Statistical Inference - Comparison of Simulated Exponential Distribution to Central Limit Theorem

David Junor

August 20, 2016

Introduction

This is the first report of two for the Coursera Statistical Inference Course Project. In this report we are asked to simulate an Exponential Distribution and compare it to the Central Limit Theorem (CLT). the CLT states that the distribution of averages of iid variables (properly normalized) becomes that of a standard normal as the sample size increases. To do this we investigate how the arithmetic mean of exponentially distributed iid variables converges to a normal distribution.

In probability theory and statistics, the exponential distribution (a.k.a. negative exponential distribution) is the probability distribution that describes the time between events in a Poisson process, i.e. a process in which events occur continuously and independently at a constant average rate. The exponential distribution is specified by the single parameter lambda. Lambda is the event rate (Wikipedia: https://en.wikipedia.org/wiki/Exponential_distribution). The mean or expected value of an exponentially distributed random variable X with rate parameter lambda is given by 1/lambda and the theoretical value for the variance of the distribution of averages is given by lambda * sqrt(n))^-2.

Methods

Simulate and Describe the Exponenetial Distribution of of 40 samples simulated 1,000 iterations.

```
# Exponential Distribution Simulation
lambda <- 0.2
n <- 40
set.seed(2534)
ExpSampMeans <- NULL
for (i in 1:1000) {
        ExpSampMeans <- c(ExpSampMeans, mean(rexp(n, rate = lambda)))
}
```

Results and Discussion

Describe Simulated Exponentil Distribution

```
#Describe Simulated Exponential Distribution
mean1k <- mean(ExpSampMeans)
sd1k <- sd(ExpSampMeans)
var1k <- var(ExpSampMeans)
```

Sample Mean: 4.9946058
Sample Variance: 0.6328144

Describe Theoretical Exponential Distribution

```
#Calculate Mean and Variance of The Theoretical Exponential Distribution
mean     <- 1/lambda
variance <- (lambda * sqrt(n))^-2
```

Population Mean: 5
Population Variance: 0.625

```
#Plot Simulated Exponential Distribution
library(ggplot2)
data <- as.data.frame(ExpSampMeans)
ggplot(data, aes(x = data$ExpSampMeans)) +
geom_density() +
geom_vline(xintercept=mean1k, size = 1, color = "green") +
xlab("Sample Mean") + ylab("Frequency")
```

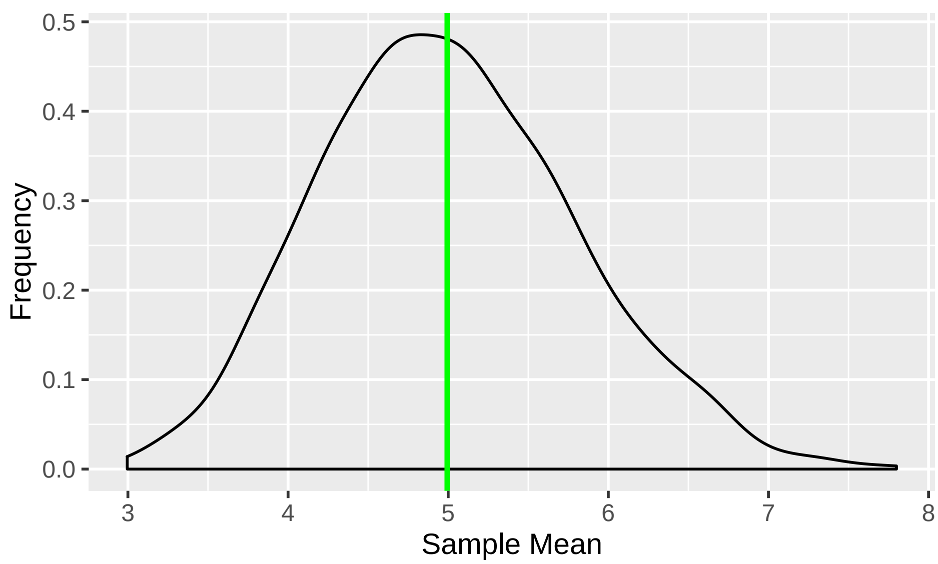

Figure 1. Distribution of the simulated sample means of the exponential distribution.

Compare Simulated Distribution of Sample Means to CLT

```
#Compare Simulated Exponential Distribution to Normal Distribution (CLT)
mu <- 1/lambda
sigma <- 1/lambda
ggplot(data=data, aes(data$ExpSampMeans)) + geom_histogram(binwidth = 0.2, color = "black", fill = "grey
stat_function(geom = "line", fun = dnorm, args = list(mean = mu, sd = sigma/sqrt(40)), size = 2, color =
xlab("Sample Mean") + ylab("Frequency")
```

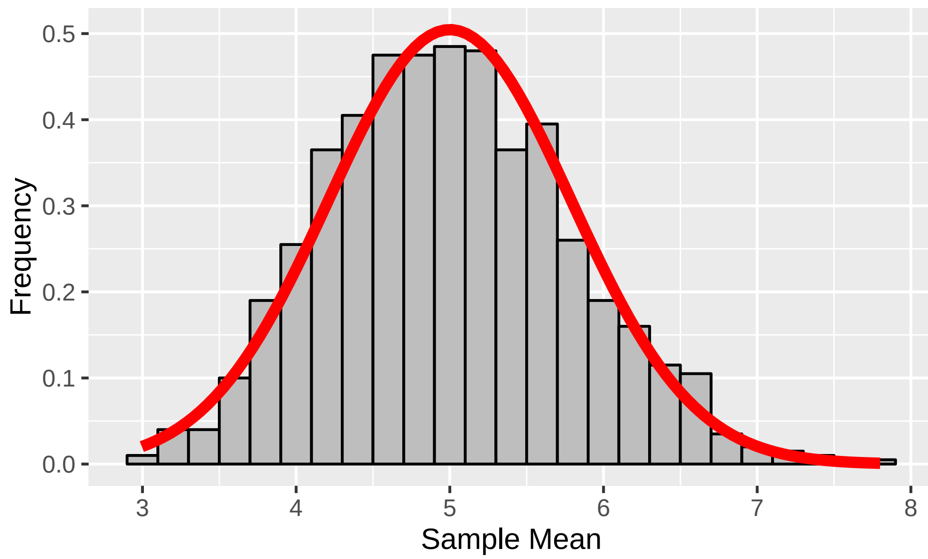

Figure 2. Comparison of the simulated exponential distribution of sample means to the normal distribution.

Test the Null Hypothesis of a Normal Distribution

```
Z <- abs((mean(data$ExpSampMeans)-mu)/(sigma/sqrt(40)))
p <- 2*pnorm(Z, lower.tail = FALSE)
c(Z,p)
```

```
## [1] 0.006823171 0.994555939
```

Given the large p-value and the unlikely event of a Type I error it is suggested that we fail to reject a Null Hypothesis that the distribution is normal and that indeed we are seeing an approximately normal distribution given the large n of the simulated exponential distribution.

Summary

The mean () and variance () of the simulated exponential distribution is comparable to the mean () and variance of the theoretical exponential distribution which indicates that indeed the simulated example does conform to precepts of the theoretical exponential distribution.

Comparison of the simulated exponential distribution of $n = 40$ samples iterated 1000 times to the theoretical normal distribution suggests that CLT holds true given the fairly large sample size and the 1000 iterations of the simulated exponential distribution. Indeed, we see that as n increases the simulated distribution approaches a normal distribution. A Null Hypothesis of a normal distribution is rejected given the large p-value and we see that under the CLT our simulated exponential distribution approximates a normal distribution given the relatively large sample size.