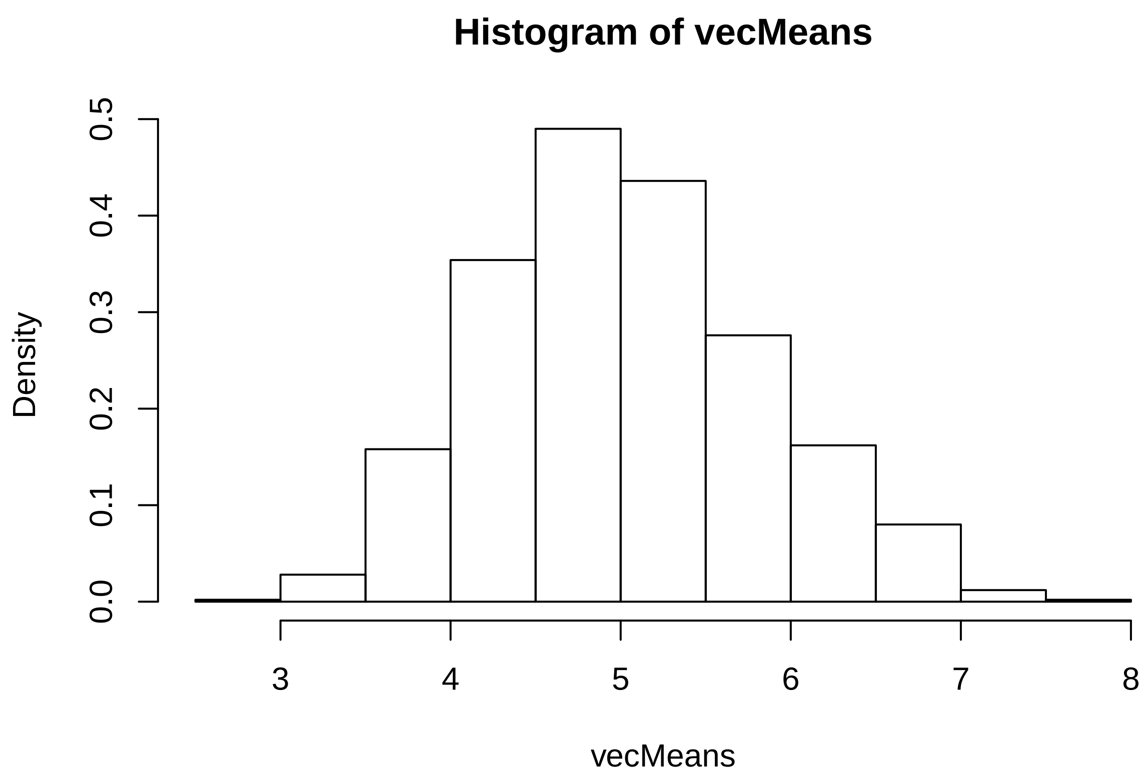
```r
set.seed(123456)
```

Run the simulation:

```r
for (i in 1 : iterates) vecMeans = c(vecMeans, mean(rexp(n,lambda)))
```

Draw a histogram showing the density (total area is equal to one). This will make it more logical when comparing to a normal distribution later on.

```r
hist(vecMeans,freq = FALSE)
```



Question 1: Show the sample mean and compare it to the theoretical mean of the distribution
Sample mean (or mean of means):

```r
mean(vecMeans)
```

```
## [1] 5.022915
```

The theoretical mean: 1/lambda

```r
1/lambda
```

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

Comparison:

```
mean(vecMeans) - 1/lambda
```

```
## [1] 0.02291512
```

It is pretty clear that the theoretical mean compares quite nice to the simulated mean.

Question 2: Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution:

The sample variance:

```
var(vecMeans)
```

```
## [1] 0.6570391
```

Theoretical variance: stdDev^2 / n

```
stdDev^2 / n
```

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

Comparison:

```
var(vecMeans) - stdDev^2 / n
```

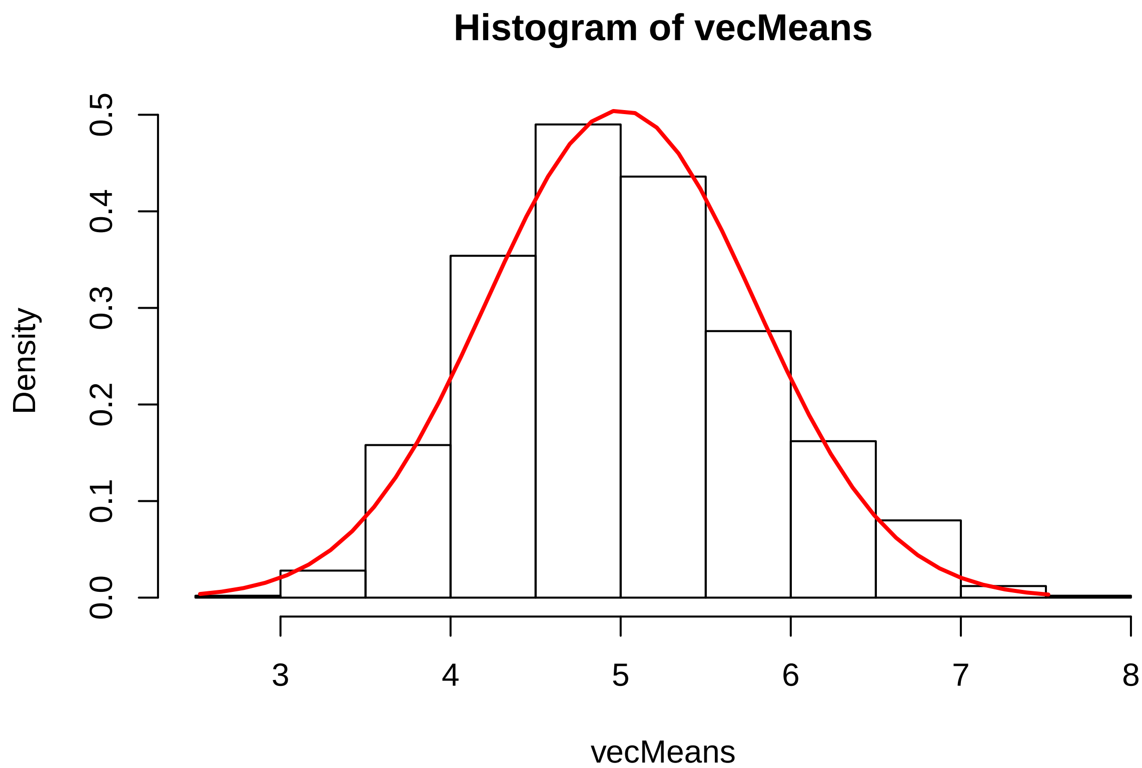
```
## [1] 0.03203913
```

Again, it is pretty clear that the theoretical variance compares quite nice to the simulated variance

Question 3: Show that the distribution is approximately normal:

CLT says: the arithmetic mean of a sufficiently large number of iterates of independent random variables will be approximately normally distributed. We do that by overlaying a normal distribution with the same mean and ranges as the histogram:

```
hist(vecMeans,freq = FALSE)
xrange <- seq(from=min(vecMeans), to=max(vecMeans), length.out=n)
yrange <- dnorm(x=xrange, mean=theoMean, sd=stdDev/sqrt(n))
lines(x=xrange, y=yrange, lty=1, col="red", lwd=2)
```



Conclusion:

Comparing the normal distribution “bell-curve” with the density histogram of the “mean of means”, shows quite illustrative the proof of the Central Limit Theorem (CLT).