Statistical Inference Course Project Part 1

Samir Yelne

Assignment Description

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.

Simulation

The following code perform the simulations to investigate the averages of 40 exponentials.

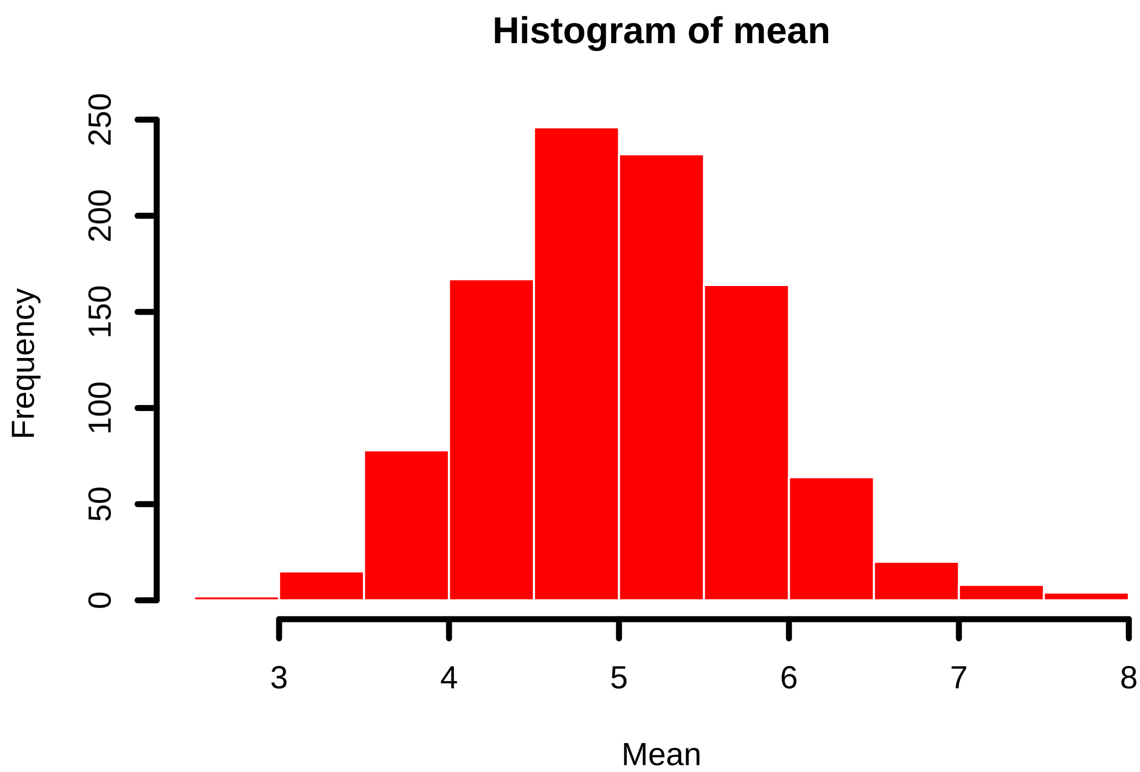
```r
library(ggplot2)

simulations <- 1000
lambda <- 0.2
n <- 40

set.seed(123)
mat <- matrix(rexp(simulations*n,rate=lambda),simulations,n)
mat_mean <- rowMeans(mat)
```

The histogram of simulation

```r
hist(mat_mean,col="red",main="Histogram of mean",xlab="Mean",border="white",type="o",lwd="3")
```



Sample Mean Comparison

The actual mean for the sample data and theoretical mean are calculated below:

```
actual_mean <- mean(mat_mean)
theory_mean <- 1/lambda
```

Actual center of the distributions based on the simulations is 5.011911, on the other hand the theoretical mean when lambda=0.2 is 5. This shows that the actual mean from the sample data is very close to the theoretical mean of normal data.

Variance Comparison

The actual variance for the sample data and thereotical variance are shown below

```
actual_v <- var(mat_mean)
theory_v <- (1/lambda)^2/n
```

Actual variance for the sample data is 0.6088292 while the theoretical variance is 0.625. Both these values are again quite close.

Approximately Normal Distribution

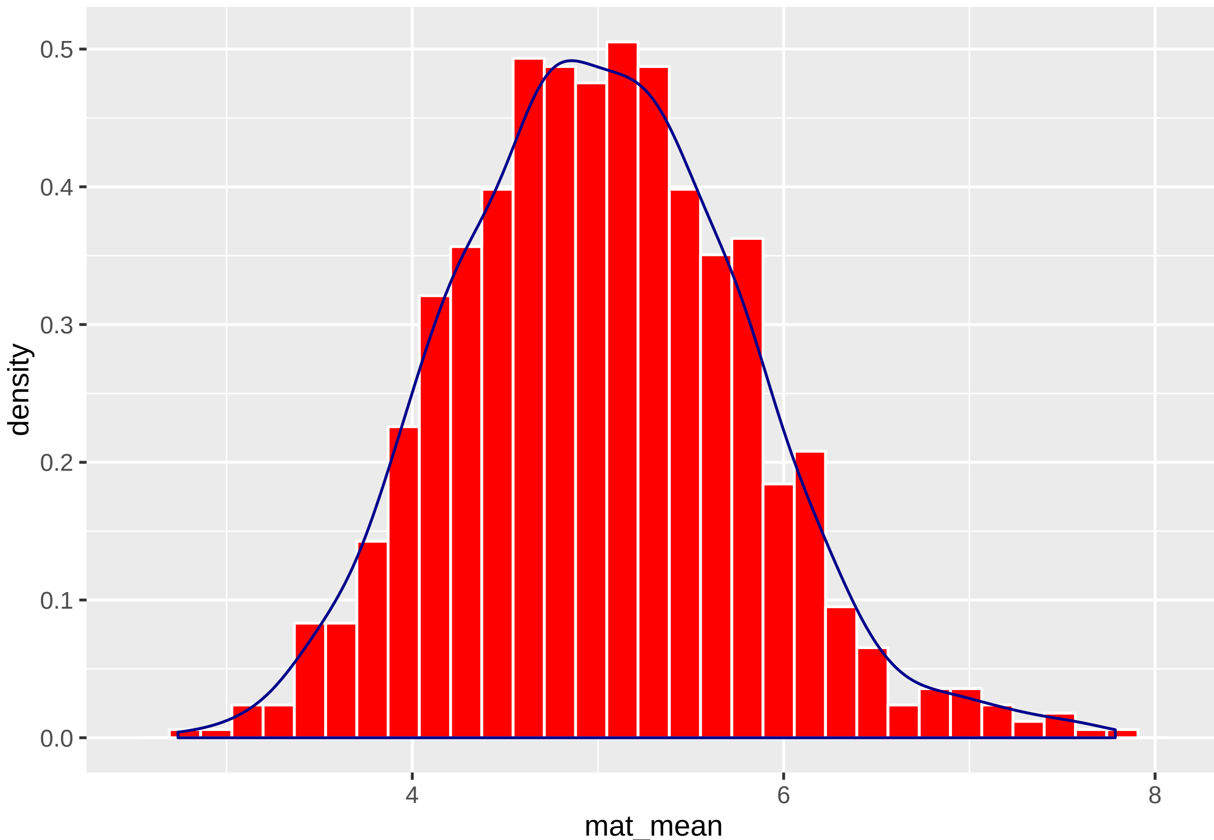
To show the distribution is approximately normal, we perform the three steps:

1. Create a normal distribution and see how the sample data aligns with it.
2. Compare the confidence interval along the mean and variance with normal distribution.
3. q-q plot for quantiles

Step 1

```
data <- data.frame(mat_mean)
ggplot(data,aes(x=mat_mean))+geom_histogram(aes(y=..density..),color="white",fill="red")+geom_density(co
```

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



The above figure shows that the distribution of the histogram approximately follows the normal distribution.

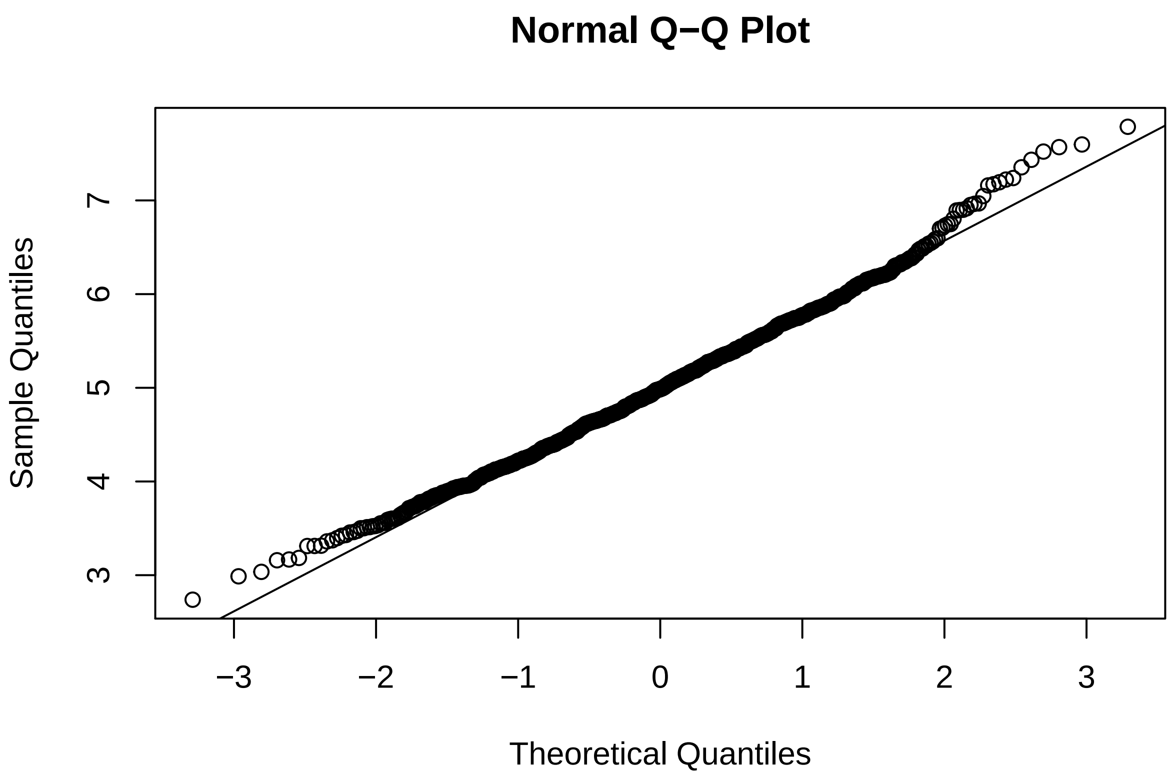
Step 2

Here we try to show that the confidence intervals match

Actual 95% confidence interval is [4.770, 5.254]. Theoretical 95% confidence interval is [4.755,5.245].

Step 3

```r
qqnorm(mat_mean)
qqline(mat_mean)
```



The thereotical quantiles also match closely with the actual quantiles. These three evidences show that the distribution is approximately normal.