

Stat Inference Part 1

Mohit Anand

Overview

This is the part 1 submission of Statistical Inference Project. In this the distribution of averages of 40 exponentials with a thousand simulation is analyzed. The distribution obtained is compared with Central Limit Theorem. The following things are shown: 1. The Sample mean compare it to theoretical mean 2. The variance of sample and compare it to the theoretical variance of the distribution 3. Distribution is approximately normal

Simulation

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.4.3
set.seed(2456)
n<-40
lambda<-0.2
sim<-1000
sim_sample<-matrix(rexp(n*sim, rate = lambda),sim)
sim_mean<- apply(sim_sample,1,mean)
```

Sample Mean Comparison

```
samp_mean <- mean(sim_mean)
```

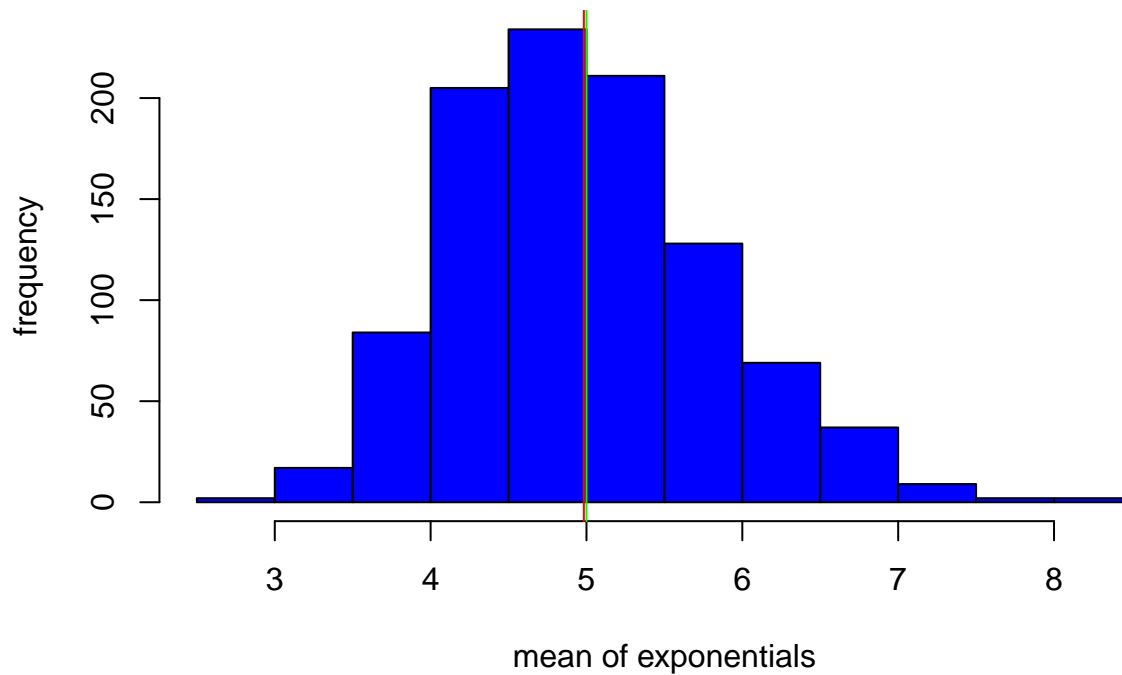
```
theo_mean <- 1/lambda
```

Here we see that the two values of Sample mean and Theoretical mean are almost equal.

The theoretical mean is shown as Green line and the sample mean as Red.

```
hist(sim_mean, xlab="mean of exponentials", ylab="frequency", main="Histogram of mean of 40 exponential.
abline(v = theo_mean, col="green")
abline(v = samp_mean, col="red")
```

Histogram of mean of 40 exponentials



Variance Comparison

Sample's Variance and standard deviation

```
#Standard deviation
sim_sd<- sd(samp_mean)

#Variance of Simulation
sim_var<-sim_sd^2
```

Theoretical Variance and Deviation

```
#Theoretical deviation
theo_sd <- (1/lambda)/sqrt(n)

#Theoretical Variance
theo_var<- theo_sd^2
```

The Variance and standard deviation values are nearly equal

The plot approximatly shows a normal distribution

```
df <- data.frame(sim_mean)
plot_data <- ggplot(df,aes(x= sim_mean))
plot_data <- plot_data + geom_histogram(aes(y=..density..),fill ="green",color = "black") +
  geom_density(colour="red")
print(plot_data)
```

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

