Assignment1_part1

Haoyi Wei

2022-12-19

Investigate the exponential distribution in R and compare it with the Central Limit Theorem.

Overview

This project investigates the exponential distribution in R and compare it with the Central Limit Theorem.

Part1: Simulation Exercise

```
#set seed for reproducability
set.seed(123)
```

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

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

```
# set the parameters per the quest of the assignment
lambda <- 0.2 # lambda
n <- 40 # number of exponentials
sim <- 1000 # number of simulations

# run the simulation
sim.exp <- replicate(sim, rexp(n, lambda))

# calculate mean of exponential simulations
mean.exp <- apply(sim.exp, 2, mean)
```

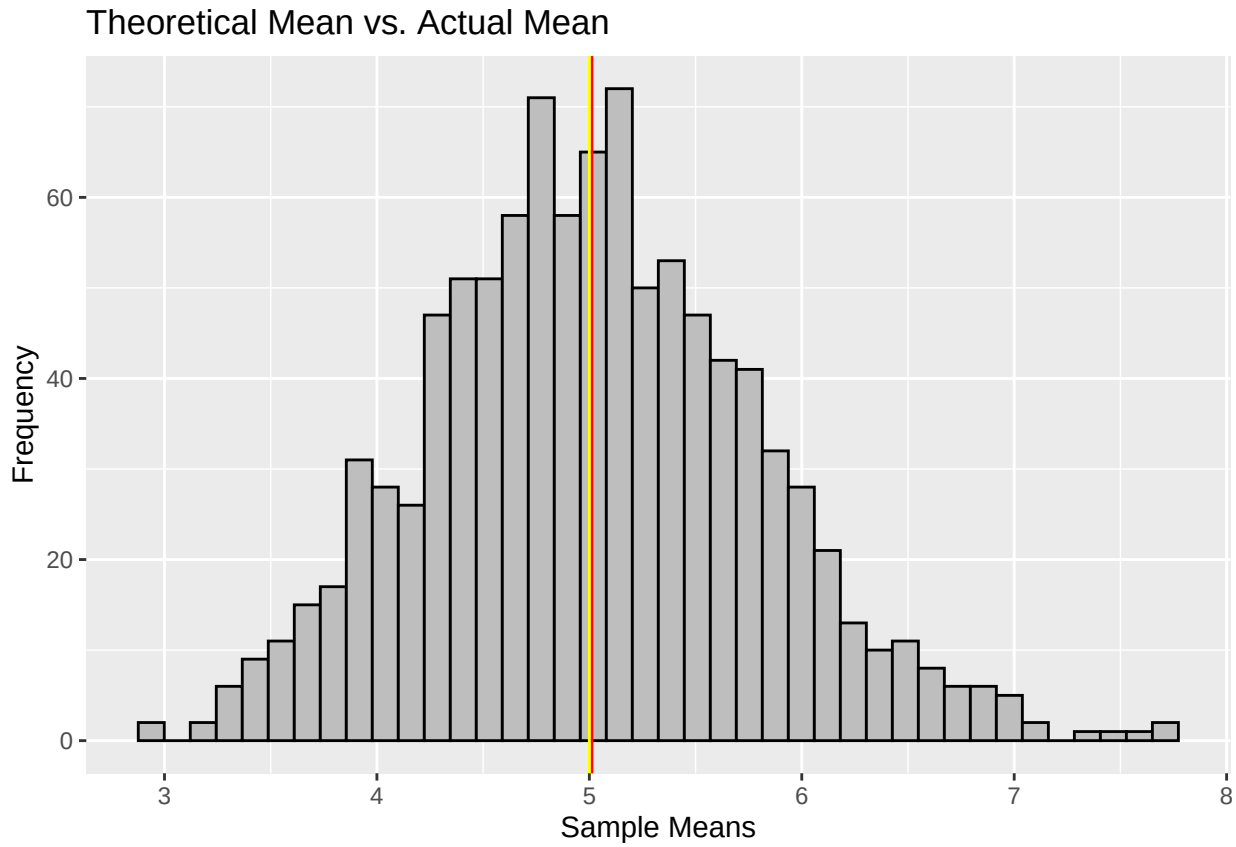
```
# mean of sample mean
mm.exp <- mean(mean.exp)
mm.exp
```

```
## [1] 5.011911
```

```r
# Theoretical Mean
1/lambda
```

```
## [1] 5
```

```r
# distribution of the sample mean
library(ggplot2)
meanexp <- data.frame(mean.exp)
ggplot(meanexp, aes(x=mean.exp)) +
          geom_histogram(bins=40, color="black", fill="grey") +
          geom_vline(aes(xintercept=mm.exp),color="red") +
          geom_vline(aes(xintercept=1/lambda),color="yellow") +
          labs(x="Sample Means", y="Frequency") +
          ggtitle("Theoretical Mean vs. Actual Mean")
```



Sample mean and theortical mean is very similar

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

```r
# Theoretical Variance
(1/lambda)^2 / n
```

```
## [1] 0.625
```

```
# Sample Variance
sd(mean.exp)^2
```

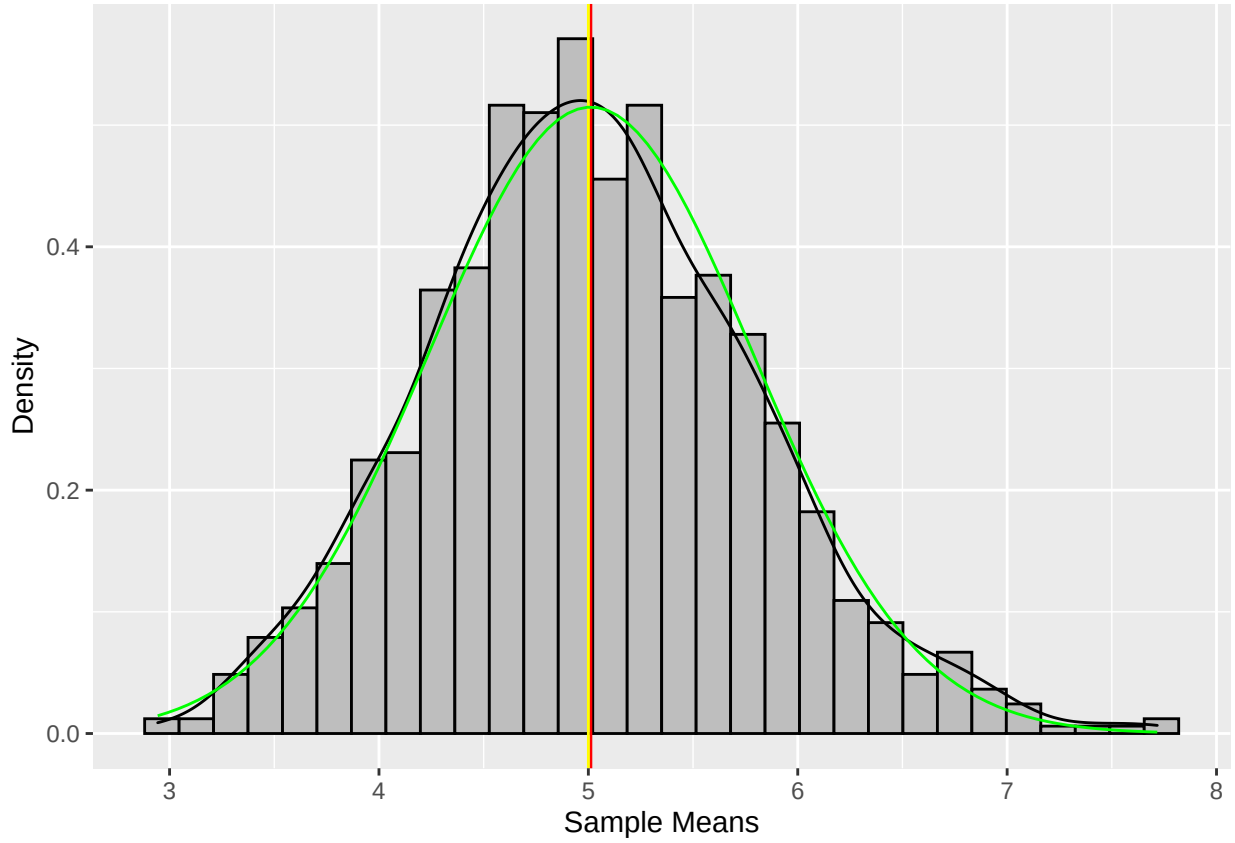
```
## [1] 0.6004928
```

Theoretical variance is close to sample varaince

3. Shoow that the distribution is approximately normal

```
ggplot(meanexp, aes(x=mean.exp)) +
          geom_histogram(aes(y=..density..), color="black", fill="grey") +
          geom_density() +
          geom_vline(aes(xintercept=mm.exp),color="red") +
          geom_vline(aes(xintercept=1/lambda),color="yellow") +
          labs(x="Sample Means", y="Density") +
          stat_function(fun=dnorm, color="green",args=list(mean=mm.exp,sd=sd(mean.exp)))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



The distribution is approximately normal