Exponential Distribution and the CLT

Efstathios Fotiadis

Overview

The scope of this project is to investigate the behaviour of the exponential distribution under Central Limit Theorem. Various experiments was run the results are compared with the the theoretical predictions of CLT. Overall it was found that the experimental results follow the theoretical predictions very well.

Exponential distribution

The exponential distribution's probability density function is given by

$$f(x) = \lambda e^{-\lambda x}, x \geq 0$$

and 0 otherwise.

The mean, or expected value, and the variance of the exponential distribution are respectively

$$mean_{exponential} = \frac{1}{\lambda}$$

$$variance_{exponential} = \frac{1}{\lambda^2}$$

We will compare those theoretical predicted values with the results from our experiments.

Central limit theorem

The assumption of CLT is that *the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined (finite) expected value and finite variance, will be approximately normally distributed, regardless of the underlying distribution* (from Wikipedia)

In our case this translates as following. You draw randomly from an exponential a large number of times (n) and compute the mean. If you repeat the process various times then the distribution of the computed means will follow a normal distribution with a mean and variance predicted from theory. This also applies to other statistics like variance etc.

This normal distribution is also called sampling distribution of the mean. The variance of the sampling distribution of the means is given by the variance of the original distribution divided by the number of the samples, n, used to calculate each mean.

$$variance_{sampling} = \frac{variance_{exponential}}{n}$$

Experimental setup

We first set the constants as the rate of the exponential λ, the number of samples in each experiment n and the number of experiments s.

```
lambda = 0.2
n = 40
s = 1000
```

Each experiment consists of drawing $n = 40$ times from the exponential distribution and compute the mean. We run $s = 1000$ experiments and the means are saved in the list called `mns`

```
mns = NULL
for (i in 1 : s)
  mns = c(mns, mean((rexp(n, lambda))))
```

Sample Mean versus Theoretical Mean

First we compute the theoretical mean as $\frac{1}{\lambda}$ and the experimental mean from the data.

```
mean_theory = 1 / (lambda)
mean_sample = mean(mns)
print(c(mean_theory, mean_sample))
```

```
## [1] 5.000000 5.051172
```

We observe that the absolute difference is less than 2%. The experimental results validate the theoretically expected value.

Sample Variance versus Theoretical Variance

We compute the variance of the mean as it is predicted by theory $\frac{variance_{exponential}}{n}$ and also the variance of the means from the experiment.

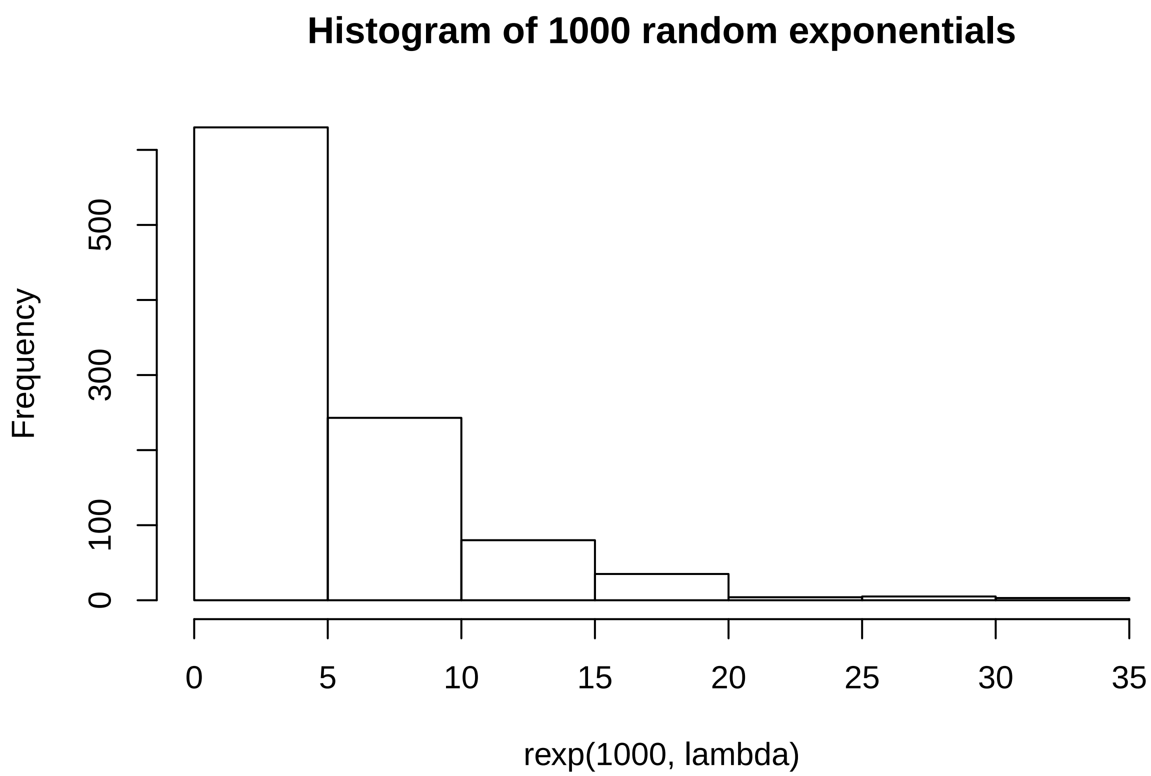
```
var_theory = (1 / (lambda^2 )) / n
var_sample = var(mns)
print(c(var_theory, var_sample))
```

```
## [1] 0.6250000 0.6471879
```

We observe that the absolute difference is less than 5%. The experimental results validate the theoretically expected value.

Histogram of random exponentials

We see the the histograms follows the exponential curve as expcted.



Distribution of means

The distribution of the 1000 means has a very different, normal, shape. It approximates well the normal distribution predicted by the Central Limit Theorem.