Statistical Inference: Final Project; Part 1

Hamed Bastan-Hagh

20 July 2016

The Exponential Distribution and the Central Limit Theorem

Overview

In this project I will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with `rexp(n, lambda)` where lambda is the rate parameter. The mean of the exponential distribution is 1/lambda and the standard deviation is also 1/lambda. I will set lambda = 0.2 for all of the simulations. I will investigate the distribution of averages of 40 exponentials. I will illustrate via 1,000 simulations and associated explanatory text the properties of the distribution of the mean of 40 exponentials.

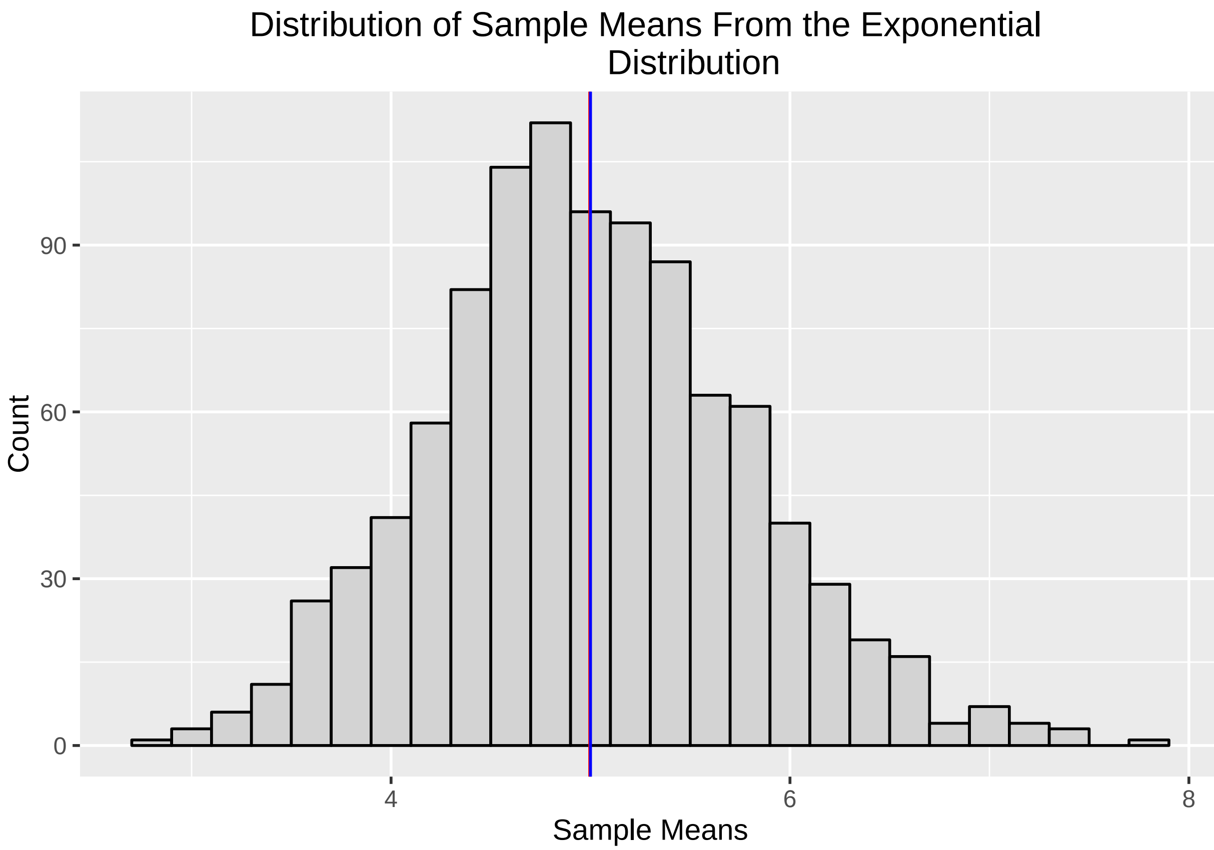
We will see that the distributions of the sample means and variance are as we would expect from the Central Limit Theorem in our simulations.

Question 1:

Compare the sample mean to the theoretical mean of the distribution.

To answer this we need 1,000 samples with size n = 40 from the exponential distribution. We can simulate this using the `rexp` function, and then plot the means of each of those samples in a histogram.

```
# Load stats library
library(stats)
# Set n to 40, number of sims to 1,000 and lambda to 0.2 as directed
n <- 40
sims <- 1000
lambda <- 0.2
# Set seed to make this reproducible
set.seed(1981)
# We will generate 40,000 random values from the exponential distribution
# using rexp. These go into a 1000 x 40 matrix: each of the 1,000 rows is one
# sample with size 40.
data <- matrix(rexp(n = n * sims, rate = lambda), sims)
# Calculate the means for each of these samples
samp.mean <- apply(data, 1, mean)
# Show histogram of the distribution, with vertical lines at the sample mean
# and the theoretical mean
library(ggplot2)
xbar <- mean(samp.mean)
mu <- 1 / lambda
ggplot(data = data.frame(Means = samp.mean), aes(x = Means)) +
    geom_histogram(colour = "black", fill = "lightgrey", binwidth = 0.2) +
    geom_vline(xintercept = xbar, colour = "red") +
    geom_vline(xintercept = mu, colour = "blue") +
    labs(title = "Distribution of Sample Means From the Exponential
         Distribution",
         x = "Sample Means",
         y = "Count")
```



NB. The red vertical line shows the sample mean, while the blue line is the theoretical mean. They are very close to each other so you may find it difficult to see the red line.

Sample Mean

The sample mean is 4.996

Theoretical Mean

The theoretical mean is `1 / lambda`, which is 5.

Comparison

The difference between the sample and theoretical means is only 0.0039157, so we can see that the sample mean is a close approximation to the theoretical mean.

Question 2:

Compare the variance of the sample to the theoretical variance of the distribution.

The variance of our distribution of sample means is **0.62.**

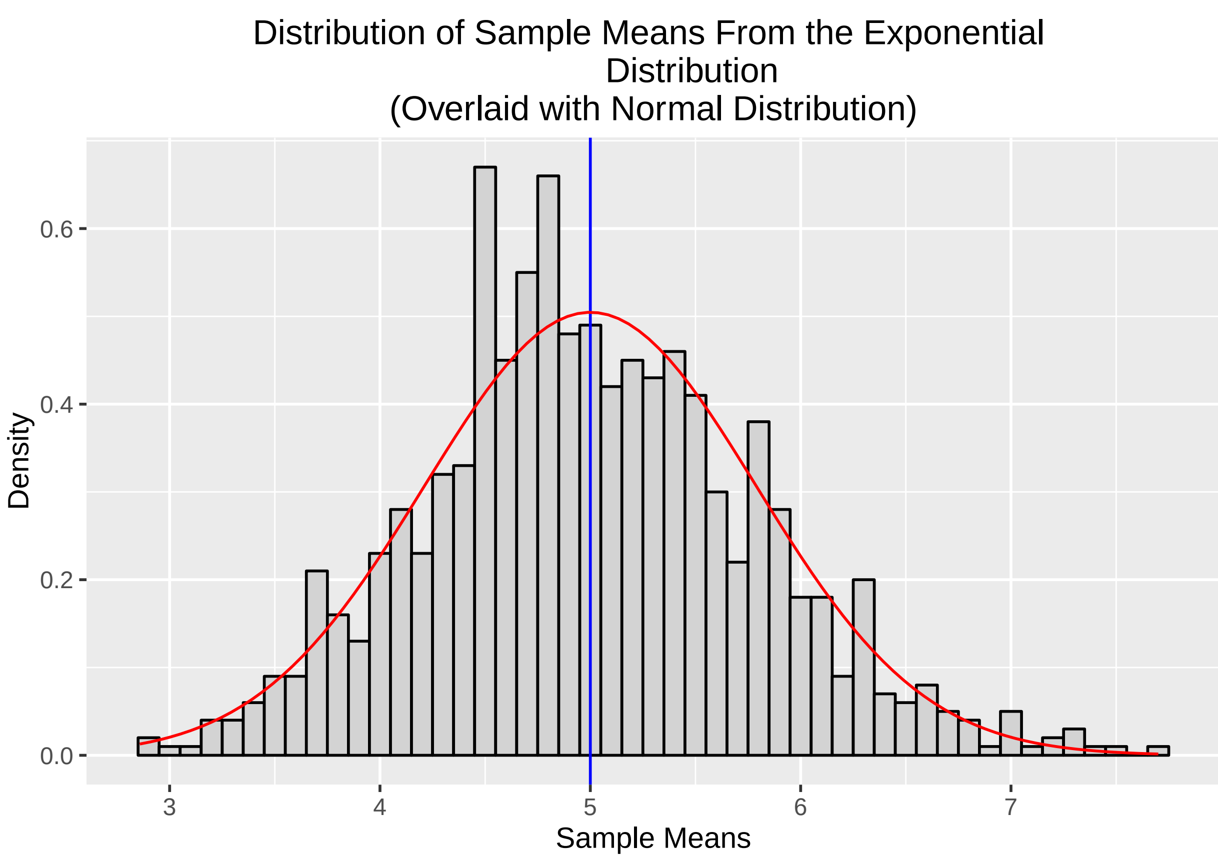
The theoretical variance of an exponential distribution is `1 / lambda` = **5**. So the expected variance of our sample means is `(1/n) * (1/lambda^2)` = **0.625**. **The difference between these is 0.009339**. We can see that the variance of our sample means is again very close to the theoretical figure, as we would expect from the Central Limit Theorem.

Question 3:

Show that the distribution is approximately normal.

We can demonstrate that the distribution is approximately normal by overlaying the curve for the normal distribution on the histogram of the distribution of the sample means.

```
# Assign the variance and standard deviation of the sample to variables
samp.var <- (1/n) * (1/lambda^2)
samp.sd <- sqrt(samp.var)
# Plot the distribution of the sample means with mean and normal curve
# overlaid
ggplot(data = data.frame(Means = samp.mean), aes(x = Means)) +
    geom_histogram(colour = "black",
                   fill = "lightgrey",
                   binwidth = 0.1,
                   aes(y = ..density..)) +
    geom_vline(xintercept = mu, colour = "blue") +
    stat_function(fun = dnorm,
                  color = "red",
                  args = list(mean = mu,
                              sd = samp.sd)) +
    labs(title = "Distribution of Sample Means From the Exponential
         Distribution \n(Overlaid with Normal Distribution)",
         x = "Sample Means",
         y = "Density")
```



This shows is that the distribution of the sample means is approximately normal.