Stat Inference Project

"Stone Cold" Tomaz Rodrigues

2025-02-21

Statistical Inference Course Project

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

Set up

```
## expected mean of one exponential distribution is 1/0.2 = 5
## expected std deviation of one exponential distribution is 1/0.2 = 5
lambda <- 0.2
n <- 40
numSims <- 1000
```

with basic parameters set we can now simulate the data we will be using

```
myVector <- 1:numSims
for(i in 1:numSims){
  myDistribution <- rexp(n, lambda) ##simulate distribution as described
  myMean <- mean(myDistribution) ##take mean of the exp distribution
  myVector[i] <- myMean ##stick the mean into myVector for analysis
}
```

Comparing sample mean to expected mean

```
##expected mean of the means of many exponential distributions with mean 5 is still 5
myVectorMean <- mean(myVector)
print(paste("my mean is ", myVectorMean, sep=""))
```

```
## [1] "my mean is 4.98176647039081"
```

Comparing sample variance to theoretical variance

```r
## expected standard deviation should be the standard error of the means
## Expected Standard Error 5/sqrt(40) ~= 0.79
myVectorDev <- sd(myVector)
print(paste("my standard error is ", myVectorDev, sep=""))
```

```
## [1] "my standard error is 0.787957326776485"
```

Is the data normally distributed at as we would expect?

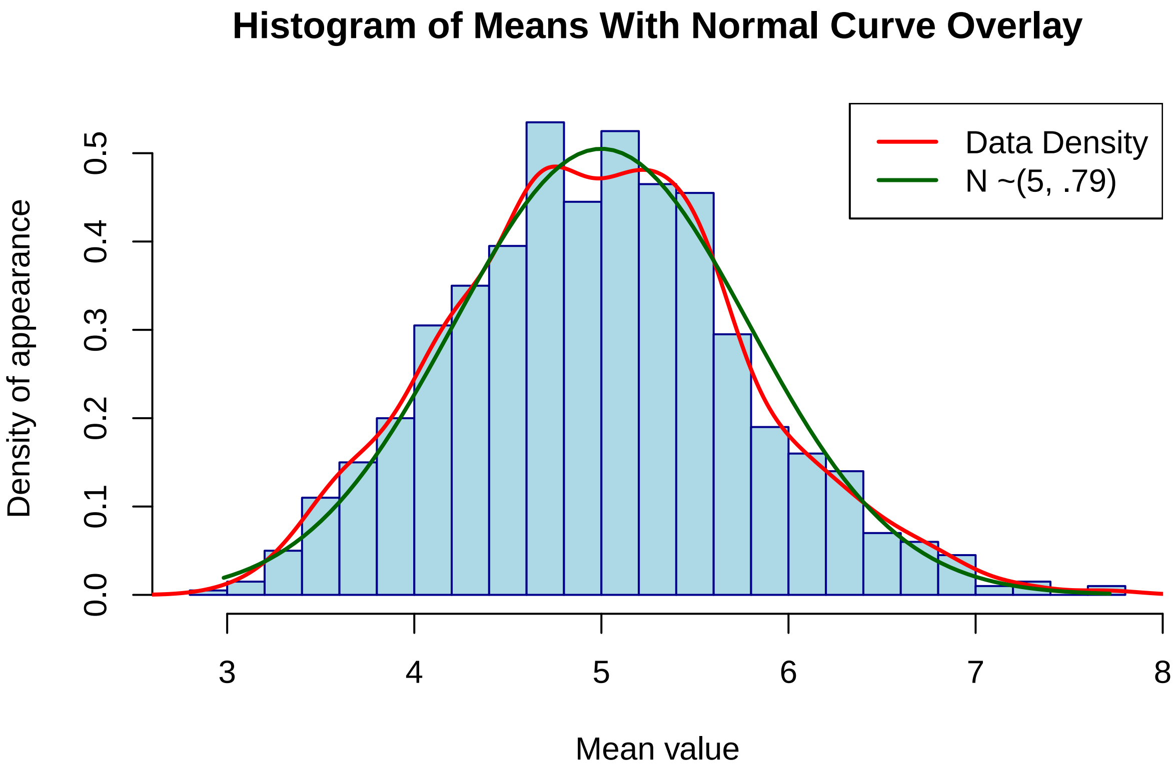
Using the central limit theorem we would expect the data to be normally distributed about mean 5 with standard error 0.79. Let's plot the data alongside a normal density curve with those parameters to see how it stacks up.

```r
hist(myVector,
     freq = FALSE,
     main = "Histogram of Means With Normal Curve Overlay",
     xlab = "Mean value",
     ylab = "Density of appearance",
     col = "lightblue",
     border = "darkblue",
     breaks = 20)

lines(density(myVector), col = "red", lwd = 2) # Adding density curve

##Add a curve with expected values and density to see if it matches up
curve(dnorm(x, mean = 5, sd = 0.79),
      from = min(myVector),  # Start from the min of your data
      to = max(myVector),    # End at the max of your data
      col = "darkgreen",
      lwd = 2,
      add = TRUE)

legend("topright",
       legend = c("Data Density", "N ~(5, .79)"),
       col = c("red", "darkgreen"),
       lwd = 2)
```



As you can plainly see from observing the density line and the normal curve line, the data closely approximates a normal distribution with mean 5 and SE 0.79. This suggests (I wouldn't go so far as to say proves) that the central limit theorem is legit. It would seem the last 2-300 hundred years of statistical theory wasn't built on a lie.