

The Exponential Distribution

Julian Hatwell

Contents

- Overview
- Simulations
- Sample Mean vs Theoretical Mean
- Sample Variance vs Theoretical Variance
- Appendix: Comparing Exponential and Normal Distributions

Overview

This document will use the Exponential distribution to investigate the Central Limit Theorem, which states that the *distribution of sample averages* of any iid variables becomes that of a standard normal as the sample size increases, no matter what the distribution of the population from which the samples are drawn.

The Exponential distribution makes a good case study as it has very different properties to the Normal distribution.

This investigation uses simulations carried out in the R programming environment.

Simulations

The following code sets up all the parameters and constants, as well as the random seed for reproducibility:

```
# here are all the standard parameters. Can be tweaked for more further investigations
set.seed(10) # reproducibility
lambda <- 0.2 # lambda for random samples
sim_size <- 1000 # simulation size
smpl_size <- 40 # sample size, each simulation
```

The following code runs the simulations. A for loop is used with the rexp (random exponential) function and the mean and var statistics are collected inside the loop.

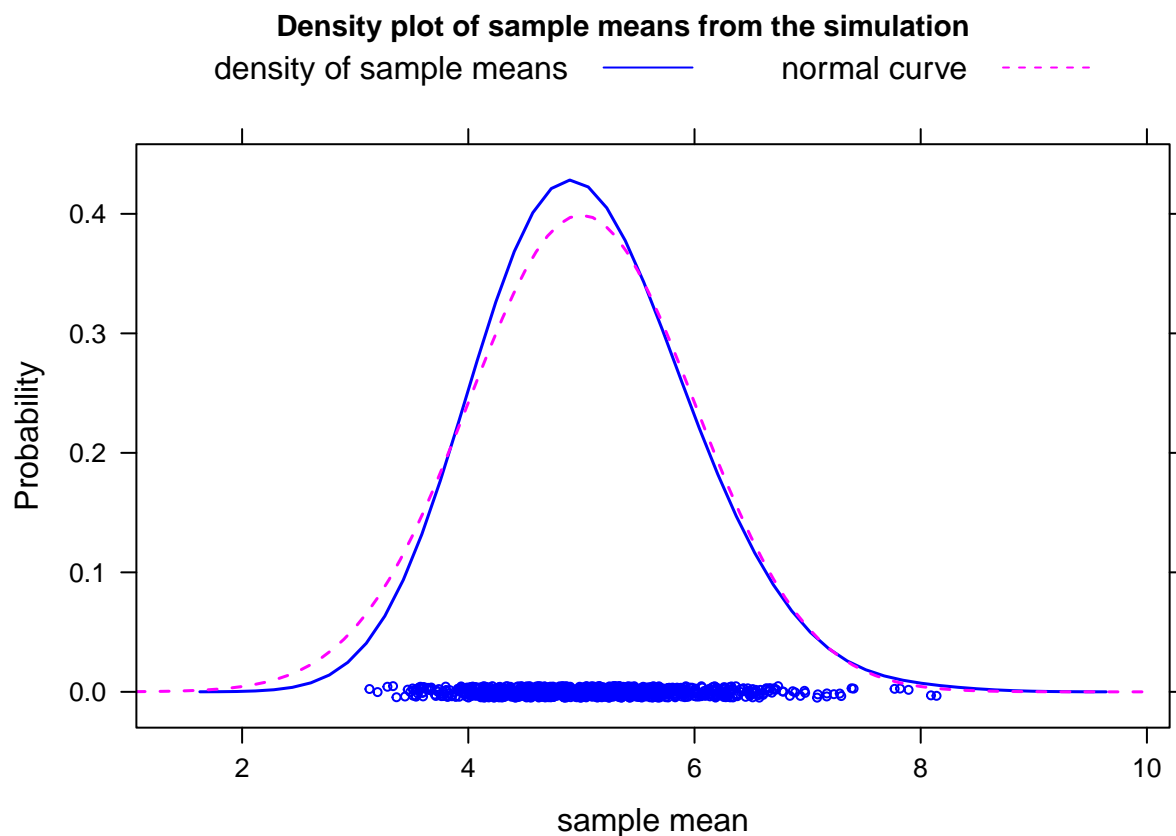
```
# create some empty vectors to fill with sim data
mn_smpl <- numeric(sim_size)
var_smpl <- numeric(sim_size)
# iterate over the random simulation and collect the results by vector index
for (i in 1:sim_size) {
  smpl <- rexp(smpl_size, lambda)
  mn_smpl[i] <- mean(smpl)
  var_smpl[i] <- var(smpl)
}
# put everything in a data frame to make it easy to work with
sim_data <- data.frame(means = mn_smpl, variances = var_smpl, st_devs = sqrt(var_smpl))
```

Sample Mean vs Theoretical Mean

Interrogating the sample mean data shows that the mean and st. dev are indeed very close to $1/\lambda$ (5), which are the expected mean and st.dev for the Exponential distribution with λ value = 0.2:

```
## $sample_mean
## [1] 5.04506
##
## $sample_variance
## [1] 25.56219
##
## $sample_st_dev
## [1] 5.055907
```

The following density plot shows the distribution of mean values from the simulation. Superimposed on this is a PDF of the Standard Normal distribution, shifted to centre on a mean of 5 ($1/\lambda$).



It is clear from the above graph that these are very similar curves. The distribution of the sample means is a very close approximation of the normal distribution.

Sample Variance vs Theoretical Variance

This step will compare the variance and standard deviation of the sample means to the theoretical mean given as the formula s/\sqrt{n} .

The variance of the sample means should be close to $5/\sqrt{40} = 5/6.3245553 = 0.7905694$

```
stats <- data.frame(
  theoretical_var = (1/lambda)^2/smpl_size
  , theoretical_sd = (1/lambda)/sqrt(smpl_size)
  , simulation_var = var(sim_data$means)
  , simulation_sd = sd(sim_data$means))
stats

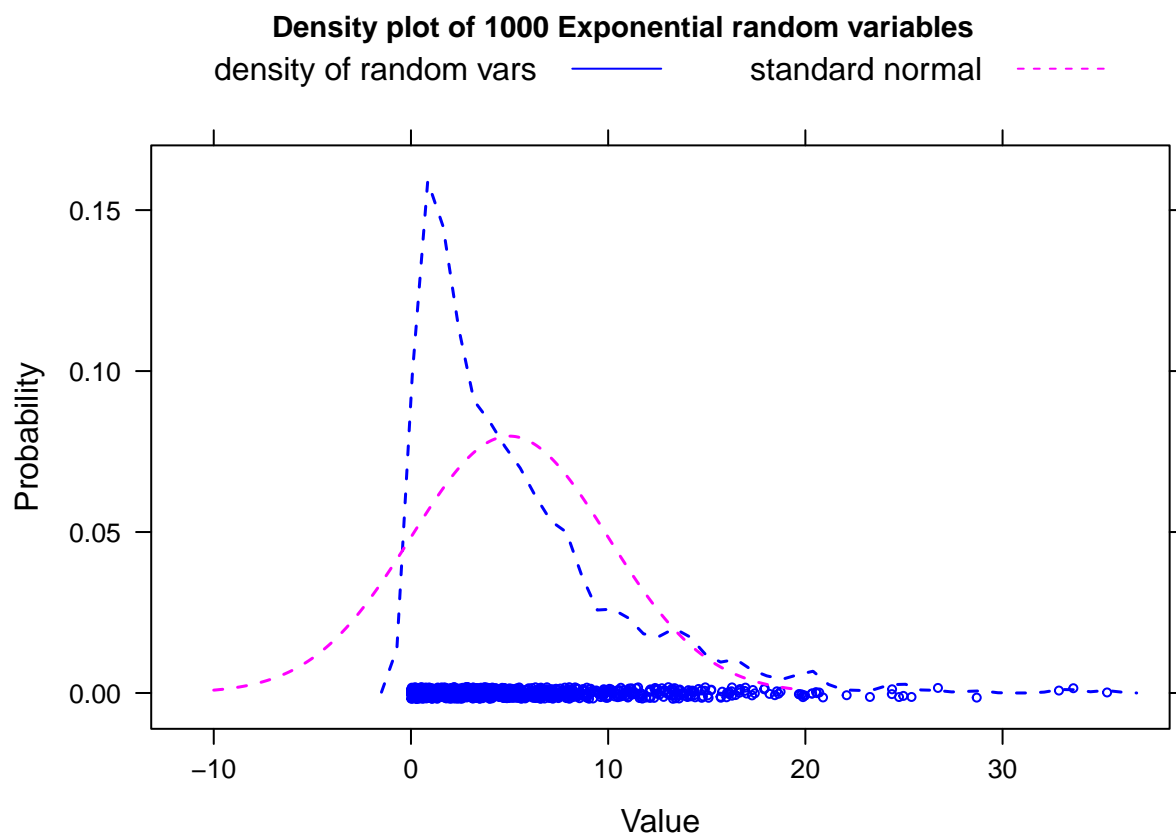
##   theoretical_var theoretical_sd simulation_var simulation_sd
## 1           0.625      0.7905694      0.6372544      0.7982821
```

The simulation statistics are very close to the theoretical statistics.

Appendix: Comparing Exponential and Normal Distributions

For completeness, more random variables are generated from the exponential function to see the difference between a large collection of random exponentials and the distribution of a large collection of averages of random variables.

The following plot shows the distribution of 1000 random variables with an exponential distribution. Superimposed on this is a PDF of the Normal distribution, shifted to centre on the mean of 5 ($1/\lambda$) and a st.dev of 5.



The differences are visually apparent. From this graph, it can be seen that the Exponential distribution has the characteristic of having $p = 0$ for values of $x < 0$. The first positive value of x has $p = \lambda$ (0.2), then p drops off sharply until a long tail of higher values.

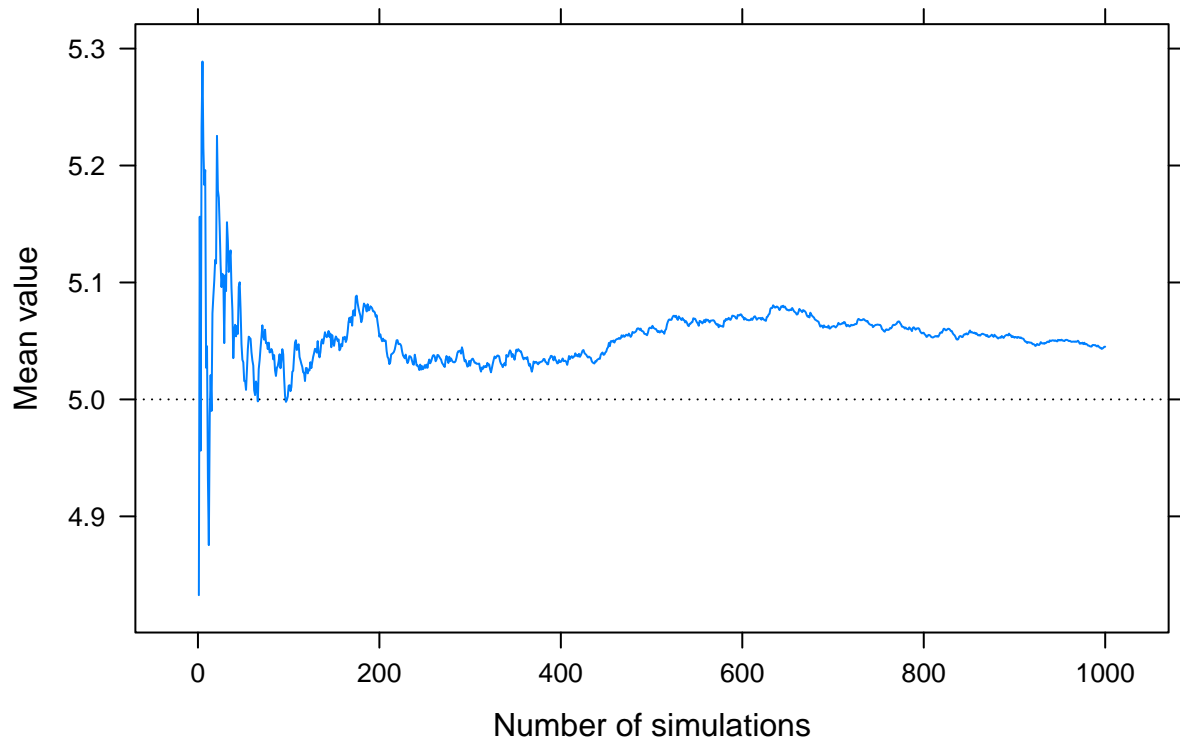
In contrast, the normal distribution is symmetrical about its mean and the standard normal has mean zero, implying negative values are just as probable as positive values.

This demonstrates the Central Limit Theorem by showing that a large sample of averages will tend to be themselves normally distributed, even though the random variables themselves are very differently distributed.

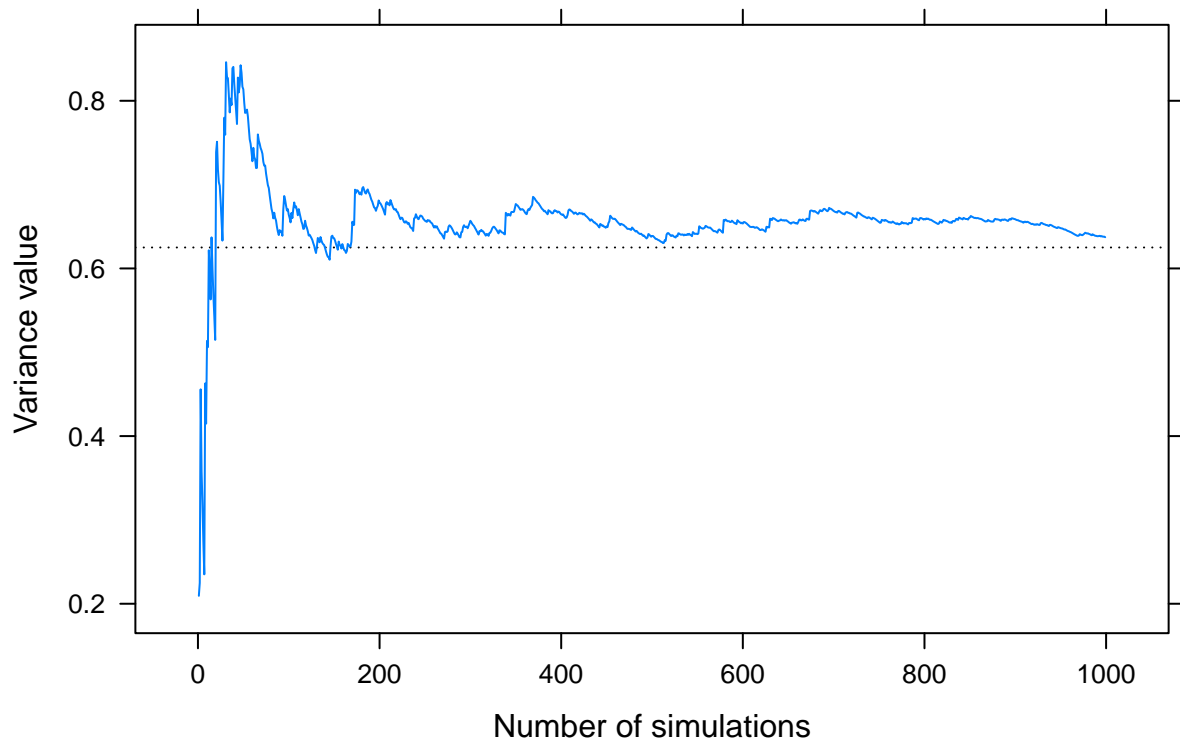
Appendix 2: Law of Large Numbers in Action

This was added after reviewing peer reports and deadlines, for the purpose of increasing learning. Credit to the other students, I don't know who they are, for coming up with this. Minor adaptations to do a lattice plot (my preference).

Sample Mean: Convergence to Theoretical Value



Variance of the Sample Mean: Convergence to Theoretical Value



Appendix 3 Using the maths to compare the simulation to the normal distribution.

This was added after reviewing peer reports and deadlines, for the purpose of increasing learning. Credit to the other students, I don't know who they are, for coming up with this. Minor adaptations to do a lattice plot (my preference).

“To [quantify] the extent to which the empirical distribution is approximately normal, the following computes how much area (amount of data) is captured by going 1 standard deviation in either direction from the mean, by going 1.96 standard deviations in either direction, and by going 2.58 standard deviations in either direction. For a normal distribution, these correspond to the 68%, 95%, and 99% confidence intervals, respectively.”

```
## [1] 0.6826213
```

```
## [1] 0.9491512
```

```
## [1] 0.98203
```