Statistical Inference Course Project

Shruti Pareek

Overview

This Project is dvided in two parts, first part will showcase a simulation exercise and second part will showcase the basic inferential data analysis This is an R Markdown document.

Part 1

Problem statement is to explore exponential distribution and then compare it with Central Limit Theory. A function rexp(n,lambda) will be used, with lambda = 0.2, a sample size of 40 and 1000 simulation.

Simulations

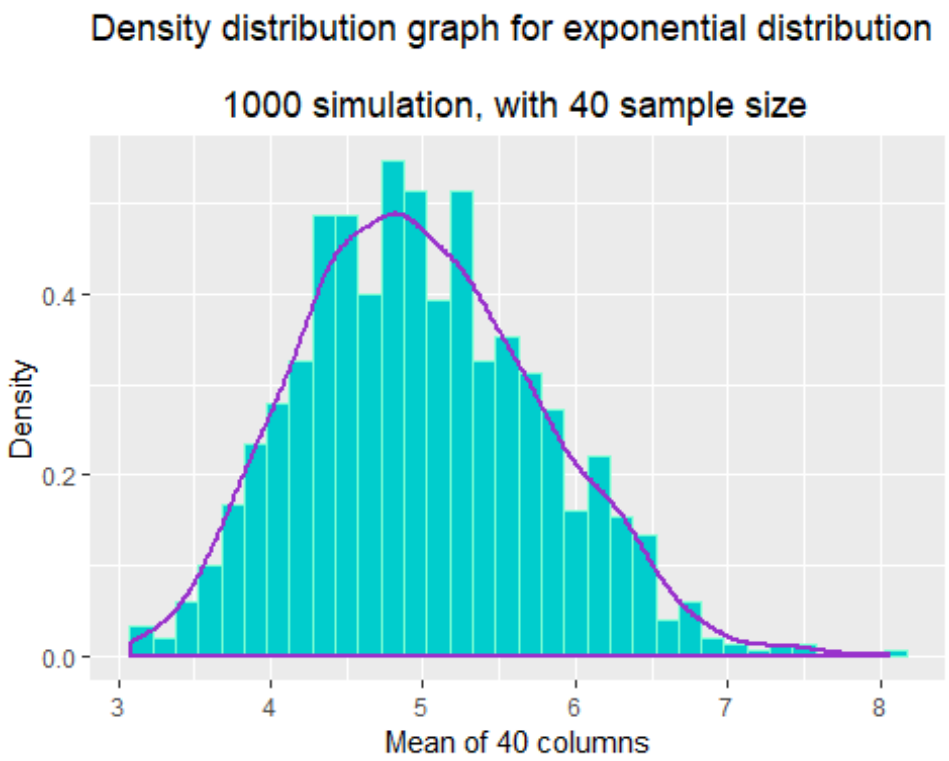
After loading the relevant libraries, variables such as lambda, sample size and simulation size have been intialized. In order to use the random exponential function, a seed with value 1 has been set. Post this simulation data has been created and shown. The simulated data has 1000 rows and 40 columns, a mean of the simulation data per ro has been caluculated and plot to show te density distribution.

```
#importing libraries required for this program
library(ggplot2)

# initializing the lambda, sample size and simulation size
lambda<-0.2;
sample_size<-40;
sim_size<-1000;

# setting the seed to 1
set.seed(1)

# creating simulation data, using the function rexp for exponential distribution
sim_data<-matrix(data=rexp(sample_size*sim_size, rate=lambda), nrow = sim_size,
ncol = sample_size)

# checking the number of rows and columns in simulation data
dim(sim_data)
```

```
## [1] 1000   40
```

```
# Calculating the mean per row for the simulation data
mean_per_row<-apply(sim_data,1,mean)

# Plotting the density distribution graph of mean values of simulated exponential
distributions
den_graph<-ggplot(data.frame(mean_per_row),aes(x=mean_per_row))
den_graph<-den_graph+ ggtitle("Density distribution graph for exponential
distribution \n
             1000 simulation, with 40 sample size")
den_graph<-den_graph+geom_histogram(binwidth =
0.15,colour="aquamarine",fill="cyan3",aes(y = ..density..))
```

```
den_graph<-den_graph+geom_density(colour="darkorchid",size=1.01)+labs(x="Mean of 40
columns",y="Density")
print(den_graph)
```



Sample Mean versus Theoretical Mean and Sample Variance versus Theoretical Variance

In this section, distribution of simulation data has been compared with central limit theory. First the statistics of normal distribution, such as mean, standard error and variance have been computed. Similar statistics have been computed for distribution of simulated data. Post this a table has been created for better visulization and side by side comparison of statistics. In the table it can be observed that theorticl mean is 5 and mean of simulated distribution is also very close to 5. Additionally the theortical mean is 0.625, which is in close agrement with variance of simulated data.

```
# Comparing the distribution of simulation data with central limit theory
# Calculating theoretical mean, standard error and vairance
theor_mean<- 1/lambda
theor_std_err<-theor_mean/sqrt(sample_size)
theor_var<-theor_std_err^2
# Calculating mean, standard error and variance of simulation data
sim_mean<-mean(mean_per_row)
sim_std_err<-sd(mean_per_row)
sim_var<-var(mean_per_row)
# creating a table for comparison of statistics
comp_table<-
cbind(c(theor_mean,theor_std_err,theor_var),c(sim_mean,sim_std_err,sim_var))
colnames(comp_table)<-c("Theory","Simulation")
```

```
rownames(comp_table)<-c("Mean","Std Error","Variance")
print(round(comp_table,3))
```

```
##           Theory Simulation
## Mean       5.000      4.990
## Std Error  0.791      0.786
## Variance   0.625      0.618
```

Distribution

In this section, with the help of a graph, it has been tried to show that distribution of sumated data is normal. First, a ggplot graph with mean of simulated data has been plot and a density curve of purple shown. On top of this curve, using stat_function, a normal distribution curve has been ploted in black. the close overlap of two curves shows that the distribution of simulated data is normal

```
den_graph2<-ggplot(data.frame(mean_per_row),aes(x=mean_per_row))
den_graph2<-den_graph2+ggtitle("How density distribution is normal")
den_graph2<-den_graph2+geom_histogram(binwidth =
0.15,colour="aquamarine",fill="cyan3",aes(y = ..density..))
den_graph2<-den_graph2+geom_density(colour="darkorchid",size=1.01)+labs(x="Mean of
40 columns",y="Density")
den_graph2<-den_graph2+stat_function(fun = dnorm,
                      args = list(mean = theor_mean, sd=theor_std_err),
                      colour="black",size=1.01,linetype=1)
den_graph2<-den_graph2+geom_vline(xintercept = theor_mean, size=1.05,
colour="black") +
  geom_vline(xintercept = sim_mean, size=1.05, colour="darkorchid")
print(den_graph2)
```