Central Limit Theorem

Oleksandr Fialko

11/28/2016

The Central Limit Theorem (CLT) states that given a sufficiently large sample size from a population with finite variance, the mean of all samples from the same population will be approximately equal to the mean of the population. The samples means will follow an approximate normal distribution pattern centered at the population mean and the variance being approximately equal to the variance of the population divided by each sample's size.

In this part I investigate the exponential distribution in R and compare it with the CLT. 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` and the standard deviation is also `1/lambda`.

Here, I investigate the distribution of averages of 40 exponentials and do a thousand simulations for `lambda=0.2`:

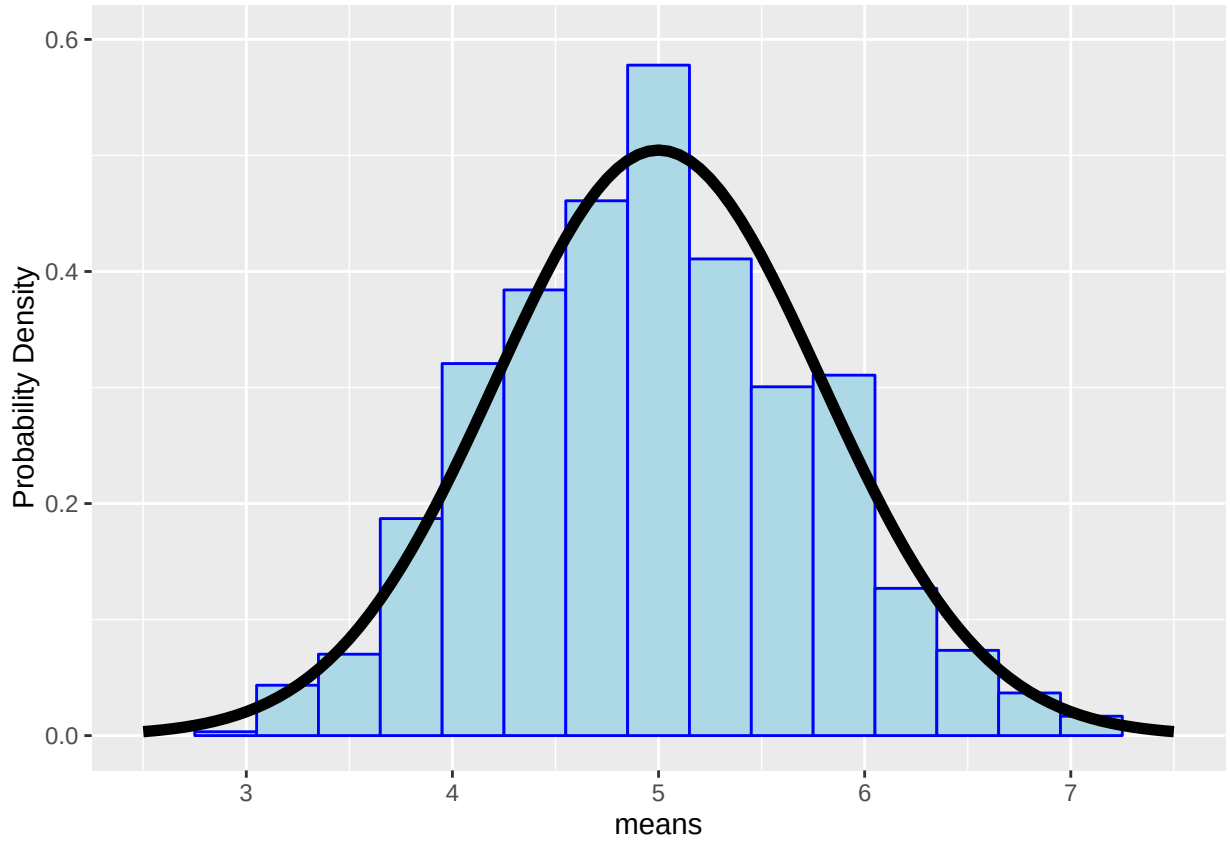
```
set.seed(12345)
lambda <- 0.2  # rate parameter
n <- 40        # a sample size
n_sim <- 1000  # number of simulations
# each row is a sample
sims <- matrix(rexp(n*n_sim,lambda),ncol = n)
```

According to the CLT, the samples means are distributed approximately normally. I create a function `norm_approx` to compare the distribution of the means with:

```
norm_approx <- function(x,mean,sd,n){
    dnorm(x,mean = mean,sd=sd/sqrt(n))
}
```

Here, I plot the normalized histogram of the means and compare it with `norm_approx`:

```
means <- apply(sims, 1, mean)
library(ggplot2)
g<-ggplot(data = data.frame(means=means),aes(x=means))
g<-g+geom_histogram(binwidth=0.3,col='blue',
                    fill='lightblue',center=1/lambda,
                    aes(y=..density..))
g+stat_function(fun=norm_approx,
                args = list(mean=1/lambda,sd=1/lambda,n=n),geom='line',size=2)+
    xlim(2.5,7.5) + ylim(0,0.6) + ylab('Probability Density')
```



Samples means and their standard deviation are

```
c(mean(means),sd(means))
```

```
## [1] 4.9719720 0.7847246
```

These values are in good agreement with the mean and the standard deviation of the `norm_approx`, namely

```
c(1/lambda,1/lambda/sqrt(40))
```

```
## [1] 5.0000000 0.7905694
```