The exponential distribution

Cristian Bidea

9 April 2016

The mean of the exponential distribution

In this project I'll investigate the exponential distribution in R and compare it with the Central Limit Theorem. We'll simulate the exponential distribution in R using the rexp(n, lambda) library function.

The mean of the exponential distribution is

$$\mu = \frac{1}{lambda}$$

And the standard deviation is also

$$\sigma = \frac{1}{lambda}$$

Now we'll use the Central Limit Theorem that states: if one takes random samples of size n from a population of mean Mu and standard deviation Sigma then as n gets large, the new random variable of sampled averages approaches the normal distribution with mean Mu and standard deviation sigma/sqrt(n).

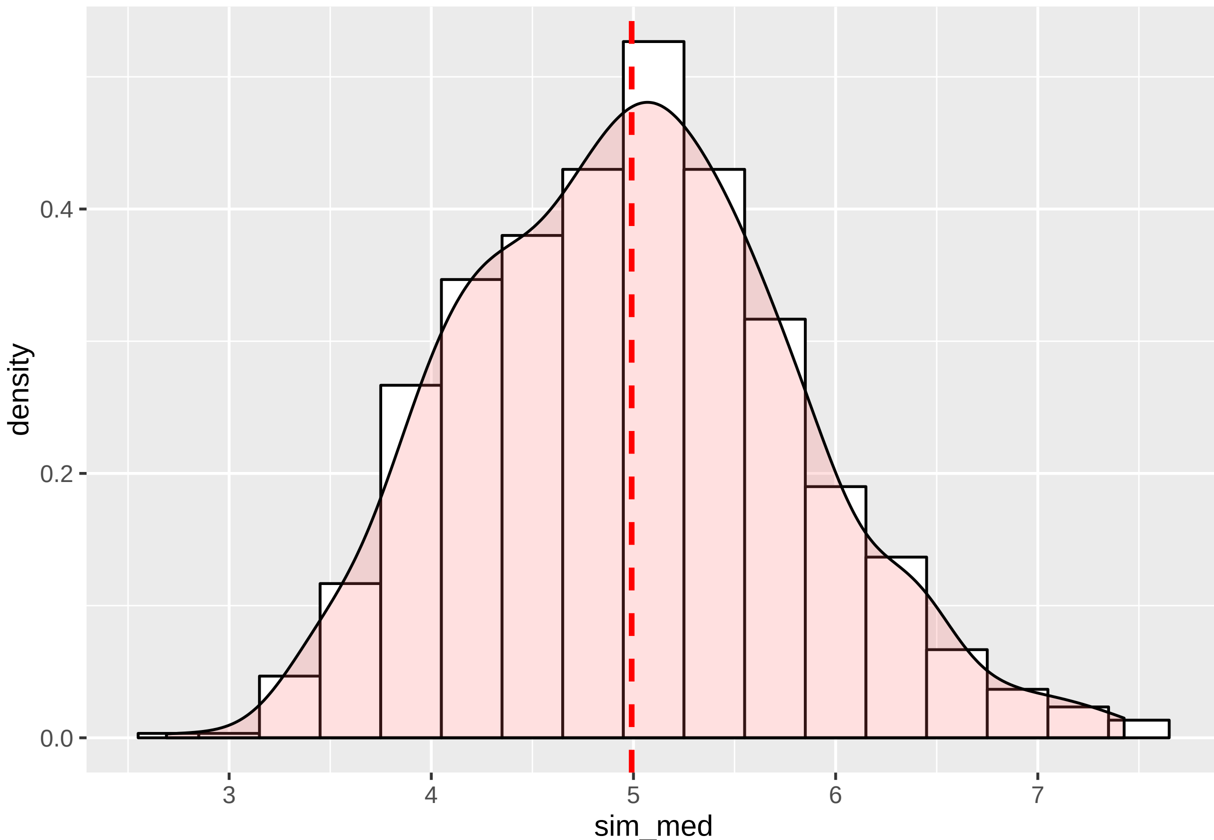
So we do the following simulation to get 1000, 40 samples averages and then plot them into a histogram, also plotting the density graph and mean.

```r
# simulation parameters
lambda <- .2         # from requirements
num_sims <- 1000     # number of simulations
num_samples <- 40    # number of samples per simulation
sim_med <- vector(mode = "numeric", length = num_sims)
sim_med <- simplify2array(lapply(sim_med,
function(n) {
    return (mean(rexp(num_samples, lambda)));
}))

# plotting the data
library (ggplot2)

dfsm <- as.data.frame(sim_med)

ggplot(dfsm, aes(x=sim_med)) +
    geom_histogram(aes(y=..density..), binwidth=.3, color="black", fill="white") +
    geom_density(alpha=.2, fill="#FF6666") +
    geom_vline(aes(xintercept=mean(sim_med)), color="red", linetype="dashed", size=1)
```



So we know the theoretical mean and standard deviation. They both are 5 in our example. We can see that the line drawn on the graph we plotted is also close to 5 (4.9568).

Variability

I look now at how variable the sample is and compare it with the theoretical variance of the distribution.

$$Variance = \sigma^2$$

```
# We know that the sampled medians will converge to a standard deviation
# equal to original standard deviation / sqrt(n)
s <- (1/lambda)/sqrt(num_sims)
var <- s^2
s
```

```
## [1] 0.1581139
```

```
var
```

```
## [1] 0.025
```

```
#The theoretical standard deviation is 1/lambda
s <- 1/lambda
var <- s^2
s
```

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

```
var
```

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

So the variance goes down for the sampled averages.

Showing that the distribution is approximately normal

```
require(gridExtra)
```

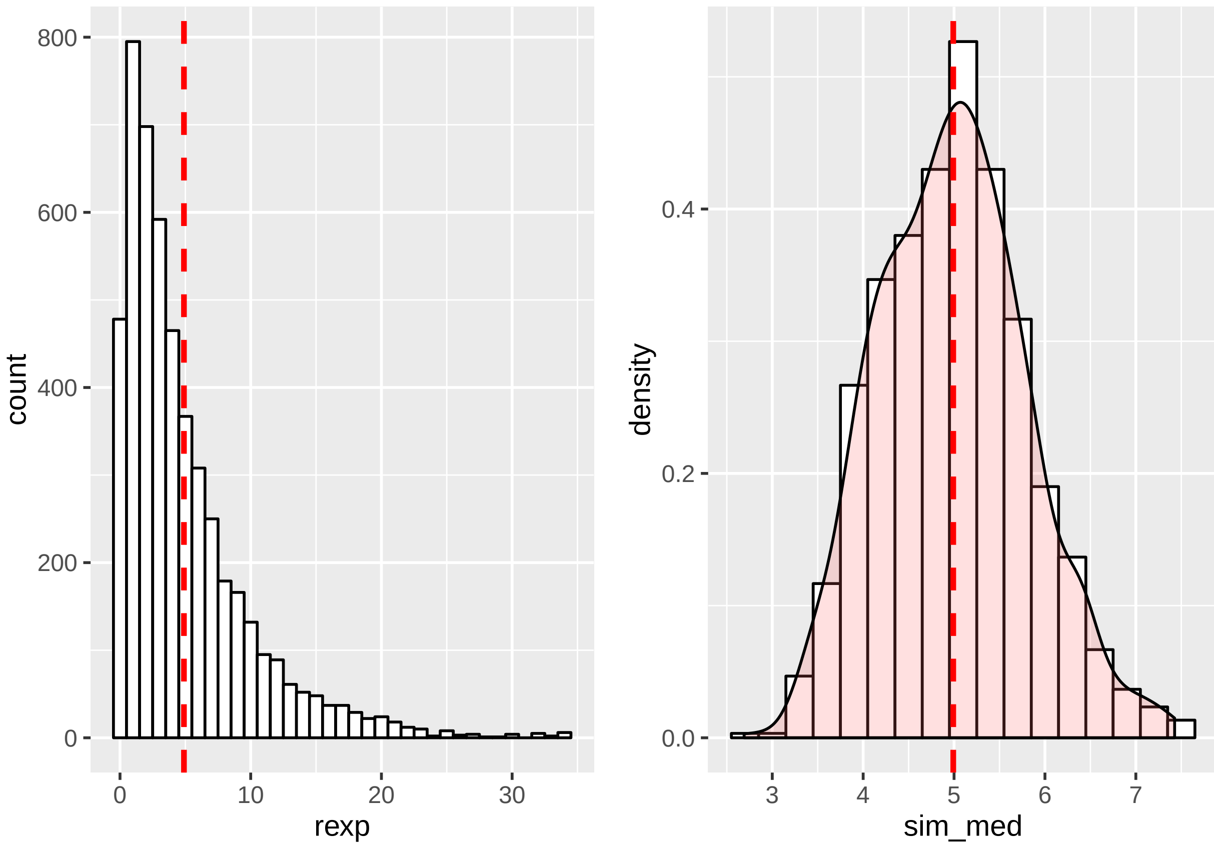
```
## Loading required package: gridExtra
```

```
rexp <- as.data.frame(rexp(5000, lambda))
names(rexp) <- "rexp"

plot1 <- ggplot(rexp, aes(x=rexp)) +
    geom_histogram(binwidth=1, color="black", fill="white") +
    geom_vline(aes(xintercept=mean(rexp)), color="red", linetype="dashed", size=1)

dfsm <- as.data.frame(sim_med)
plot2 <- ggplot(dfsm, aes(x=sim_med)) +
    geom_histogram(aes(y=..density..), binwidth=.3, color="black", fill="white") +
    geom_density(alpha=.2, fill="#FF6666") +
    geom_vline(aes(xintercept=mean(sim_med)), color="red", linetype="dashed", size=1)

grid.arrange(plot1, plot2, ncol=2)
```



The CLT is telling us that even if the random variable from which we sample averages isn't distributed normally then the distribution of the new random variable of the sampled averages we'll be distributed aproximately normal. This is obvious from the graphic above, looking at the shape in the right, we can see that it is bell shaped (more closely resembling a normal distribution).