Statistical Inference Project Part I

Ronda Hutchin

8/22/2020

Part I Assignment Instructions

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:

(1) Show the sample mean and compare it to the theoretical mean of the distribution.
(2) Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
(3) Show that the distribution is approximately normal.

Part I Overview

The code performs calculations, plots and compares confidence intervals to prove the distribution is a normal distribution.

Load library & Set Variables

Set the seed and required variable values per the assignment instructions written above. The ggplot2 library is loaded for generating histograms.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
set.seed(100)
lambda <- 0.2
NumberExp <- 40
NumberSimulations <- 1000
```

Step 1 for Part I (1)

Run the simulation 1000 times and get the sample mean and compare it to the theoretical mean of the distribution.

```
MySimulatedSample <- replicate(NumberSimulations, rexp(NumberExp, lambda))
MySimSampleMean <- apply(MySimulatedSample, 2, mean)
SampleMean <- mean(MySimSampleMean)
TheoreticalMean <- 1 / lambda
```

Comparison summary. The sample and theoretical mean are very close with a different of -0.0002980731 !

```
SampleMean
```

```
## [1] 4.999702
```

```
TheoreticalMean
```

```
## [1] 5
```

```
SampleMean - TheoreticalMean
```

```
## [1] -0.0002980731
```

Step 2 for Part I (2)

Show how variable the sample is (via variance and standard deviation) and compare it to the theoretical variance of the distribution.

```
SampleVariance <- var(MySimSampleMean)
TheoreticalVariance  <- (1 / lambda)^2 / (NumberExp)

SampleStdDev <- sd(MySimSampleMean)
TheoreticalStdDev  <- 1/(lambda * sqrt(NumberExp))
```

Comparison summary. The sample Variances and Standard Deviation are close.

```
SampleVariance
```

```
## [1] 0.6432442
```

```
TheoreticalVariance
```

```
## [1] 0.625
```

```
SampleStdDev
```

```
## [1] 0.8020251
```

```
TheoreticalStdDev
```

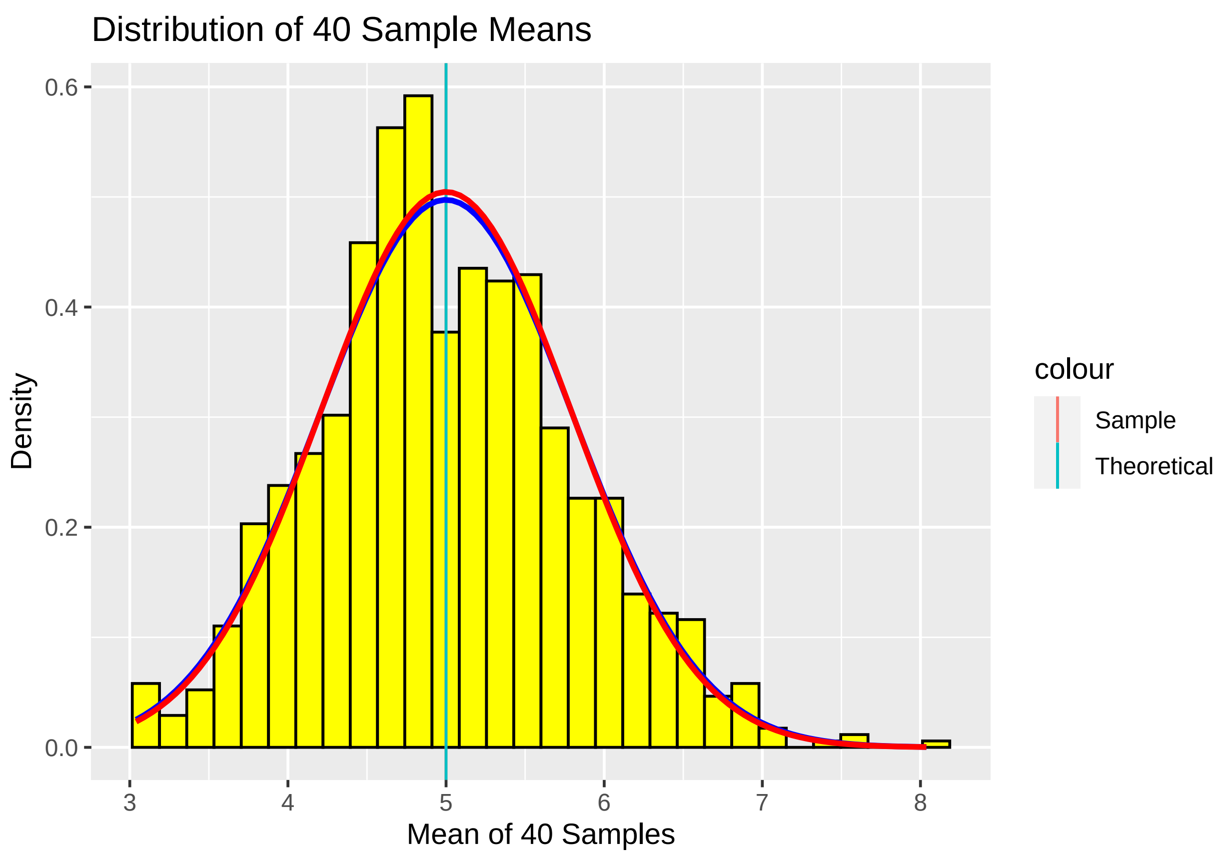
```
## [1] 0.7905694
```

Step 3 for Part I (3)

Show that the distribution is approximately normal.

```
plotdata <- data.frame(MySimSampleMean)
Myplot <- ggplot(plotdata, aes(x =MySimSampleMean))
Myplot <- Myplot + geom_histogram(aes(y=..density..), color="black",
                        fill = "yellow")
Myplot <- Myplot + labs(title = "Distribution of 40 Sample Means", x = "Mean of 40 Samples", y = "Densi
Myplot <- Myplot + geom_vline(aes(xintercept = SampleMean, color = "Sample"))
Myplot <- Myplot + geom_vline(aes(xintercept = TheoreticalMean, color = "Theoretical"))
Myplot <- Myplot + stat_function(fun = dnorm, args = list(mean = SampleMean, sd = SampleStdDev), color =
Myplot <- Myplot + stat_function(fun = dnorm, args = list(mean = TheoreticalMean, sd = TheoreticalStdDe
Myplot
```

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



Comparison summary. The sample Variances and Standard Deviation are close. The previous chart show how close the distribution is to normal.

```
SampleConfidence <- round (mean(MySimSampleMean) + c(-1,1)*1.96*sd(MySimSampleMean)/sqrt(NumberExp),3)
TheoreticalConfidence <- TheoreticalMean + c(-1,1) * 1.96 * sqrt(TheoreticalVariance)/sqrt(NumberExp)

SampleConfidence
```

```
## [1] 4.751 5.248
```

```
TheoreticalConfidence
```

```
## [1] 4.755 5.245
```