Statistical Inference Project - Part 1

Bryan Willauer

4/23/2020

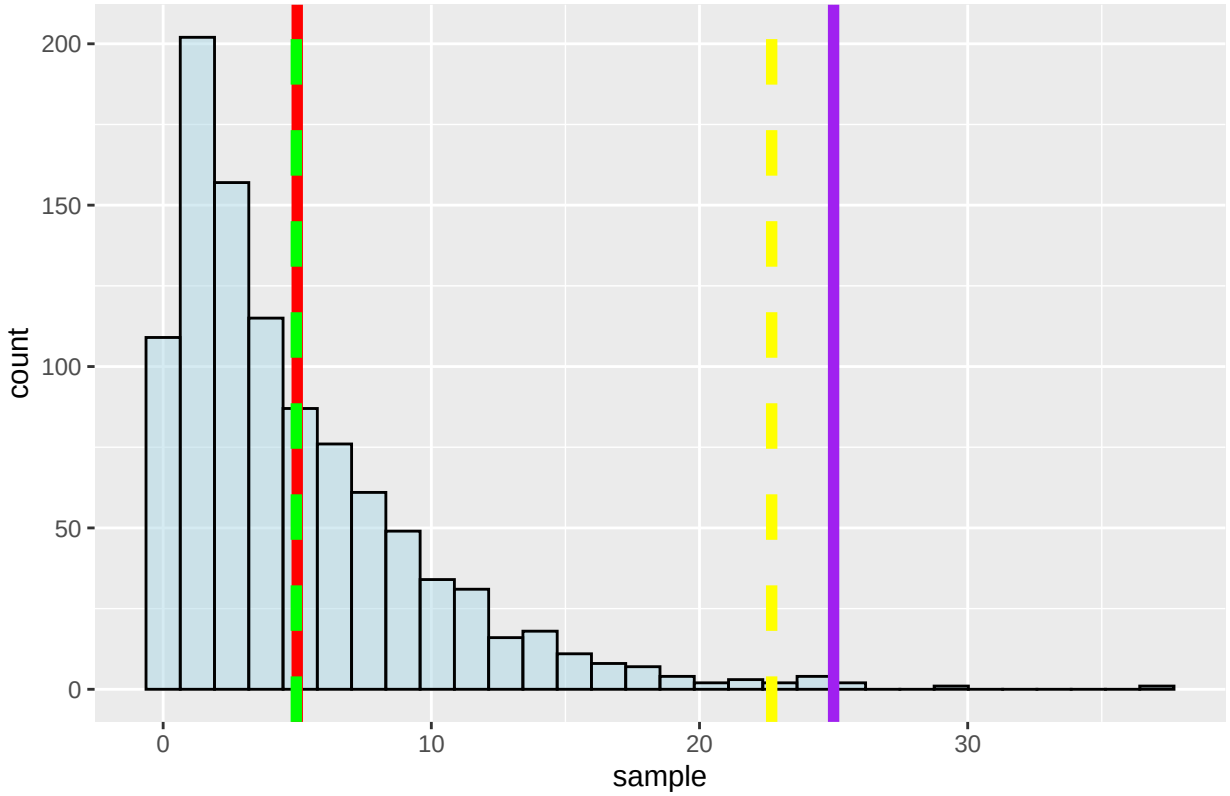
Overview

In this project, the exponential distribution will be compared to the Central Limit Theorem. This will be accomplished using simulations, graphs, and theoretical computations.

Simulations

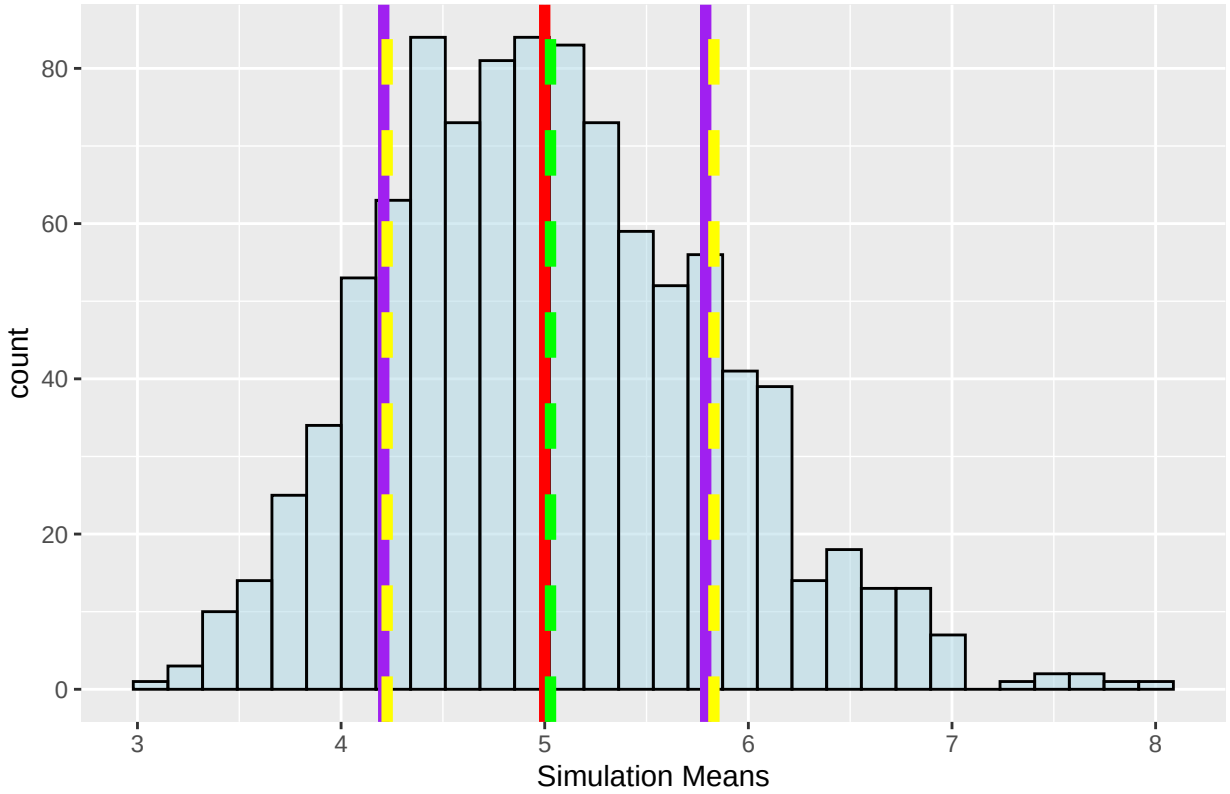
```
library(tidyverse)
set.seed(1968)
n <- 40
lambda <- 0.2
theoretical_mean <- 1/lambda
sd <- 1/lambda
variance <- (1/lambda)^2
sem <- sd / sqrt(n)

# take a sample of rexp variables
sample <- rexp(1000, lambda)

# report measured mean and variance
sample_mean <- mean(sample)
sample_var <- var(sample)

# histogram of 1000 randow exp variables
data <- data.frame(sample)
data %>% ggplot(aes(sample)) +
        geom_histogram(alpha = 0.5, col = "Black", fill = "lightblue") +
        geom_vline(size = 2,xintercept = theoretical_mean, colour = "Red", lty = 1) +
        geom_vline(size = 2,xintercept = sample_mean, colour = "Green", lty = 2) +
        geom_vline(size = 2,xintercept = variance, colour = "Purple", lty = 1) +
        geom_vline(size = 2,xintercept = sample_var, colour = "Yellow", lty = 2) +
        ggtitle("Figure 1: Histogram of 1000 rexp variables (lambda=0.2)")
```

Figure 1: Histogram of 1000 rexp variables (lambda=0.2)




```
# Histogram of means of 1000 40 sample exp variables
lambda <- 0.2   # lambda

n <- 40         # number of exponentials

sims <- 1000    # number of simulations

#Run simulations
sim_exp <- replicate(sims, rexp(n, lambda))

#Calc the means of the exponential simulations
mu_xbar <- apply(sim_exp, 2, mean)

data1 <- data.frame(mu_xbar)
data1 %>% ggplot(aes(mu_xbar)) +
        geom_histogram(alpha = 0.5, col = "Black", fill = "lightblue") +
        geom_vline(size = 2,xintercept = theoretical_mean, colour = "Red", lty = 1) +
        labs(x = "Simulation Means") +
        geom_vline(size = 2,xintercept = mean(mu_xbar), colour = "Green", lty = 2) +
        geom_vline(size = 2,xintercept = theoretical_mean + c(-1,1) * sem, colour = "Purple", lty = 1) +
        geom_vline(size = 2,xintercept = mean(mu_xbar)+ c(-1,1)*sd(mu_xbar), colour = "Yellow", lty = 2
        ggtitle("Figure 2: Histogram of 1000 averages of 40 rexp variables (lambda=0.2)")
```

Figure 2: Histogram of 1000 averages of 40 rexp variables (lambda=0.2)



Sample Mean versus Theoretical Mean

Figure #1 shows a distribution of 1000 random exponential variables. The solid red vertical line represents the theoretical mean of the distribution and the solid purple line represents the theoretical variance. The dashed green line is the sample mean and the dashed yellow is the sample variance.

The theoretical mean for an exponential distribution is 1 / lambda, and in our case lambda is 0.2, making the theoretical mean equal to 5. The sample mean in our case ended up as 4.965.

Figure #2 displays the results of running 1000 sets of 40 exponential variables and taking the mean of each of the 1000 sets. The colored vertical lines represent the theoretical and sample meansand variances as in figure #1.

The Central Limit Theorem states that the averages are centered around the theoretical mean. Our sample mean was 5.028, which is very close to the theoretical mean of 5.

Sample Variance versus Theoretical Variance

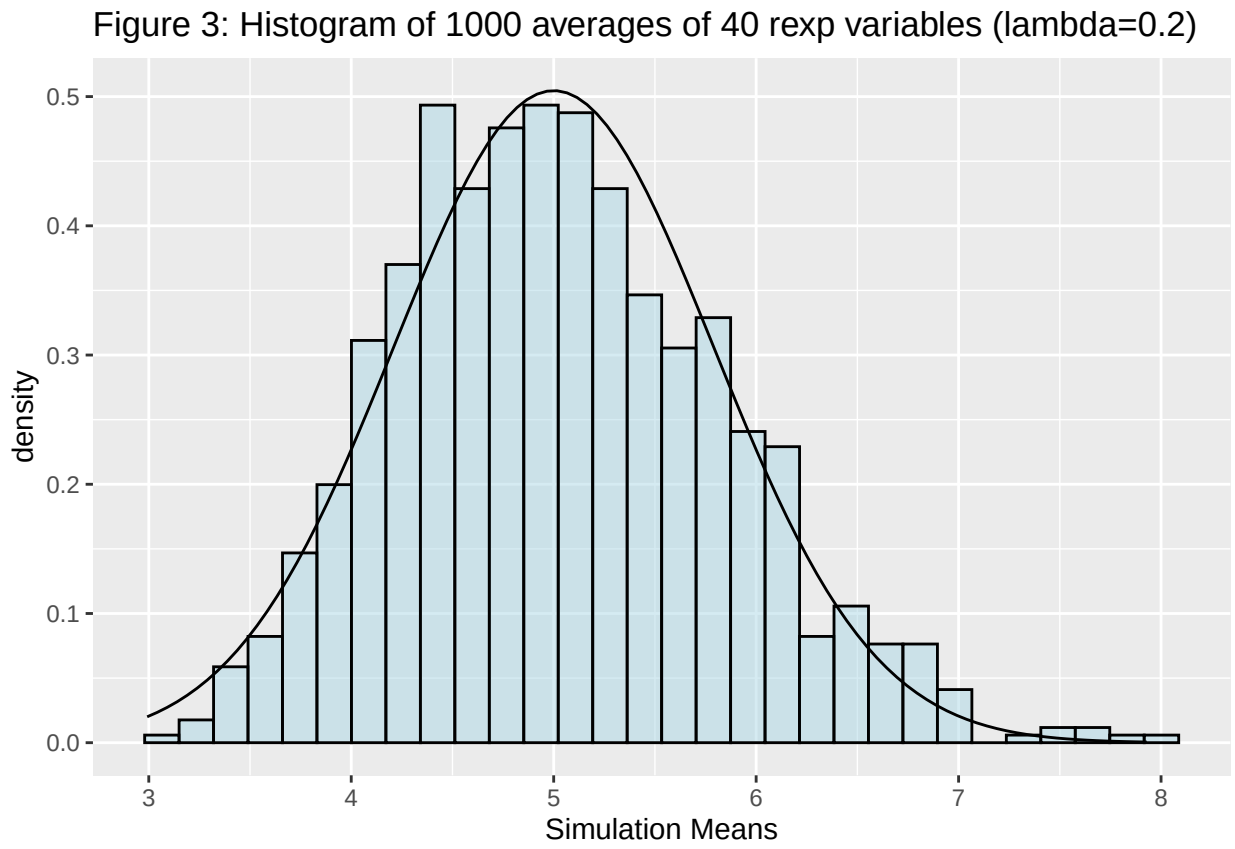
As stated previously, the yellow and purple vertical lines in t figures #1 and #2 show sample and theoretical variance, respectively. In figure #1, theoretical variance was calculated using the formula (1 / lambda) ^ 2, for a value of 25. The sample variance was 22.691.

In figure #2, the theoretical variance is defined by the formula (1/lambda) / sqrt(n) for a value of 0.791. The sample variance for the distribution was 0.802. The variances are all very close

Distribution

The Central Limit Theorem states that the average of the simulations should follow a normal distribution.

```
data1 %>% ggplot(aes(mu_xbar)) +
        geom_histogram(alpha = 0.5, col = "Black", fill = "lightblue", aes(y = ..density..)) +
        stat_function(fun = dnorm, args = list(mean = theoretical_mean, sd = sem)) +
        labs(x = "Simulation Means") +
        ggtitle("Figure 3: Histogram of 1000 averages of 40 rexp variables (lambda=0.2)")
```



In figure #3, a normal distribution curve was added on top of the histogram. The histogram fits closely to the normal distribution curve as stated by the Central Limit Theorem.