Exponential Distribution and the CLT

Emanuele Di Saverio

21 Mar 2015

We will explore ad document the Central Limit Theorem. A simplified and short formulation of the CLT that I find easier to memorize is

“the distribution of **averages** of any series of **iid** variables is **normal**”.

(That means that can be normalized towards a **standard normal distribution** using mu and the standard error.)

We will go on testing this using a (series of) exponential distribution with rate **lambda = 0.2**.

Question 1

First let’s compare the theoretical mean and the sample mean

```
# parameters
lambda = 0.2
experiments = 1:1000 #lots od experiments
sample_size = 40
theoretical_mean = 1/lambda

#for each of the experiments, compute the mean of sample_size random values taken from ex
p(lambda)
sample_means <- sapply(experiments, function(x) mean(rexp(sample_size, lambda)))
print(mean(sample_means))
```

```
## [1] 5.009986
```

```
print(theoretical_mean)
```

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

These are the mean of the means of the experiments Theoretical mean is 1/lambda, which is 5.

Question 2

Now we can move to the **variance** of the sample

```
theoretical_var = (1/lambda)^2
sample_vars <- sapply(experiments, function(x) var(rexp(sample_size, lambda)))
print(mean(sample_vars))
```

```
## [1] 24.05309
```

```
print(theoretical_var)
```

```
## [1] 25
```

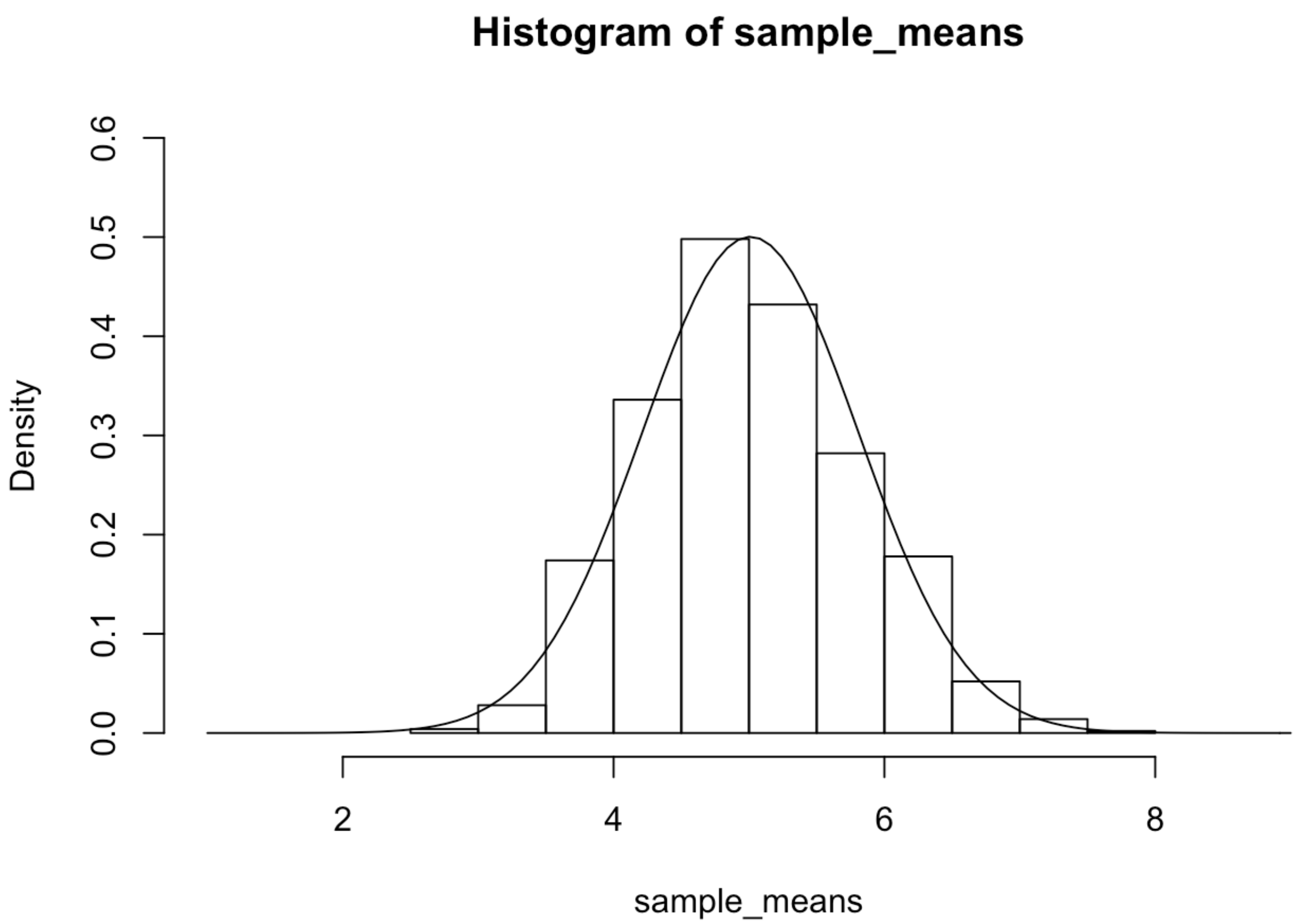
Theoretical variance for an exponential is square of standard deviation, which is also 1/lambda, hence 25.

Now we can see (re-stating CLT here in a more actionable way) that the mean of the means of samples of iid (any distriution) is an **unbiased estimator** of the mean of the population.

Question 3

Next, we move to show the CLT in action. Thanks to theorem we say that the distribution of the many means of the 40-sized samples is normal, and the distribution is approximately similar to a normal distribution with mean equal to the population mean, and standard deviation equal to the standard deviation of the means.

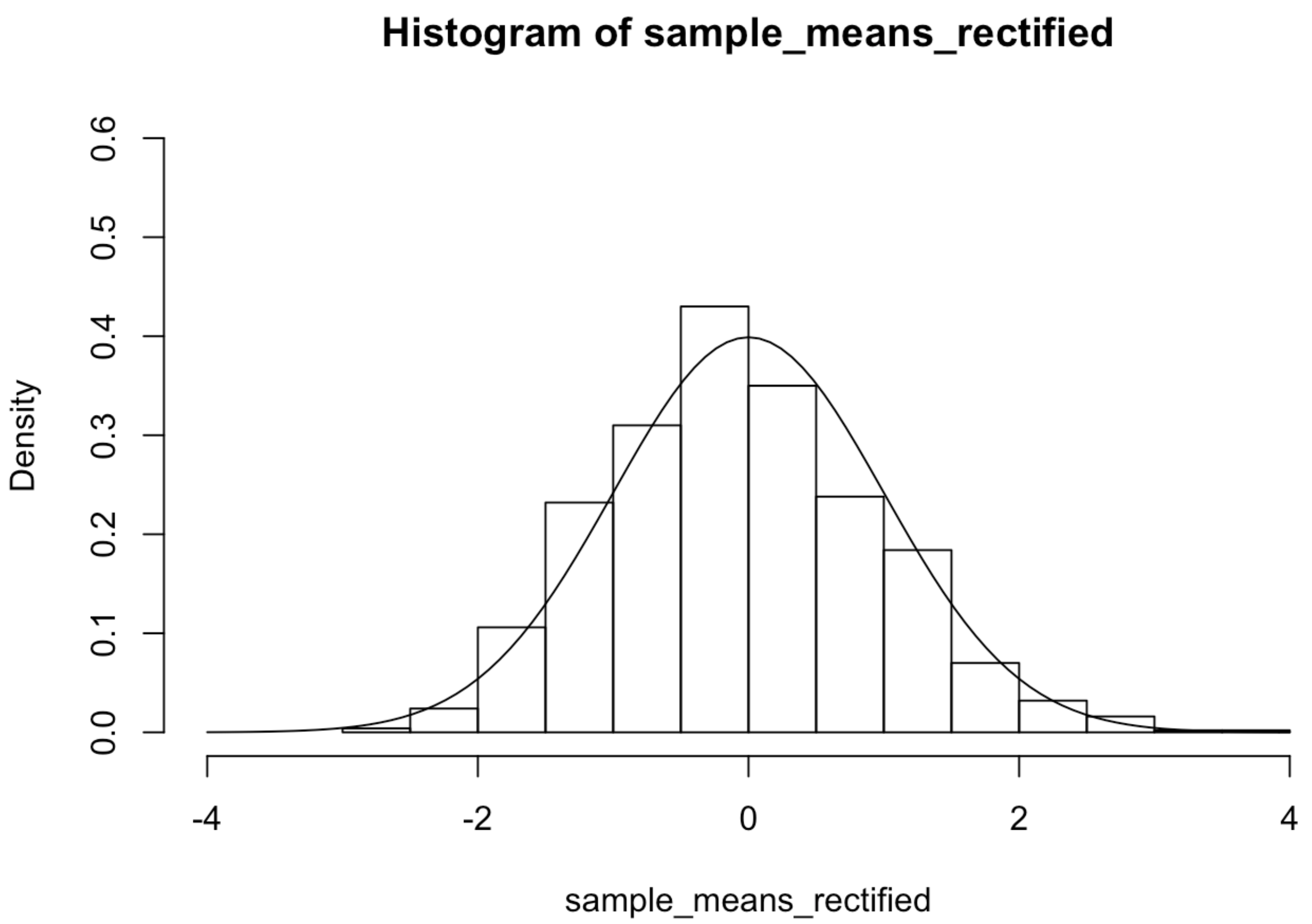
```
hist(sample_means, freq = F, ylim = c(0,0.6), xlim=c(1,9), breaks=12)
curve(dnorm(x, mean=mean(sample_means), sd = sd(sample_means)), add=T)
```



Apprendix - CLT by definition

If we want to stay closer to the original formulation of the CLT, we can re-compute the experiment to translate the result by the population mean and scale by the standard error in every measurement, and then re-plot the chart against a standard Normal Density (with mean 0 and standard deviation 1)

```
sample_means_rectified <- sapply(experiments, function(x) {
    vars <- rexp(sample_size, lambda) - theoretical_mean
    vars <- vars * sqrt(sample_size / theoretical_var)
    mean(vars)
})
hist(sample_means_rectified, freq= F, xlim=c(-4,4),ylim = c(0,0.6), breaks=12)
curve(dnorm(x, mean=0, sd=1), add=T)
```



and still find that the CLT applies.