Statistical Inference Course Project Part A

Aayush Shah

Overview

This is the first part of the course project for Statistical Inference course.In this project investigation of the exponential distribution and its comparison with the Central Limit Theorem is performed.Analysis is performed with help of plots and confidence intervals and the normal behaviour of the distribution is proved.

Tasks

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.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

Show the sample mean and compare it to the theoretical mean of the distribution. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution. Show that the distribution is approximately normal. In point 3, focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

Simulation

Simulating the exponential distribution

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 also 1/lambda.

```
set.seed(100)
lambda <- 0.2
n <- 40
sample_size <- 1000
simulation <- replicate(sample_size, rexp(n, lambda))
means_expdistribution <- apply(simulation, 2, mean)
```

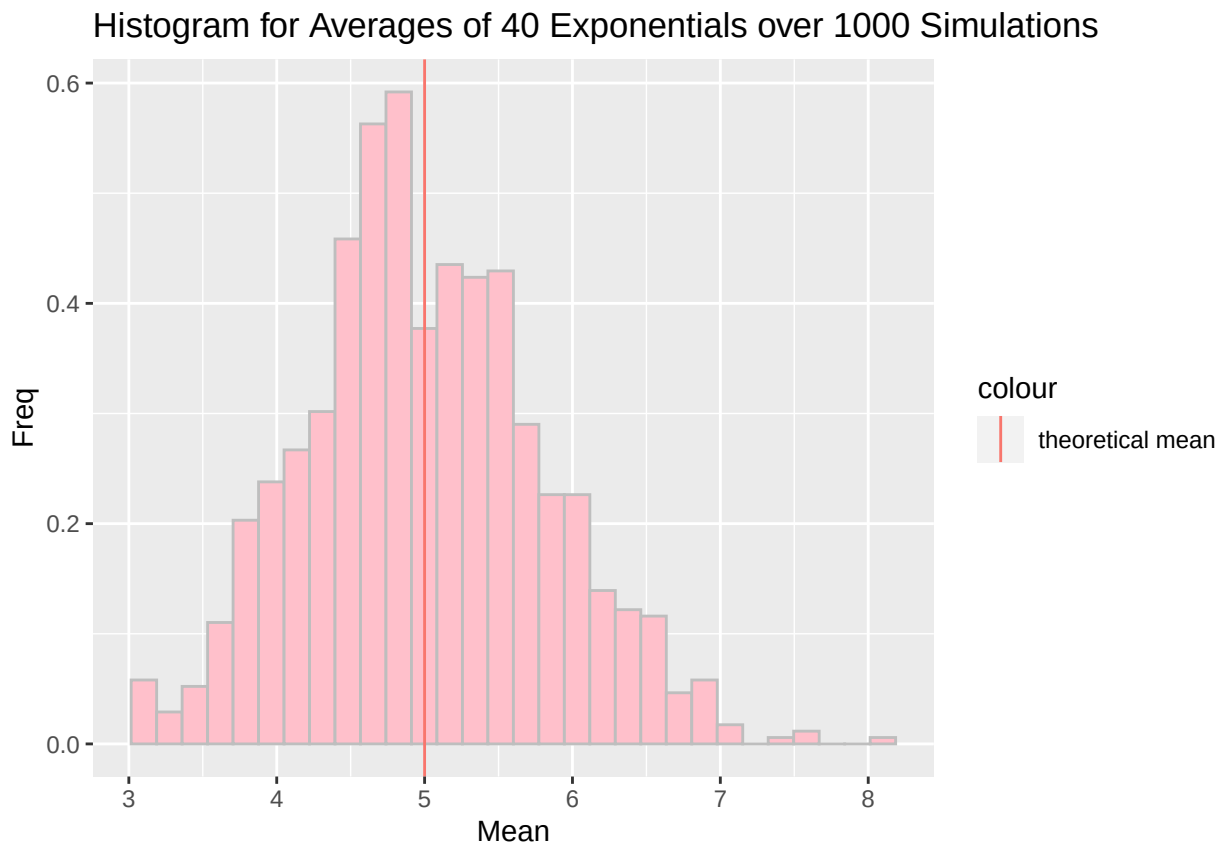
Comparing sample mean and theoretical mean

```
sample_mean<-mean(means_expdistribution)
theoretical_mean<-1/lambda
```

So the sample mean is 4.9997019 and the theoretical mean is 5.As seen, the two values are very close.

```
library(ggplot2)
datamean <- data.frame(means_expdistribution)
means_plot<-ggplot(datamean, aes(x =means_expdistribution))+geom_histogram(aes(y=..density..), colour="
means_plot
```

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



The pink line indicates the theoretical mean which is 5

Comparing sample variance+standard deviation and theoretical variance+standard deviation

```
sample_variance <- var(means_expdistribution)
theoretical_variance  <- (1 / lambda)^2 / (n)
sample_std <- sd(means_expdistribution)
theoretical_std  <- 1/(lambda * sqrt(n))
```

So the sample and theoretical **variances** are *0.6432442* and *0.625* respectively and sample and theoretical **standard deviations** are *0.8020251* and *0.7905694* respectively.So as seen, these values are also very close to each other.

Confidence Interval of 95%

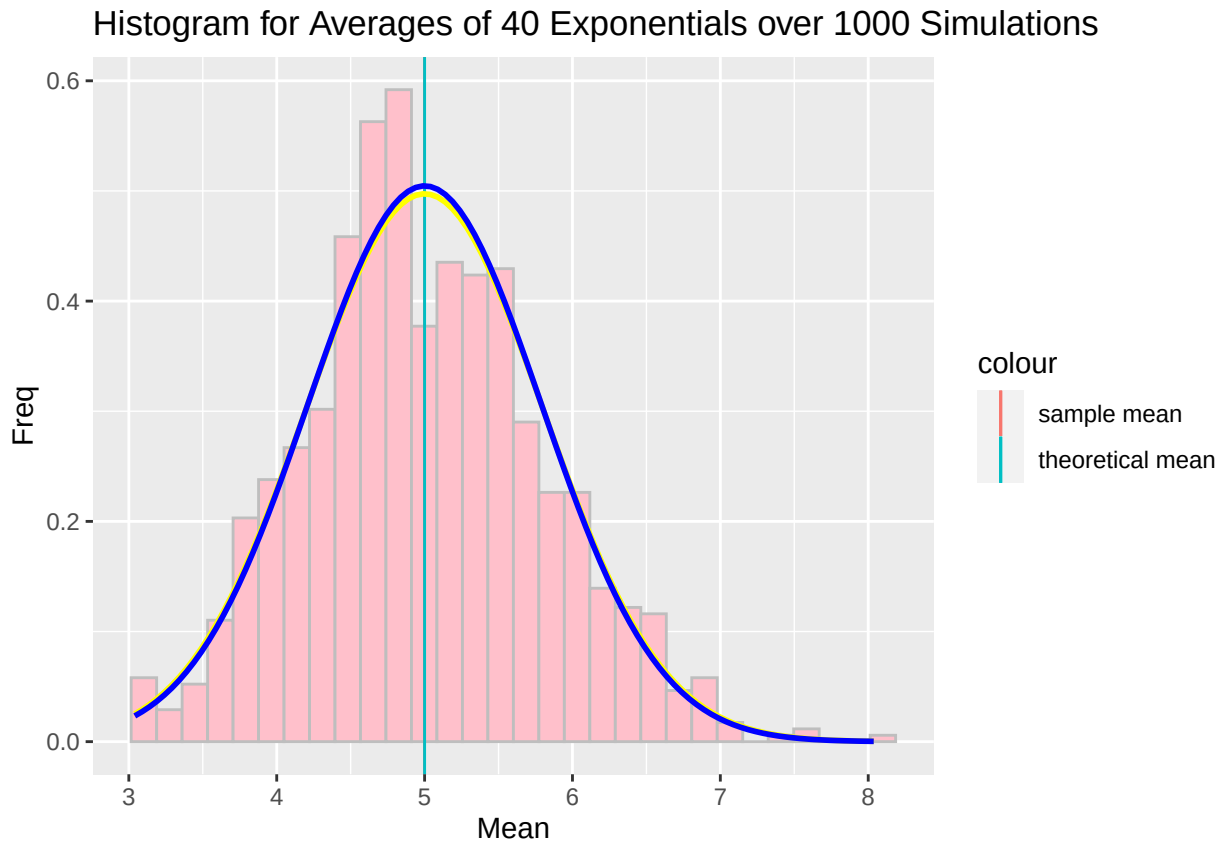
```
samplemean_cinterval <- round (mean(means_expdistribution) + c(-1,1)*1.96*sd(means_expdistribution)/sqr
```

The sample confidence interval is (4.751, 5.248).So we are 95% confident that the true mean lies in this interval.

Distribution

```
library(ggplot2)
gg<-means_plot+geom_vline(aes(xintercept = sample_mean, colour = "sample mean"))+ geom_vline(aes(xinter
gg
```

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



The theoretical mean and the sample mean are very close. The ***blue*** line shows the normal curve formed by the the theoretical mean and standard deviation. The ***yellow*** line shows the curve formed by the sample mean and standard deviation.

As seen from the graph, the distribution of means of 40 exponential distributions is close to the normal distribution.