Central Limit Theorem

Statistical Inference Project Part I, Class 6 in data science series

Ann Crawford

Snopiss

Central Limit Theorem

The Central Limit Theorem (CLT) states that the distribution of averages of independent and identically distributed (iid) variables becomes that of a **standard normal** as the sample size increases even if the original variables are not normally distrubted. This document investigates the exponetial distribution, generated using R function rexp and compares it to the CLT.
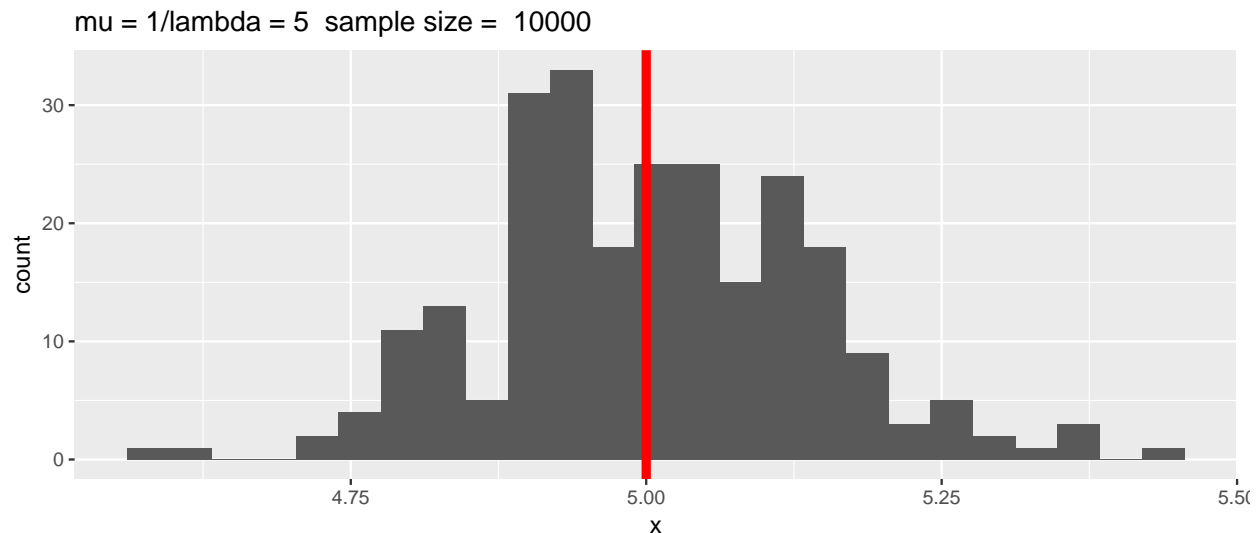
$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\text{Estimate} - \text{Mean of estimate}}{\text{Std. Err. of estimate}}.$$

```
# Clear our workspace
rm(list=ls())
```

The sample mean compared to the theoritical mean

```
## Warning: package 'ggplot2' was built under R version 3.3.3
```
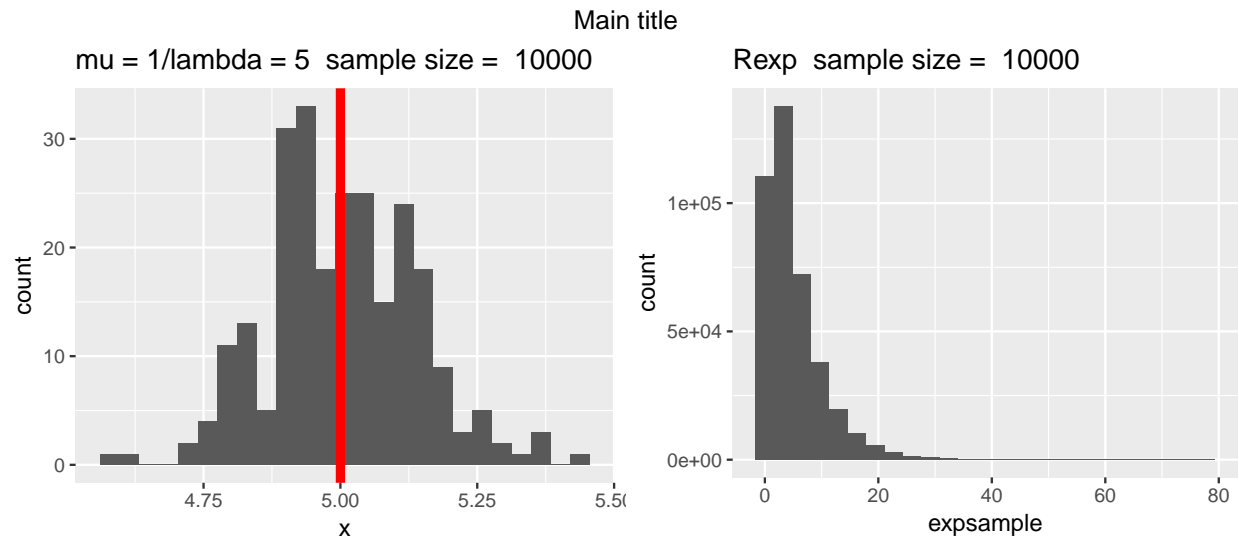
```
## Warning: package 'gridExtra' was built under R version 3.3.3
```

mu = 1/lambda = 5 sample size = 10000


```
## Don't know how to automatically pick scale for object of type data.frame. Defaulting to continuous.
```

Rexp sample size = 10000

Don't know how to automatically pick scale for object of type data.frame. Defaulting to continuous.

Main title
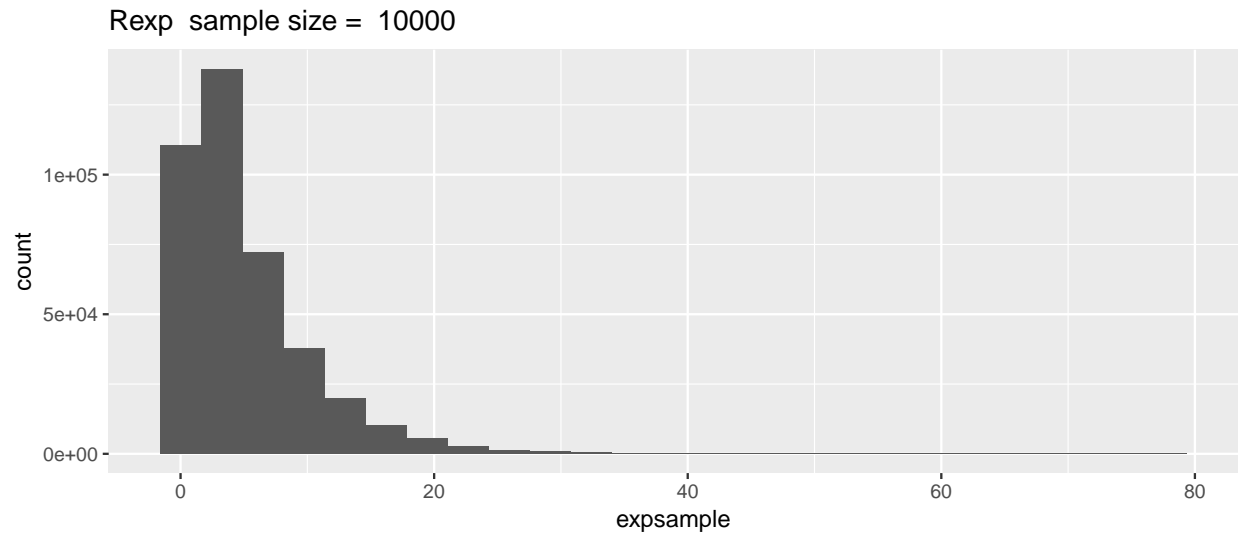
Show that the distribution of the sample means is normal

The histogram shows a distribution with a shape similar to the normal curve. The density proportions shown in the table below are approximately normal.