Exponential Distribution and the Central Limit Theorem

Ion Scerbatiuc

September 26, 2015

Synopsis

In the following analysis we will explore the exponential distribution and it's relationship with the Central Limit Theorem (or CLT). CLT states that the distribution of averages of iid variables becomes that of a standard normal as the sample size increases.

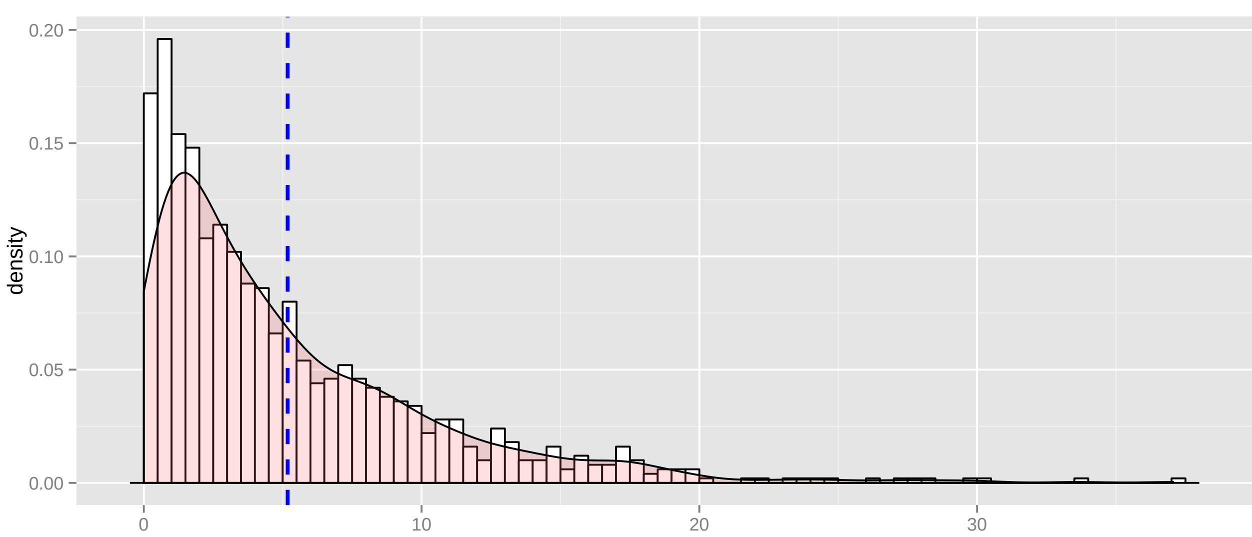
By using a big number of simulations we will show that the CLT applies also for exponential distributions. That is, we will show that the distribution of sample means is approximately normal.

```
library(xtable)
library(ggplot2)
set.seed(1001)
options(xtable.comment = FALSE)
```

The Exponential Distribution

Let's first start by visualizing a distribution of `n = 1000` iid variables drawn from an exponential distribution with `lambda = 0.2`.

```
n <- 1000
lambda <- 0.2
exp_dist <- data.frame(var = rexp(n, lambda))
ggplot(exp_dist, aes(x=var)) +
    geom_histogram(aes(y=..density..), binwidth=.5, color="black", fill="white") +
    geom_density(alpha=.2, fill="#FF6666") +
    geom_vline(aes(xintercept=mean(var)), color="blue", linetype="dashed", size=1)
```

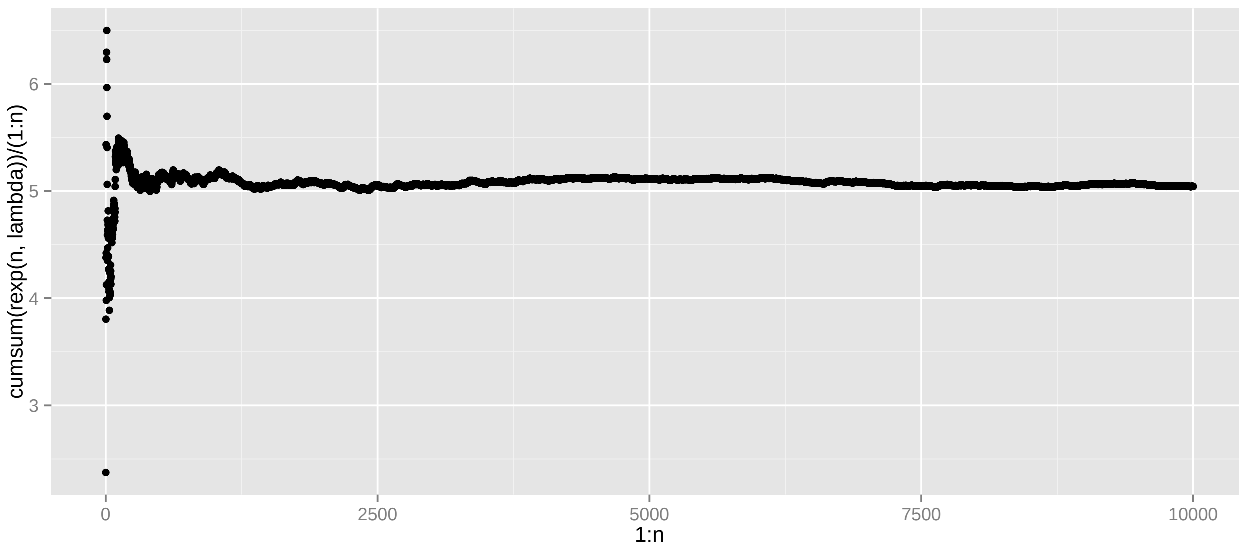


As we can see from the graph above, the exponential distribution has a shape that is very differently than the normal distribution. We also plotted the mean of the sample with a dotted blue line. The value is consistent with the theoretical mean of the distribution of `1 / lambda`.

The Central Limit Theorem

Let's now apply CLT to the exponential distribution and visualize what the sample mean looks like as the sample size increase. For this simulation we will draw `n = 10000` iid variables, will get the sample means for `1, 2, ..., n` variables and will plot them.

```
n <- 10000
qplot(1:n, cumsum(rexp(n, lambda)) / (1:n))
```



As we can see, at the beginning the sample means have a lot of variability, but as the sample size increases, the mean converges to the true mean, which is expected to `1 / lambda` in our case.

Exploring the distribution of sample means

Next, let's set-up the following experiment. We will repeatedly draw `n = 40` iid variables from an exponential distribution with `lambda = 0.2` and record the sample mean. We will repeat the simulation `count = 1000` times and will plot the distribution of the sample means.

```
n = 40
count <- 1000
exp_means <- replicate(count, {mean(rexp(n, lambda))})
```

Estimated Mean vs Theoretical Mean

As we know, the exponential distribution has a mean of `1 / lambda`. The estimated mean of a sample of iid variables from an exponential distribution is expected to be the same as the population distribution.

We can see in the table below that the center of mass of the distribution of sample means converges to the theoretical mean of the original distribution, thus reinforcing the Central Limit Theorem.

estimated	theoretical
4.96	5.00

Estimated Variance vs Theoretical Variance

The expected standard deviation of a sample is equal to the population standard deviation divided by the square root of the sample size. Let's now compare the variance of the distribution of sample means and the expected variance.

We can see in the table below, that the variance of the distribution of sample means converges to the theoretical variance of the original distribution.

estimated	theoretical
0.61	0.62

The sample means distribution

We expect the distribution of the sample means to be approximately normal, so we should see an approximately normal histogram, with the mean around `1 / lambda`.

In the following diagram we plot the distribution histogram and the estimated mean, as a vertical dotted blue line. We also plot the expect theoretical density function of a normal distribution with the mean `1 / lambda` and the standard deviation of `1 / lambda`.

We can easily see that the distribution of sample means is approximating pretty well the density function of the normal distribution.

```
t_mean <- 1 / lambda
t_sd <- (1 / lambda) / sqrt(n)
ggplot(data.frame(mu = exp_means), aes(x=mu)) +
    geom_histogram(aes(y=..density..), binwidth=.5, color="black", fill="white") +
    stat_function(fun = dnorm, arg = list(mean = t_mean, sd = t_sd), colour="red") +
    geom_vline(aes(xintercept=mean(mu)), color="blue", linetype="dashed", size=1)
```