Statistical inference Assignment Part 1

Aravind S

24/03/2020

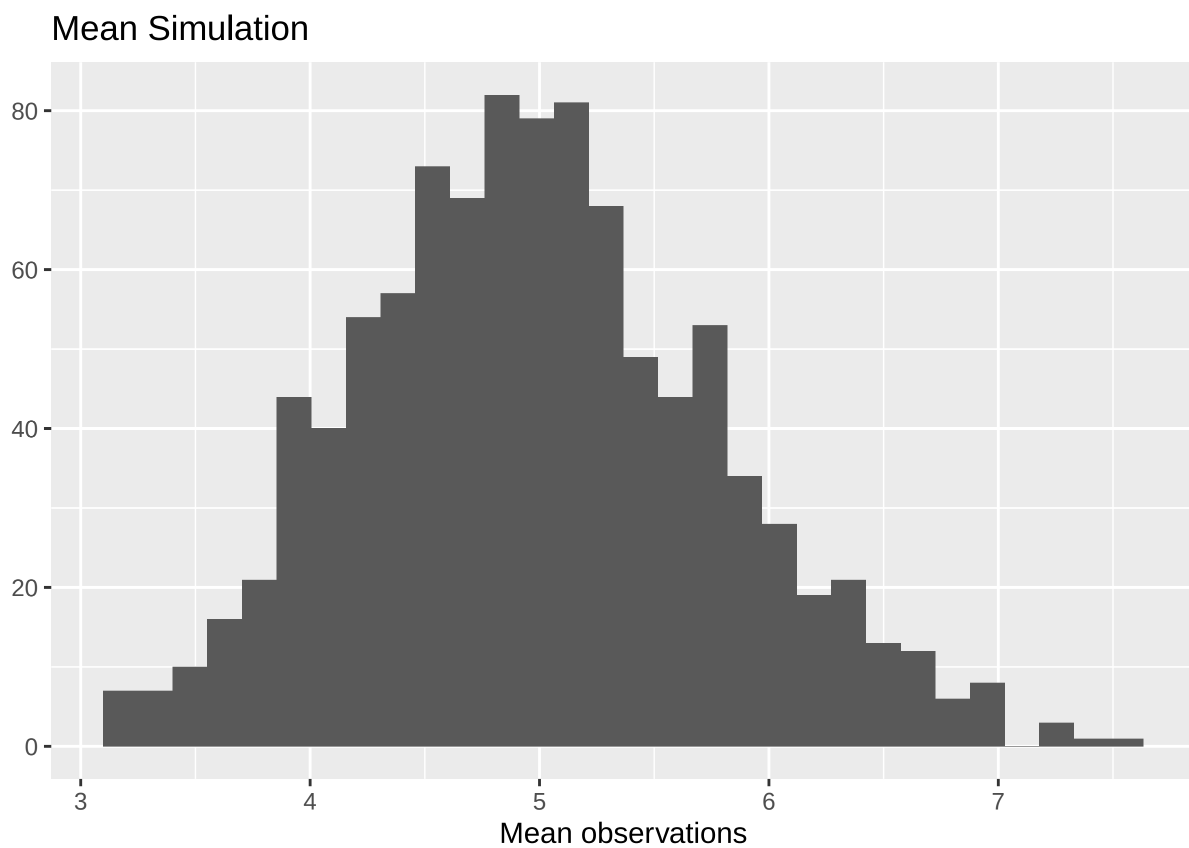
PROJECT DESCRIPTION

This project investigates the exponential distribution in R and compares it with the Central Limit Theorem. The exponential distribution is 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.

The lambda value is fixed at 0.2 for all of the simulations. The distribution of averages of 40 exponentials is investigated by simulating it a 1000 times.

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

```r
library(ggplot2)
set.seed(1) ## to make this reproducable
n <- 40
lambda <- 0.2
simulation <- replicate(1000, rexp(n, .2))
mean_simulation <- apply(simulation, 2, mean)
qplot(mean_simulation, geom = "histogram",
      main = "Mean Simulation",
      xlab = "Mean observations")
```




```
mean2 <- mean(mean_simulation)
print(mean2)
```

```
## [1] 4.990025
```

The theoretical mean is equal to 1/lambda

```
Tmean <- 1/0.2
print(Tmean)
```

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

The theoretical mean and sample mean are shown to be very similar.

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

The expected standard deviation is equal to (1/lambda)/sqrt(n) or Esd. The variance is just the standard deviation squared.

```
Esd <- (1/.2)/sqrt(n)
Evar <- Esd^2
```

We can then calculate the standard deviation and variance of the sample.

```
Ssd <- sd(mean_simulation)
Svar <- var(mean_simulation)
```

The standard deviation of the sample is Ssd and the variance of the sample is Svar. These are both close approximations of the expected standard deviation (Esd) and the expected variance (Evar).

3. Show that the distribution is approximately normal.

The graph shows the simulation means distribution with the normal distribution curve on top. Illustrating that it fits relatively well.

```
fit <- seq(min(mean_simulation), max(mean_simulation), length=100)
standard_fit <- dnorm(fit, mean=Tmean, sd=Esd)
hist(mean_simulation, breaks = n, prob=T, xlab = "means", ylab = "count", main = "Density of Means")
lines(fit, standard_fit, pch=2, col="red", lty=5)
```