Statistical Inference Project - Part 1: Simulation Exercise

Lena Horsley

Overview/Instructions

taken from the assignement page:

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.

Setup

```
sampleSize <- 40
lambda <- 0.2
numOfSimulations <- 1000
quantile <- 1.96
set.seed(1984)
```

Data Creation and Simulation

```
simulationData <- matrix(rexp(sampleSize * numOfSimulations, rate = lambda),
numOfSimulations, sampleSize)
simulationMean <- rowMeans(simulationData)

#Let's take a peak
head(simulationData)
```

```
##           [,1]      [,2]       [,3]     [,4]     [,5]      [,6]      [,7]
## [1,] 1.5880473 1.7917730  0.5417411 15.53612 3.066071 2.7587792  8.636012
## [2,] 4.6127430 0.2857810  3.3640577  1.88829 1.830337 9.1432803  6.683237
## [3,] 2.3756366 2.8361541  6.1349024 11.41090 2.820193 2.8685734  7.061109
## [4,] 0.1124172 0.7847665  8.9503474 26.73106 3.503630 0.9380747 25.489629
## [5,] 3.2986892 6.0033770 17.6998946 11.30092 4.080523 2.5065714  3.681563
## [6,] 6.9729743 2.1570919  1.2499536 14.34263 2.122486 7.7708459 15.137679
##           [,8]       [,9]      [,10]     [,11]     [,12]     [,13]
## [1,] 1.5885185 19.6753355 0.01187565  5.099723 8.2411748  7.216157
## [2,] 1.0524345  0.9495248 3.55186941  7.554435 0.3220688  4.932518
## [3,] 0.9090108  1.8847490 0.31233845  6.264267 1.0896915 11.032281
## [4,] 9.1509624  5.1427812 3.33019528  8.089875 0.4101066  7.403714
## [5,] 6.6076286 10.9444500 2.19095102 17.845131 1.4294936  7.139901
## [6,] 1.8320902 12.0094665 0.39121610  2.550085 2.8732409  1.482500
##           [,14]     [,15]       [,16]      [,17]     [,18]      [,19]
## [1,]  0.58841278 14.239824  2.29325332  2.0360853  2.825817  1.4804273
## [2,] 15.17400073  2.054941  1.00691090  1.5914042 10.599647  9.1825299
## [3,]  3.54664620  6.658168 24.61131298  0.3019942  5.049182  2.7956966
```

```
## [4,]  0.04981892  2.246627  7.17657439  5.0962957  3.285412 10.6602086
## [5,]  0.72601423  9.493416  4.50298280 17.3637268  3.629832  1.4872502
## [6,]  1.09541122  4.061235  0.09394217  0.4031143 14.333723  0.9512317
##          [,20]     [,21]     [,22]      [,23]     [,24]      [,25]
## [1,] 0.2554073  6.549394  7.175034  8.0230898  5.979162  5.4223427
## [2,] 3.9658898  5.826490  9.732782  5.5577387  3.900711 11.7091879
## [3,] 0.4457545  5.651081  2.303850 15.1851889  1.119194  2.9638817
## [4,] 0.5468199 11.659088 11.633129  4.1953273 10.710007  0.7336178
## [5,] 0.4046304  3.084913 12.593915  0.3458649  1.984486  6.5582682
## [6,] 5.0487912 10.753569  1.670899  1.3901485  4.470450 13.4461429
##          [,26]      [,27]     [,28]     [,29]     [,30]     [,31]
## [1,]  1.998965  7.0605284 10.289907 2.3257022  3.522795  1.366102
## [2,]  9.780611  1.3699972  7.107224 1.7620510  7.551905  4.598623
## [3,] 12.681525  0.2582432  5.841863 5.5127871 16.555917  2.006972
## [4,]  8.910978  0.2411209  2.448739 0.6980031  1.314989  4.550418
## [5,]  5.121373 11.4371064 14.574273 3.9365977  8.956330 12.219839
## [6,] 12.326746  1.6964790  3.826455 3.1149687  1.409037  2.625311
##           [,32]      [,33]      [,34]     [,35]      [,36]     [,37]
## [1,]  4.95324410 2.83789936  3.1549116 5.0584885 14.9771146 1.4996507
## [2,]  5.02149843 2.64227128  0.6504391 7.1056764  6.7627509 0.8805932
## [3,]  1.88080351 0.42140028  1.3456615 1.8080194  2.5736761 1.0713728
## [4,]  0.04956456 6.74614675  0.1504861 0.9236111  9.4551476 2.8696486
## [5,]  2.59011162 4.71731407  1.7842772 1.4416602  0.5946620 8.2888814
## [6,] 12.28897731 0.09713004 10.8075563 0.1223492  0.7556099 0.3461867
##           [,38]    [,39]     [,40]
## [1,] 14.2969807 3.400018 1.6759730
## [2,] 10.8171338 3.405607 1.8158727
## [3,]  0.2785366 5.779179 3.0765646
## [4,] 12.7358314 2.744657 0.5944983
## [5,]  2.0480891 5.064706 3.5097153
## [6,]  1.2883089 1.650712 1.8711100
```

Analysis

Show the sample mean and compare it to the theoretical mean of the distribution.

```r
sampleMean <- mean(simulationMean)
theoreticalMean <- 1/lambda
meanDifference <- theoreticalMean - sampleMean

#Display the sampleMean, theoreticalMean, and meanDifference
library(knitr)
dfMean <- data.frame(Sample = c(sampleMean),
          Theoretical = c(theoreticalMean),
          Difference=c(meanDifference))
row.names(dfMean) <- c("Mean")
kable(dfMean)
```

	Sample	Theoretical	Difference
Mean	4.981324	5	0.018676

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

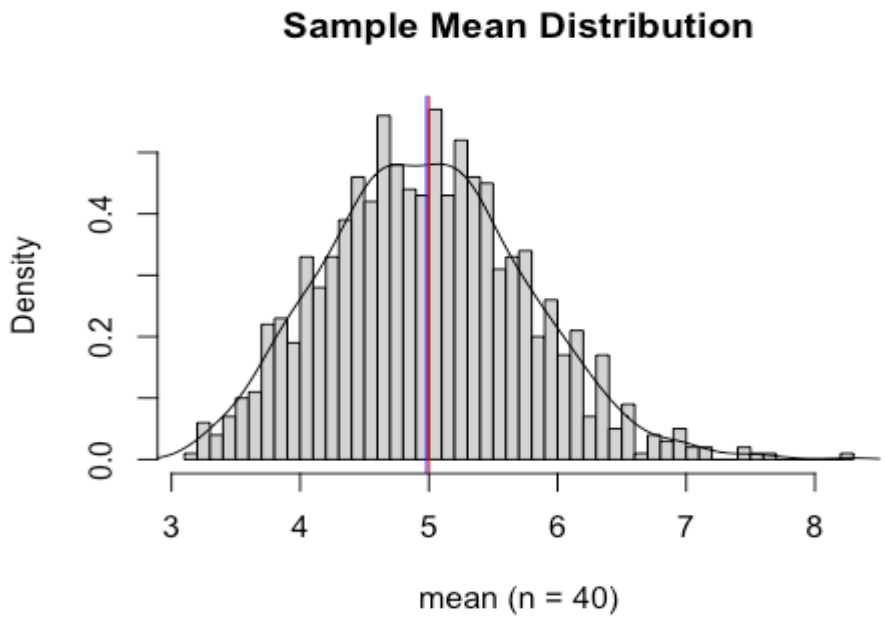
```
sampleVariance <- var(simulationMean)
theoreticalVariance <- ((1/lambda)^2)/sampleSize
varianceDifference <- theoreticalVariance - sampleVariance

#Display the sampleVariance, theorecticaVariance, and varianceDifference
dfVariance <- data.frame(Sample = c(sampleVariance),
          Theoretical = c(theoreticalVariance),
          Difference=c(varianceDifference))
row.names(dfVariance) <- c("Variance")
kable(dfVariance)
```

	Sample	Theoretical	Difference
Variance	0.612919	0.625	0.012081

Show that the distribution is approximately normal.

```
hist(simulationMean, xlab="mean (n = 40)", ylab="Density", main="Sample Mean
Distribution", breaks=sampleSize, probability = TRUE, col="light gray")
lines(density(simulationMean), col="black")
abline(v=mean(sampleMean), col="blue")
abline(v=mean(theoreticalMean), col="red")
```




```
#The sample mean is the blue line. The theoretical mean in the red line.
```

Now, let's look at the confidence intervals.

Theoretical Confidence Interval

```
#Step 1 - calculate the theorectical standard deviation
theoreticalStandardDeviation <- (1/lambda)/(sqrt(sampleSize))
theoreticalStandardDeviation

## [1] 0.7905694

#Step 2 - calculate the theorectical Confidence Interval
theoreticalConfidenceInterval <- theoreticalMean + c(-1,1) * quantile *
(sqrt(theoreticalStandardDeviation)/sqrt(sampleSize))
theoreticalConfidenceInterval

## [1] 4.724453 5.275547
```

Sample Confidence Interval

```
#Step 1- calculate the sample standard deviation
sampleStandardDeviation <- sd(simulationMean)
sampleStandardDeviation

## [1] 0.7828914

#Step 2 - calculate the sample Confidence Interval
sampleConfidenceInterval <- sampleMean + c(-1,1) * quantile *
(sampleStandardDeviation/(sqrt(sampleSize)))
sampleConfidenceInterval

## [1] 4.738703 5.223945
```

Comparison

```
dfConfidenceInterval <- data.frame(Sample = c(sampleConfidenceInterval),
                                   Theoretical = c(theoreticalConfidenceInterval))
row.names(dfConfidenceInterval) <- c("low","high")
kable(dfConfidenceInterval)
```

	Sample	Theoretical
low	4.738703	4.724453
high	5.223945	5.275547

Analysis

Despite the relatively close theoretical and sample values (mean, standard deviation, variance, confidence intervals), the graph shows the mean of 40 exponentials does not fit a "perfect" normal distribution. However, an increase in the sample size would resolve this.