Statistical Inference Course Project - Part 1

Jonathan Chang

February 28, 2017

Overview

In this project, we compare the exponential distribution in R with the Central Limit Theorem, which tells us "that the distribution of averages of iid variables becomes that of a standard normal as the sample size increases." We illustrate the following:

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

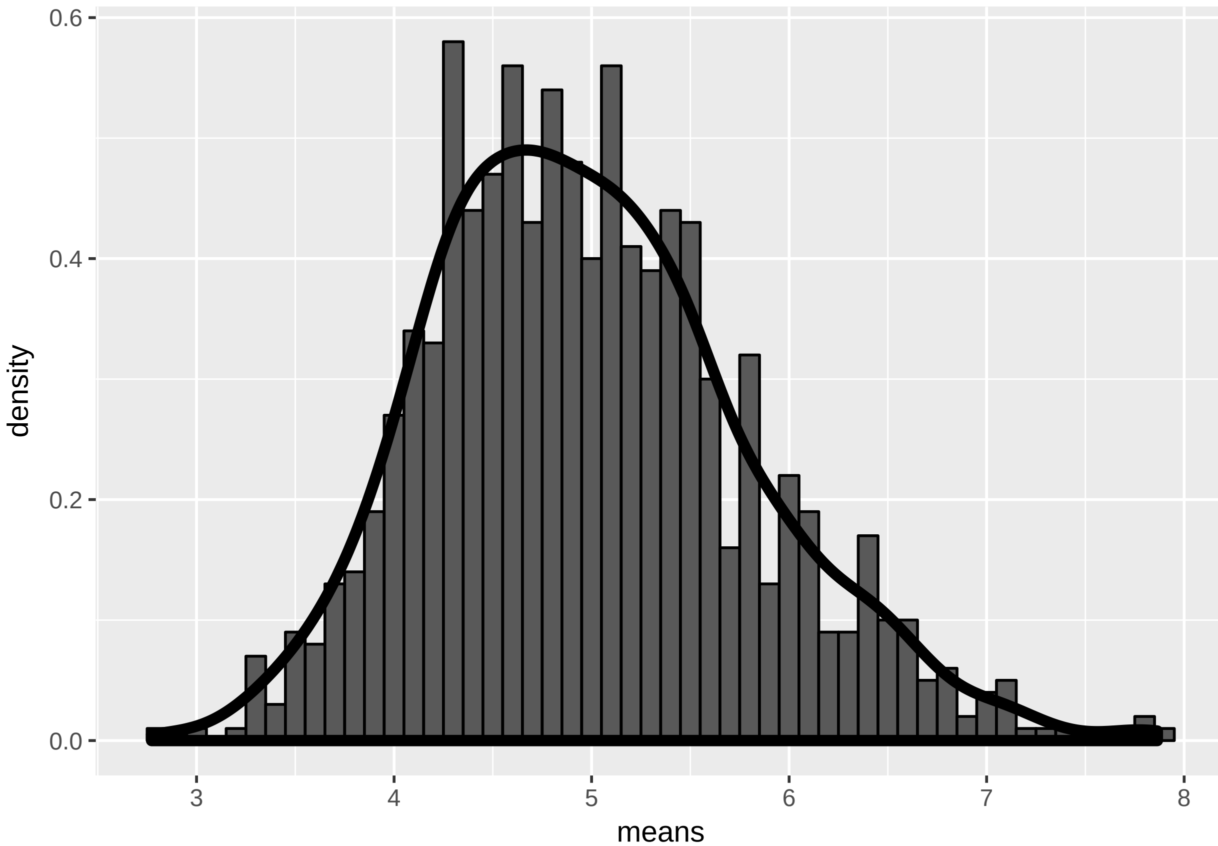
Simulations

We use lambda = 0.2 for all of the simulations, n = 40 to investigate the distribution of averages for 40 exponentials, and perform one thousand simulations using rexp(n, lambda).

```r
# load library for plotting
library(ggplot2)

# set seed for reproducibility
set.seed(973)

# set constants
lambda <- 0.2
n <- 40
num_sims <- 1000

# perform simulations and store in a dataframe
sim_means <- data.frame(means = sapply(1:num_sims, function(x) {mean(rexp(n, lambda))}))
```

Initial Plot of Simulation

```r
g <- ggplot(data = sim_means, aes(x = means)) +
    geom_histogram(aes(y = ..density..), binwidth=.1, color = "black") +
    geom_density(size = 2)
g
```



Sample Mean versus Theoretical Mean

The sample mean of our distribution:

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

```
## [1] 4.979909
```

The theoretical mean of our distribution is 1/lambda:

```
theoretical_mean <- 1/lambda
theoretical_mean
```

```
## [1] 5
```

The sample mean is fairly close to the theoretical mean.

Sample Variance versus Theoretical Variance

We use R to find the sample variance of our distribution:

```
sample_variance <- var(sim_means$means)
sample_variance
```

```
## [1] 0.6483053
```

The standard deviation of exponential distribution is 1/lambda, and standard deviation is the square root of variance. We also calculate the theoretical variance of our distribution:

```r
theoretical_variance <- ((1/lambda)^2)/40
theoretical_variance
```

```
## [1] 0.625
```

Because variance equals the standard deviation squared, our sample variance and theoretical variance are close, but still off.

Distribution

In the graph below, we plot the results of our simulation, which we already saw previously. We also superimpose a plot of the normal distribution as well vertical lines of the sample and theoretical means. The theoretical values are red and the experimental (simulation) results are blue. We can see that the distribution is approximately normal, but the Central Limit Theorem tells us that we could still do better with an even higher number of simulations.

```r
g <- ggplot(data = sim_means, aes(x = means)) +
    geom_histogram(aes(y = ..density..), fill = "lightgray", binwidth=.1) +
    geom_density(color = "blue", size = 2) +
    # superimpose the theoretical plot of the normal distribution
    stat_function(fun = dnorm, args = list(mean = theoretical_mean, sd = sqrt(theoretical_variance)), co
    geom_vline(xintercept = sample_mean, size = 0.5, color = "blue") +
    geom_vline(xintercept = theoretical_mean, size = 0.5, color = "red") +
    labs(title = "Simulation Plot w/ Normal Distribution Superimposed")
g
```