

Statistical Inference Course Project (Part 1)

Tong Tat

December 22, 2017

Rpub Link: [Click Here](#)

Part 1

1.1. Overview

This is a document for Coursera Statistical Inference Course Final Project. This project will investigate the exponential distribution in R and compare it with the Central Limit Theorem. Given that $\lambda = 0.2$ for all of the simulations. Part 1 of the project will investigate the distribution of averages of 40 exponentials over a thousand simulations.

1.2. Simulations

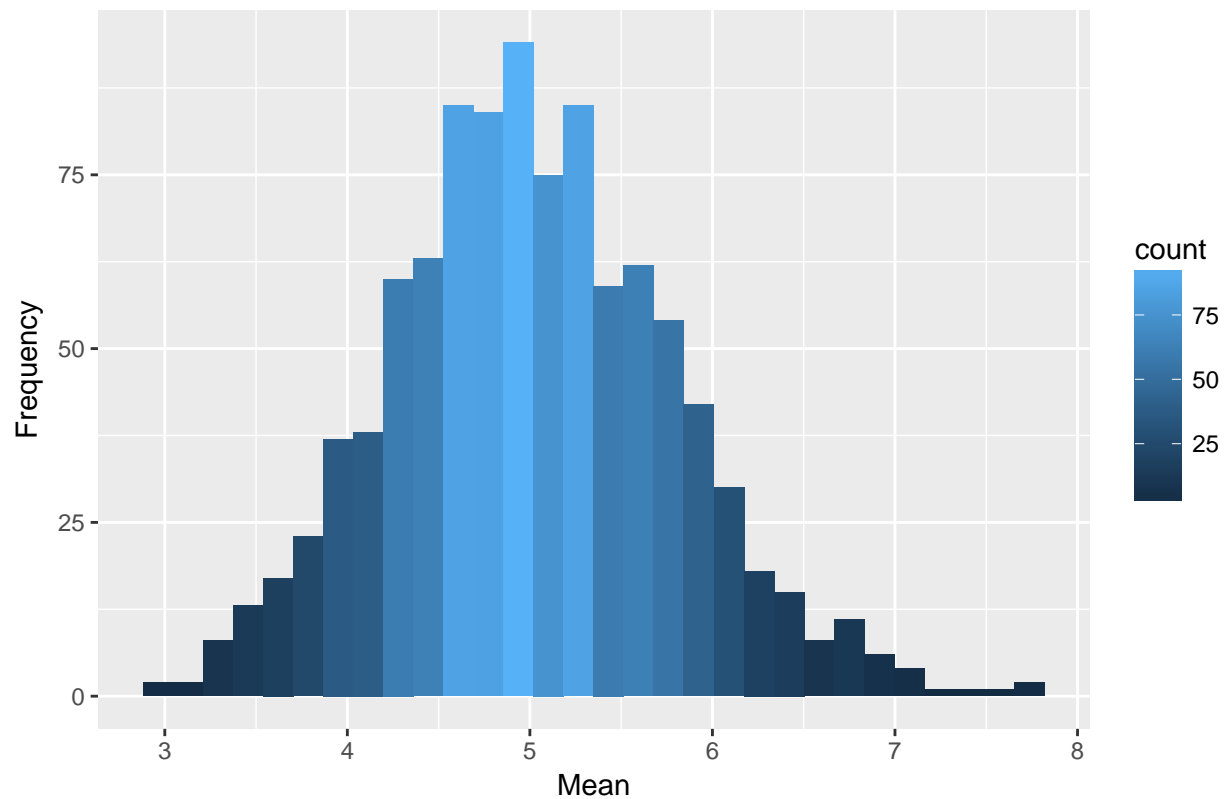
```
# Using pre-defined parameters
lambda <- 0.2
n <- 40
sims <- 1:1000
set.seed(123)

# Check for missing dependencies and load necessary R packages
if(!require(ggplot2)){install.packages('ggplot2')}; library(ggplot2)

# Simulate the population
population <- data.frame(x=sapply(sims, function(x) {mean(rexp(n, lambda))}))

# Plot the histogram
hist.pop <- ggplot(population, aes(x=x)) +
  geom_histogram(aes(y=..count.., fill=..count..)) +
  labs(title="Histogram for Averages of 40 Exponentials over 1000 Simulations", y="Frequency", x="Mean")
hist.pop
```

Histogram for Averages of 40 Exponentials over 1000 Simulations



1.3. Sample Mean versus Theoretical Mean

As we can see below, both sample mean and theoretical mean are very close.

```
# Tabulating the Sample Mean & Theoretical Mean
sample.mean <- mean(population$x)
theoretical.mean <- 1/lambda
cbind(sample.mean, theoretical.mean)

##      sample.mean theoretical.mean
## [1,]      5.011911              5

# Checking 95% confidence interval for Sample Mean
t.test(population$x)[4]

## $conf.int
## [1] 4.963824 5.059998
## attr(,"conf.level")
## [1] 0.95
```

At 95% confidence interval, the sampled mean is between 4.9638242 and 5.0599984.

1.4. Sample Variance versus Theoretical Variance

As we can see below both Sample Variance and Theoretical Variance are very close.

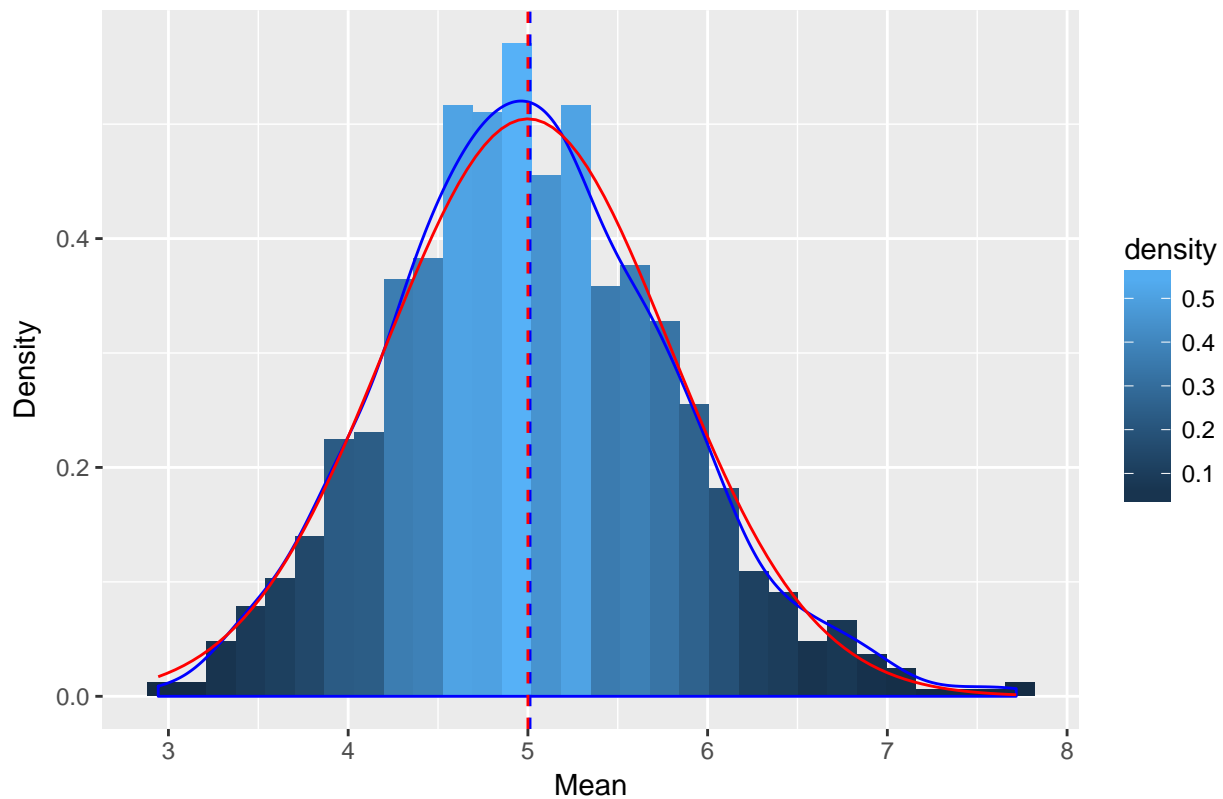
```
sample.variance <- var(population$x)
theoretical.variance <- ((1/lambda)^2)/n
cbind(sample.variance, theoretical.variance)
```

```
##      sample.variance theoretical.variance
## [1,]      0.6004928           0.625
```

1.5. Distribution

```
# Plotting Sample Mean & Variance vs Theoretical Mean & Variance
gg <- ggplot(population, aes(x=x)) +
  geom_histogram(aes(y=..density.., fill=..density..)) +
  labs(title="Histogram of Averages of 40 Exponentials over 1000 Simulations", y="Density", x="Mean") +
  geom_density(colour="blue") +
  geom_vline(xintercept=sample.mean, colour="blue", linetype="dashed") +
  stat_function(fun=dnorm, args=list(mean=1/lambda, sd=sqrt(theoretical.variance)), colour="red") +
  geom_vline(xintercept=theoretical.mean, colour="red", linetype="dashed")
gg
```

Histogram of Averages of 40 Exponentials over 1000 Simulations



As we can see, the Sampled mean for 40 exponentials simulated 1000 times are very close to the Theoretical mean for a normal distribution.

Please note the assumptions is we are sampling without replacement and set.seed is at 123.