Statistical Inference Course Project: A simulation exercise.

Stephen Yang

August 7, 2019

Overview

This project is to investigate how the exponential distribution is to be compared with Central Limit Theorem. The analysis set lambda = 0.2 for all of the simulations. And we will look into the distribution of averages of 40 exponentials over 1,000 simulations.

Simulations

Set the lambda and the simulations variables

```
lambda <- 0.2
n <- 40
simulations <- 1000
```

Show the sample mean and compare it to the theoretical mean of the distribution.

Calculate the mean, standard deviation of the theoretical distribution then check the comparison with sample distribution. We can see there's a slight difference between the two distributions.

```
library(ggplot2)
theoretical_mean <- 1/lambda
theoretical_sd <- 1/lambda
exponential_distribution <- data.frame(x=apply(matrix(rexp((n*simulations), lambda),1000),1 ,mean))
mean_plot <- ggplot(exponential_distribution, aes(x=x)) +
  geom_histogram(binwidth=0.1, aes(y=..density..), color="blue", fill="white") +
  geom_density(alpha=0.2, fill="blue", color="blue") +
  geom_vline(aes(xintercept = theoretical_mean), color = "red", lty = "twodash")
```

AS we increase the amount of simulations, the distribution will be more likely centered at the central given the law of large number and the Center Limit Theorem.

Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

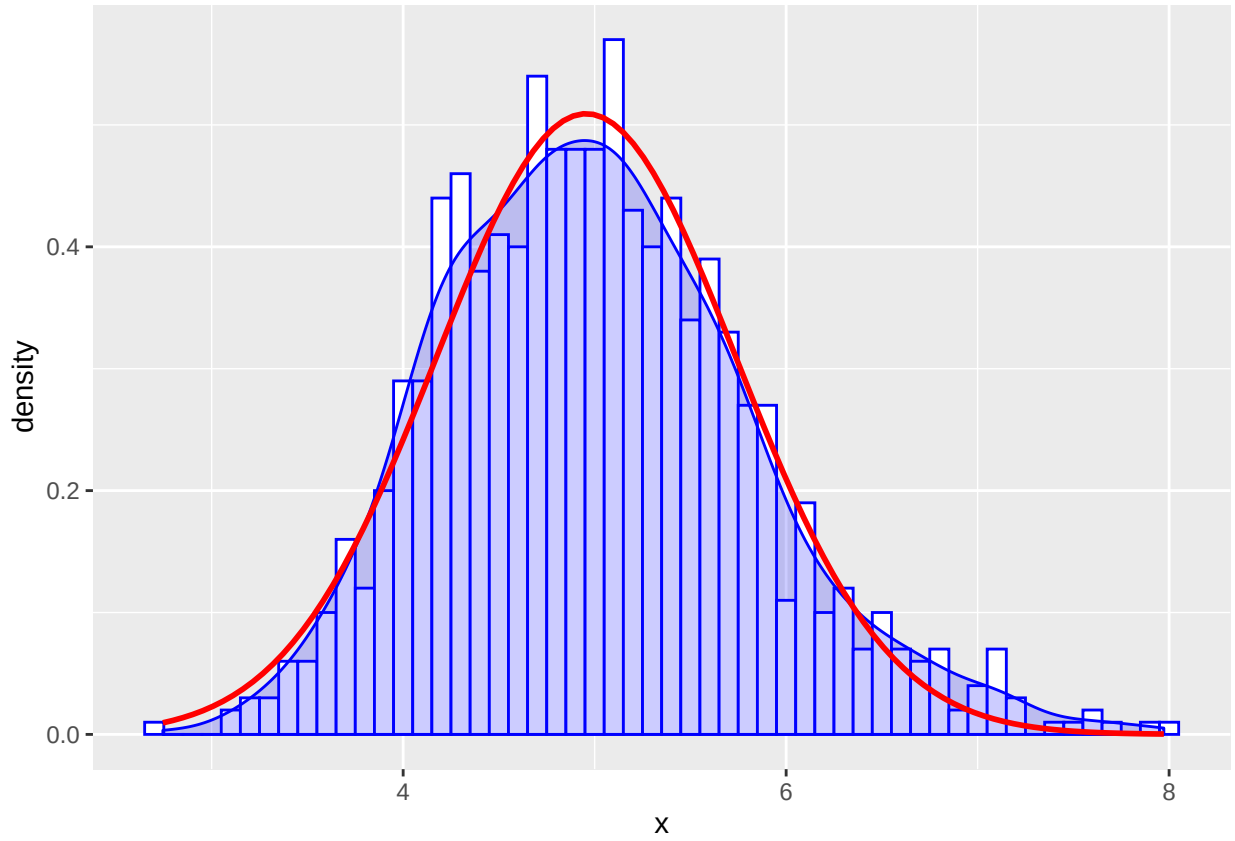
Fist thing is to calculate the theortical standard deviation,then we are able to get the variance. Also the standard deviation of the sample mean we have to caluclate.

```
theoretical_var <- (theoretical_sd/sqrt(n))^2
sd_simu <- sd(apply(matrix(rexp((n*simulations), lambda),1000), 1, mean))
md_simu <- mean(apply(matrix(rexp((n*simulations), lambda),1000), 1, mean))
var_simu <- sd_simu^2
```

It's pretty clear that the variance of the theoretical distribution is really close to the sample one.

#Show that the distribution is approximately normal. Last thing we have to do is to compare the two distribution directly. So i will plot two overlapping density plots on the same graph.

```
ggplot(exponential_distribution, aes(x=x)) +
  geom_histogram(binwidth=0.1, aes(y=..density..), color="blue", fill = "white") +
  geom_density(alpha=0.2, fill="blue", color="blue") +
  stat_function(fun=dnorm, args = list(mean=md_simu, sd=sd_simu), color="red", size=1)
```



Here is the answer, the sample mean distribution is approximately normal.