Statistical Inference Project Part 1 - Saran Shanmugam

Overview

This Coursera-Statistical Inference(Data Science Specialization) project illustrates 1. A Simulation exercise. 2. Basic Inferential Data Analysis.

Synopsis

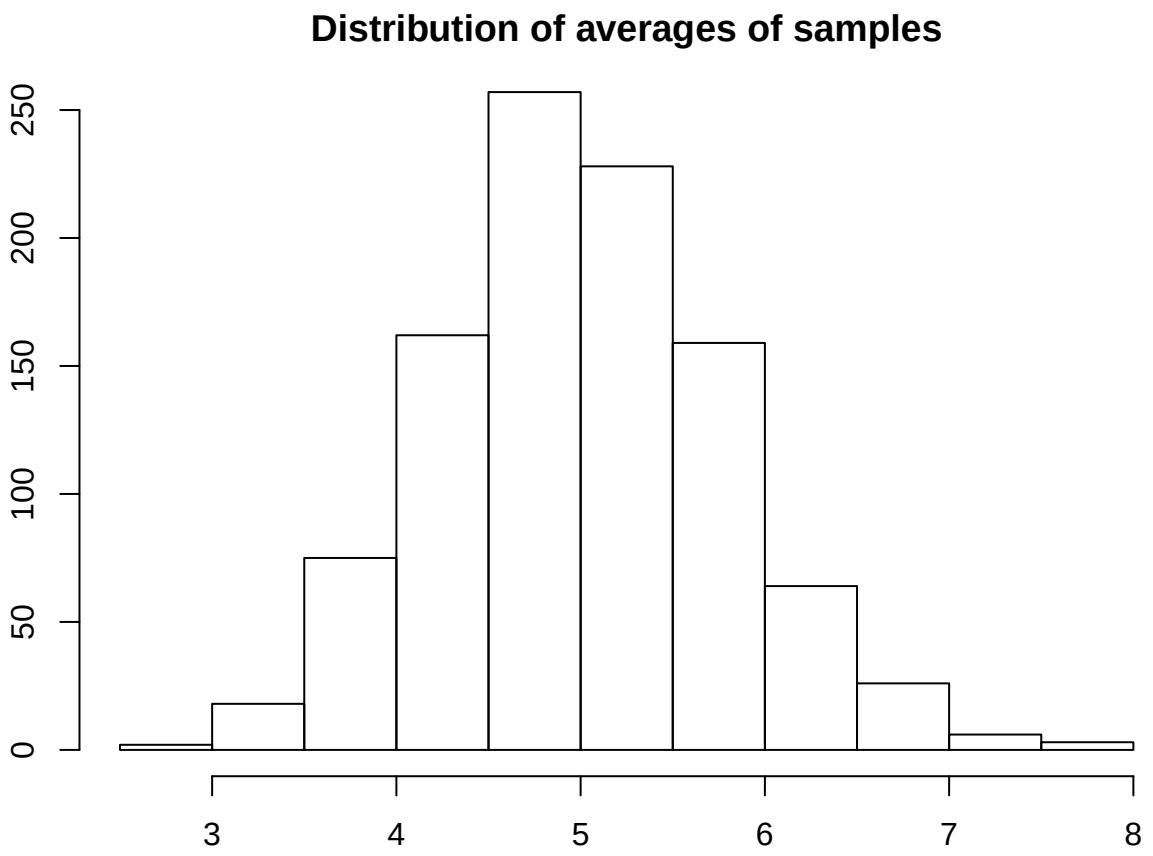
The exponential distribution in R here is demonstrated using rexp() function with exponentials of 40 and lambda of 0.2 for a 1000 observations.

Results

Simulations

Lets generate 1000 simulated averages for 40 exponentials and lambda of 0.2 using rexp()

```
library(ggplot2)
library(knitr)
numsim <- 1000
lambda <- 0.2
sample <- 40
set.seed(123)
mns=NULL
for(i in 1:1000) mns=c(mns,mean(rexp(sample,lambda)))
par(mar=rep(2,4))
hist(mns,main="Distribution of averages of samples")
```



1. Sample Mean versus Theoretical Mean

The next step is comparing the theortical and sample mean

```
actual_mean <- round(mean(mns),3)
theory_mean <- 1/lambda
print(paste("Sample Mean=",actual_mean))
```

```
## [1] "Sample Mean= 5.012"
```

```
print(paste("Theoretical Mean=",theory_mean))
```

```
## [1] "Theoretical Mean= 5"
```

Looking at both the means,they are approximately the same.

2. Sample Variance versus Theoretical Variance

The next step is comparing the theortical and sample variance

```
actual_var <- round(var(mns),3)
theory_var <- (1/lambda)^2/sample
print(paste("Sample Variance=",actual_var))
```

```
## [1] "Sample Variance= 0.6"
```

```
print(paste("Theoretical Variance=",theory_var))
```

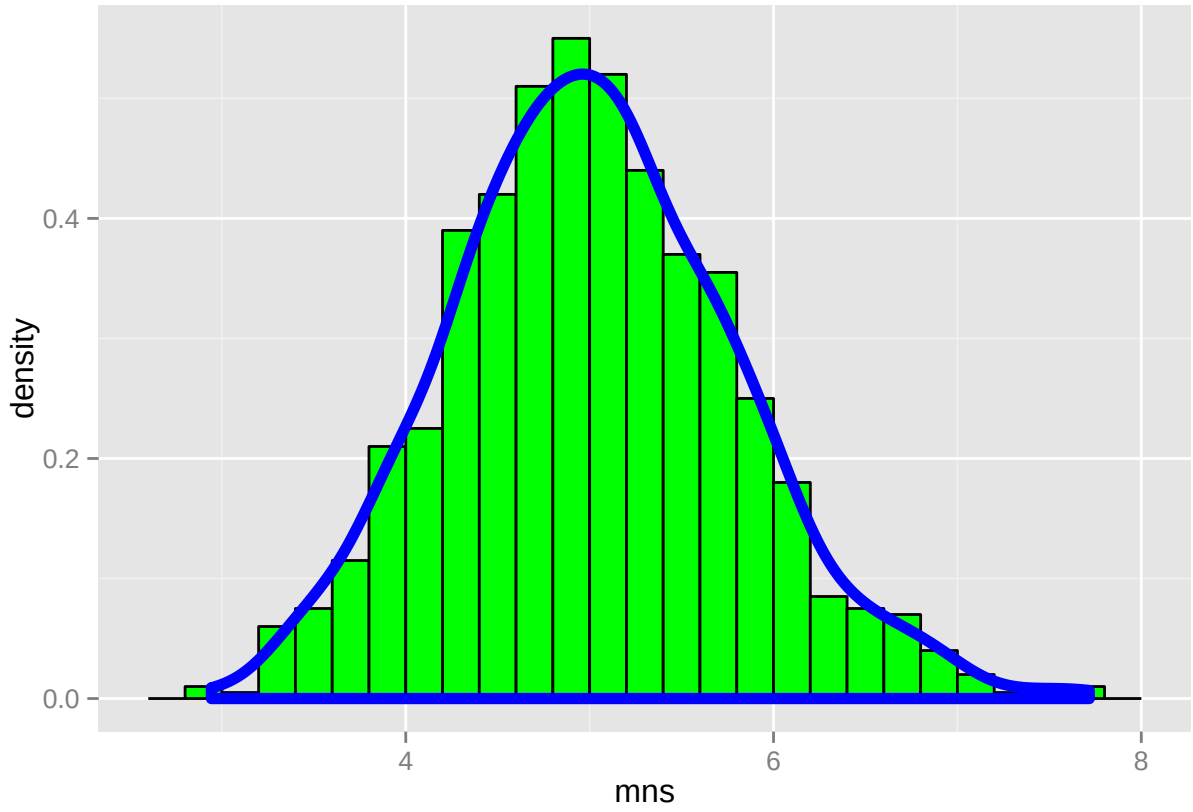
```
## [1] "Theoretical Variance= 0.625"
```

From above, it can be infered that the Sample and Theoretical variances are almost the same.

3. Demonstration that the distribution is approximately normal

The last step is demonstrating that the distribution is approximately normal

```
mns_df <- data.frame(mns)
gg <- ggplot(mns_df,aes(x=mns))
gg <- gg + geom_bar(aes(y=..density..),colour="black",fill="green",binwidth=.2)
gg <- gg + geom_density(colour="blue",size=2)
gg
```



The above plot depicts a distribution of approximately normal