The Exponential Distribution vs the Central Limit Theorem

Zoheb Nensey

Overview

This document is intended to show the differences between the exponential distribution and the distribution that results as a result of the central limit theorem (the normal distribution).

Simulations

Loading necessary libraries.

```
library(knitr)
library(ggplot2)
```

First, I set a number of constants in concurrence with the assignment. For this assignment, the lambda is the rate constant, and is uniform for all simulations.

```
# Setting a seed so that this is reproducible.
set.seed(234234)

# Setting lamba.
lambda <- 0.2

# Setting number of simulations.
nosim <- 1000

# Number of samples (per assignment)
n <- 40
```

Then, I ran the simulations.

```
exponential_distributions <- matrix(rexp(nosim*n, rate=lambda),nosim,n)
exponential_means <- data.frame(means=apply(exponential_distributions, 1, mean))
```

Comparing the sample mean to the theoretical mean.

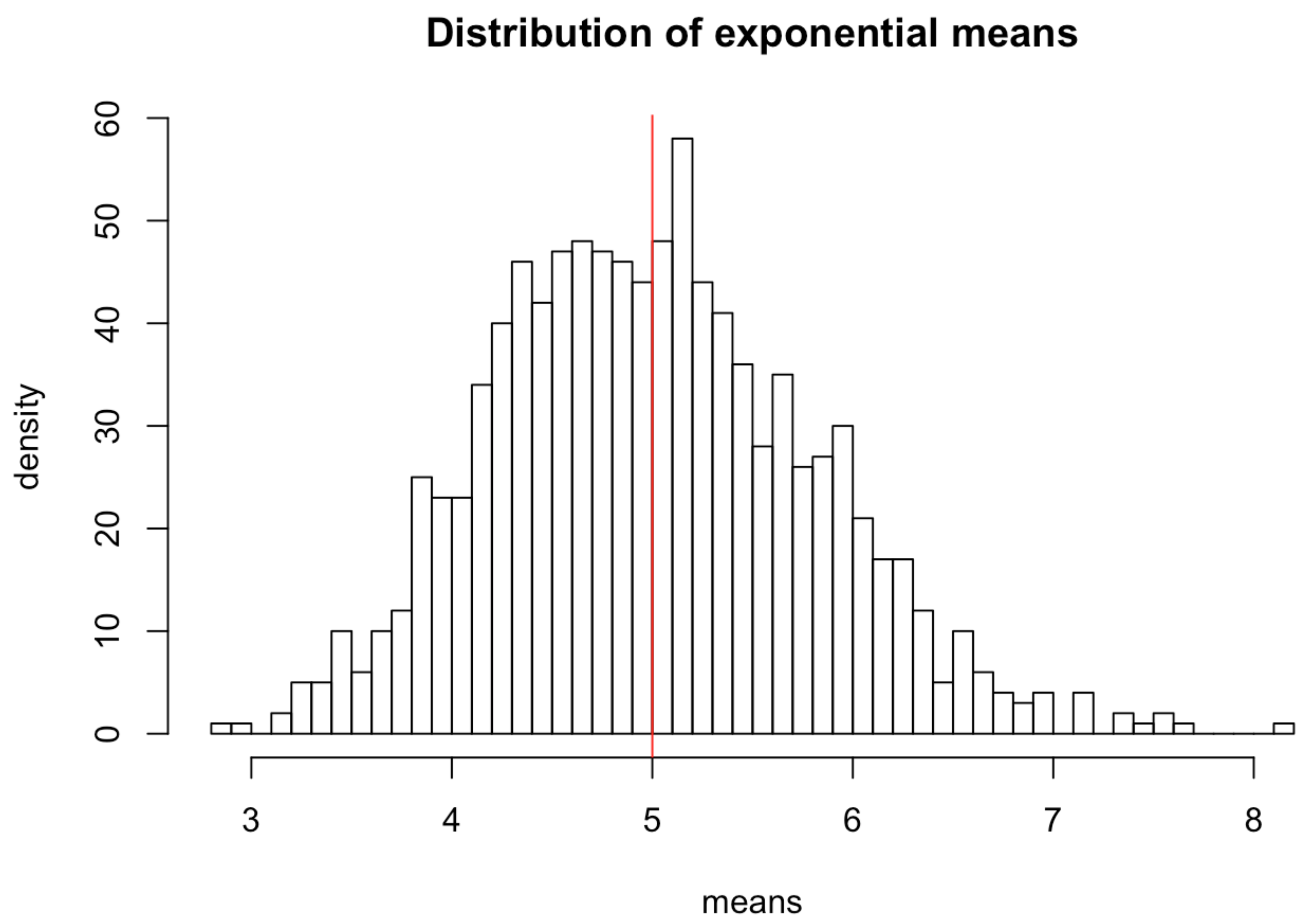
At the theoretical level, the mean of an exponential distribution is 5, or at least for our purposes, 1/lambda.

Upon calculating the mean of the sample distribution as shown below, we can see that the sample mean is very close to the theoretical mean.

```
sample_mean <- mean(exponential_means$means)
```

The sample mean is 4.9984444.

This can also be shown in the form of a histogram. The line in the center is the theoretical mean. Note that the highest density is located around the theoretical mean.



Though the peak of the distribution is just to the right of 5, increasing the number of simulations would bring the distribution closer and closer to the theoretical mean.

Comparing the theoretical variance to the sample variance.

The givens indicate that the theoretical standard deviation is 1/lambda divided by the square root of the sample size. The theoretical variance, which is simply the standard deviation squared, can be derived from this.

```
theoretical_sd <- (1/lambda)/sqrt(n)
theoretical_var <- theoretical_sd^2
```

The theoretical standard deviation is 0.7905694 and the theoretical variance is 0.625.

Now we can calculate the standard deviation and variance of the sample itself by taking the standard deviation of the sample means.

```
sample_sd <- sd(exponential_means$means)
sample_var <- sample_sd^2
```

The sample's standard deviation is 0.8040381 and the variance is 0.6464773.

These numbers, due to the large sample size, are not particularly different. The sample variance of 0.6464773 is not very different from the theoretical value of 0.625.

Comparing to the normal distribution.

The plot below shows the normal distribution (red line) superimposed on top of the distribution of the sample mean (blue line). Given that there are few differences between the sample and the theoretical distribution, it can serve as a useful approximation for comparison to the normal distribution.

```
ggplot(data=exponential_means, aes(x=means)) +
      geom_histogram(binwidth=0.1,aes(y=..density..),alpha=0.2) +
      stat_function(fun = dnorm, arg=list(mean = theoretical_mean, sd = theoretica
l_sd)
                    , col = "red") +
      geom_density(col="blue")
```