Applying the Central Limit Theorem to the exponential distribution

Pier Lorenzo Paracchini

Overview

The Central Limit Theorem states that *"the distribution of averages of indipendent and identically distributed (iid) variables becomes that of a standard normal as the sample size increase."* so in this assignment it will be investigate the behaviour of the **exponential distribution** when looking at distribution of averages and compare it with **the Central Limit Theorem**.

Simulations

Lets simulate an experiment with a random variable X with an **exponential distribution** with the following characteristics

- rate parameter $\lambda = 0.2$
- mean $\mu = E[X_i] = 1/\lambda = 5$
- standard deviation $\sigma = \sqrt{2xVar(X_i)} = 1/\lambda = 5$
- standard error (SE) $\sigma/\sqrt{2n}$

For each simulation

- run the experiment n time where `n = 40`,
- be X_i the outcome for the experiment $n = i$.
- calculate the mean $\overline{X}_n$ using $X_1, X_2, .., X_{40}$
- calculate $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$

and repeat thousands of times, keeping the calculated values. Below the code snippet performing the steps previously defined.

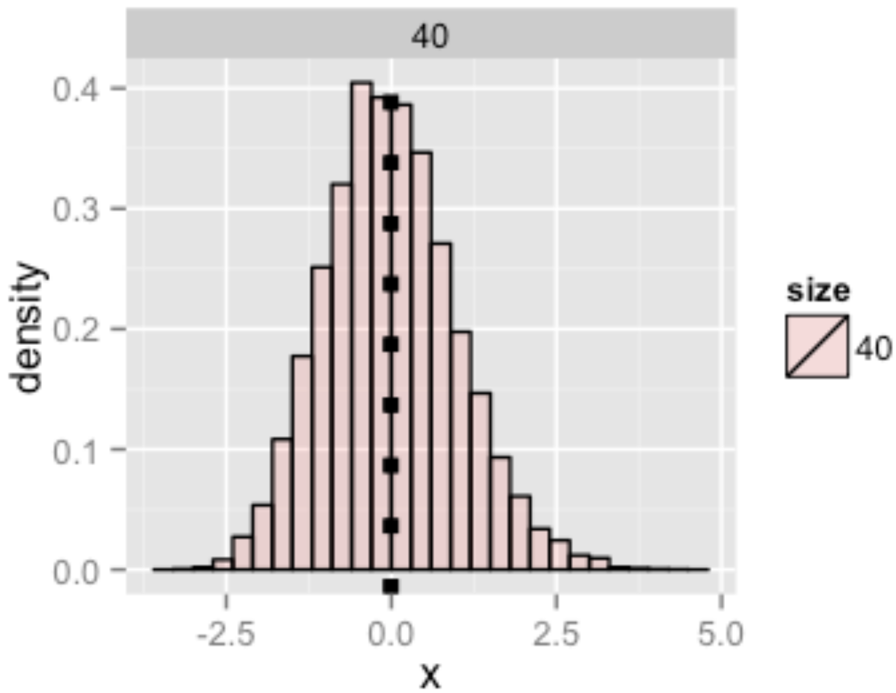
```
no_of_simulation = 10000
n = 40
clt_func <- function(x, n) sqrt(n) * (mean(x) - 5) / 5

dat <- data.frame(
  x = c(apply(matrix(rexp(n * no_of_simulation, 0.2), nrow =
no_of_simulation), MARGIN = 1, clt_func, n)),
  size = factor(rep(40, no_of_simulation))
  )
```

Sample Mean versus Theoretical Mean

Using the data from the simulations lets focus on the mean. For the distribution of averages previously calculated we know that the expected **theoretical mean** is $\mu = 0$ and the **sample mean** should have the same value. The current value of the sample mean based on the data from simulations is **-0.0127413**.

Lets highlight the **sample mean** with a dotted vertical line in **-0.0127413**, we can see from the plot below that it is almost centered in 0.

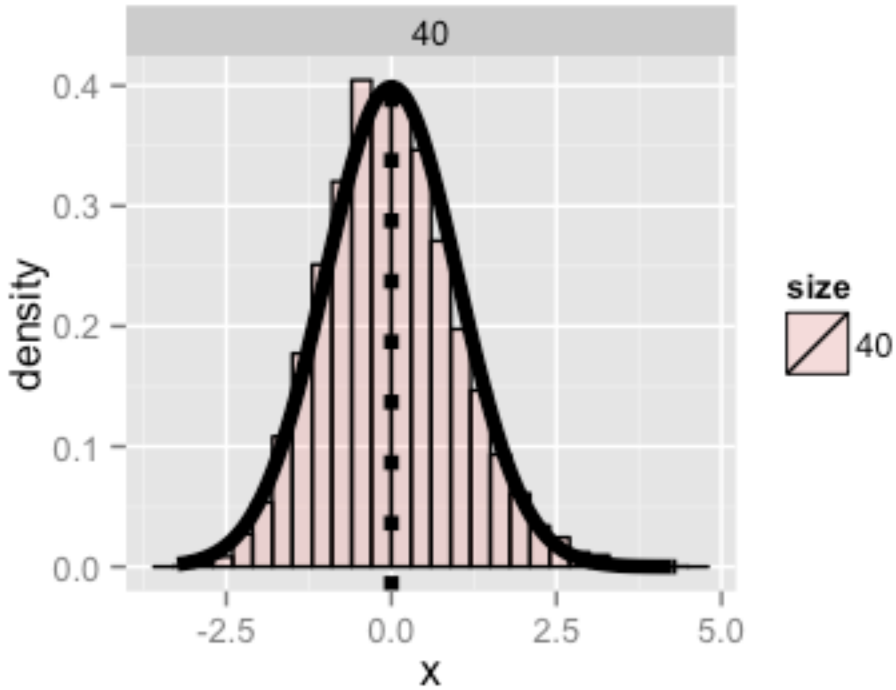


Sample Variance versus Theoretical Variance Include figures (output from R)

with titles. Highlight the variances you are comparing. Include text that explains your understanding of the differences of the variances.

Distribution

Lets plot the distribution of the averages previously calculated over all of the simulations together with the standard normal distribution N(0,1). We can see from the plot below, the distribution of averages ($\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$) is a good approximation of a standard normal distribution N(0,1) as expected from the Central Limit .


40
0.4
0.3
0.2
0.1
0.0
density
-2.5
0.0
2.5
5.0
x
size
40