

Statistical Inference Course Project

Kurt Fitz

April 21, 2019

Overview

The following report is divided into two parts. The first explores the exponential distribution in relation to the Central Limit Theorem, and the second involves some basic inference using the ToothGrowth data in R.

Part 1 - Simulation Exercise

The Exponential distribution is simulated and compared with the Central Limit Theorem. The simulation is performed 1000 times with a sample size $n = 40$

Sample and theoretical mean

Following we see the distribution of 1000 exponentials:

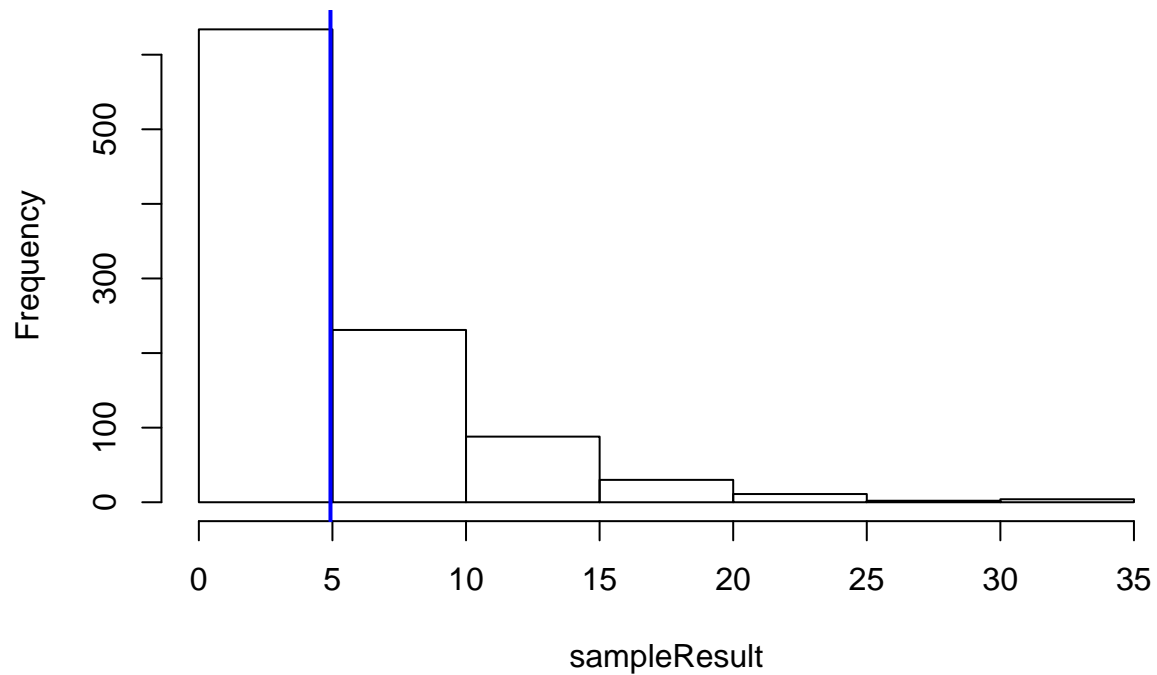
```
set.seed(33)
lambda <- 0.2 # The rate parameter
n <- 40 # sample size

sampleResult <- (rexp(1000, lambda))

hist(sampleResult, main = "distribution of 1000 exponentials")

sampleMean <- mean(sampleResult)
abline(v = sampleMean, col = "blue", lwd = 2)
```

distribution of 1000 exponentials



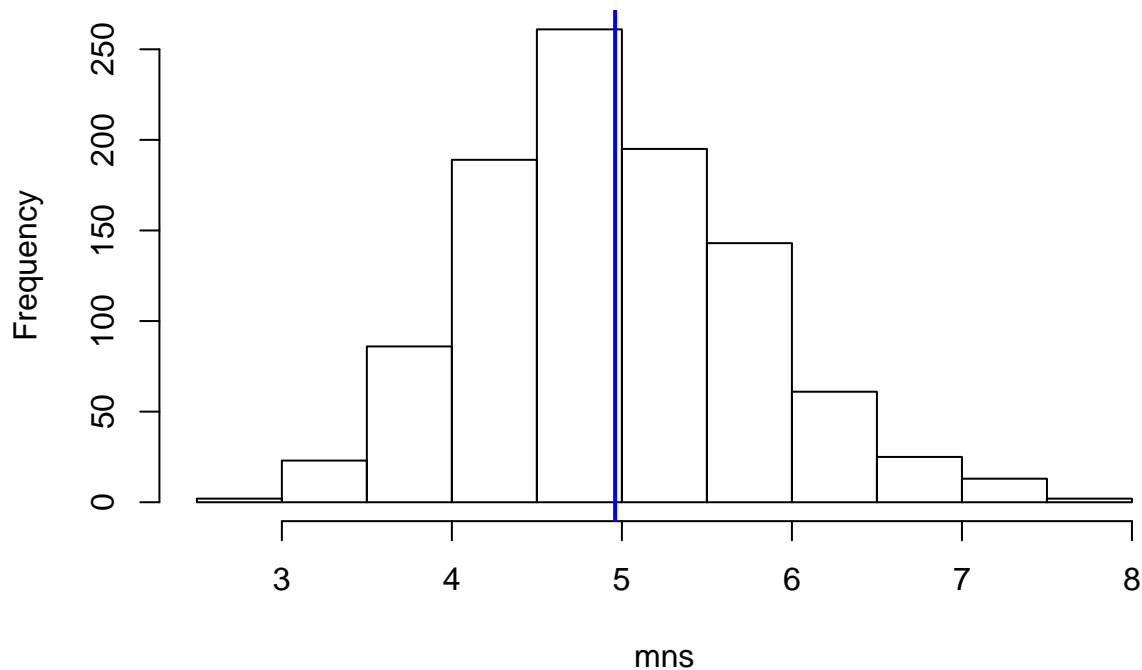
Next, there is the distribution of means of 40 exponentials sampled 1000 times.

```
lambda <- 0.2 # The rate parameter
n <- 40 # sample size
mns = NULL

for (i in 1 : 1000) mns = c(mns, mean(rexp(n, lambda)))
hist(mns, main = "Simulated Dist of means n=40 1000 times")

simMean <- mean(mns)
abline(v = simMean, col = "blue", lwd = 2)
```

Simulated Dist of means n=40 1000 times



```
sampleMean <- mean(sampleResult)
sampleVariance <- var(mns)
sampleStdDev <- (1/lambda)/sqrt(n)

simMean <- mean(mns)
simVariance <- (1/lambda)^2/n
simStdDev <- (1/lambda)/sqrt(n)
```

The sample mean from above is 4.9265823 compared to a simulated theoretical mean of 4.9601793. Notice how the distribution is centered around 5. When we compare the distribution of means from our 40 exponentials simulated 1000 times, we can see that the sample mean from above tends to be centered around the same point.

Sample variance and Standard deviations and theoretical variance of distribution

The variance of sample distribution is 0.6513642, while the variance of the means of the simulated samples is 0.625. The sample standard deviation is 0.7905694 and the theoretical standard deviation is 0.7905694.

Normality of distribution

The following graph demonstrates that the simulated distribution of the means of 40 exponentials (1000 times) approximates a normal distribution. This is consistent with the Central Limit Theorem in that the distribution of the averages of samples are normally distributed.

```
hist(mns, prob=TRUE, col="green", main="", breaks=10)  
lines(density(mns), lwd=3, col="blue")
```

