Exponential Distribution and a Basic Inferential Data Analysis

Abhinav Agarwal
25 November 2018

Overview

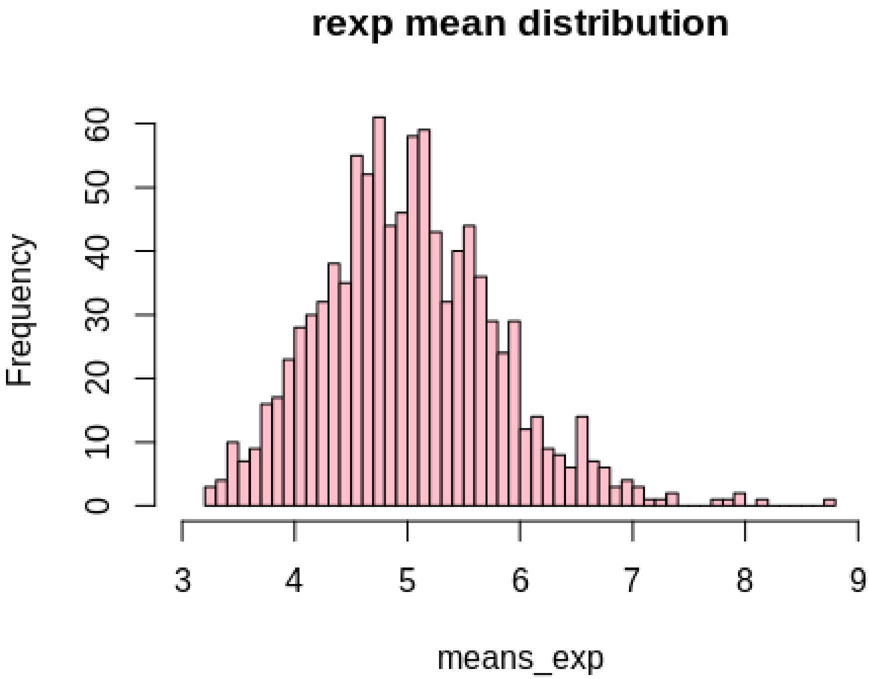
In the first part of this two-part project, we will compare the mean of simulated Exponential distributions in R and the expected theoretical values provided by the Central Limit Theorem (CLT), using the distribution of averages of 40 exponentials across 1,000 simulations. This assignment will make calculations and plots, and eventually prove that the distribution is approximately normal.

Simulations

The following parameters are used to create a sample data set that we can compare to expected values provided by the Central Limit Theorem.

Simulation: Distribution of 40 exponentials with a rate (lambda) of 0.2, simulated 1,000 times using the 'rexp' formula.

```r
# set seed for reproducability
set.seed(2018)

# Set sampling values as described in the project instructions
lambda <- 0.2   # lambda

n <- 40         # no of exponentials

sims <- 1000    # no of simulations

#Run simulations
sim_exp <- replicate(sims, rexp(n, lambda))

#Calc the means of the exponential simulations
means_exp <- apply(sim_exp, 2, mean)

#Histogram of the means

hist(means_exp, breaks=40, xlim = c(3,9), main="rexp mean
distribution", col = "pink")
```



Sample Mean versus Theoretical Mean

The mean of the exponential distribution is in theory 1/lambda. Since lambda is 0.2, the theoretical mean should render 5

```
sample_mean <- mean(means_exp)
sample_mean

## [1] 5.020107
```

The code above shows us that our sample mean is 5.02 which is pretty close to our theoretical mean of 5.

Sample Variance vs Theoretical Variance

The standard deviation of the exponential distribution is (1/lambda)/sqrt(n).

```
sample_var <- var(means_exp)
print("Sampl var:")

## [1] "Sampl var:"

sample_var
```

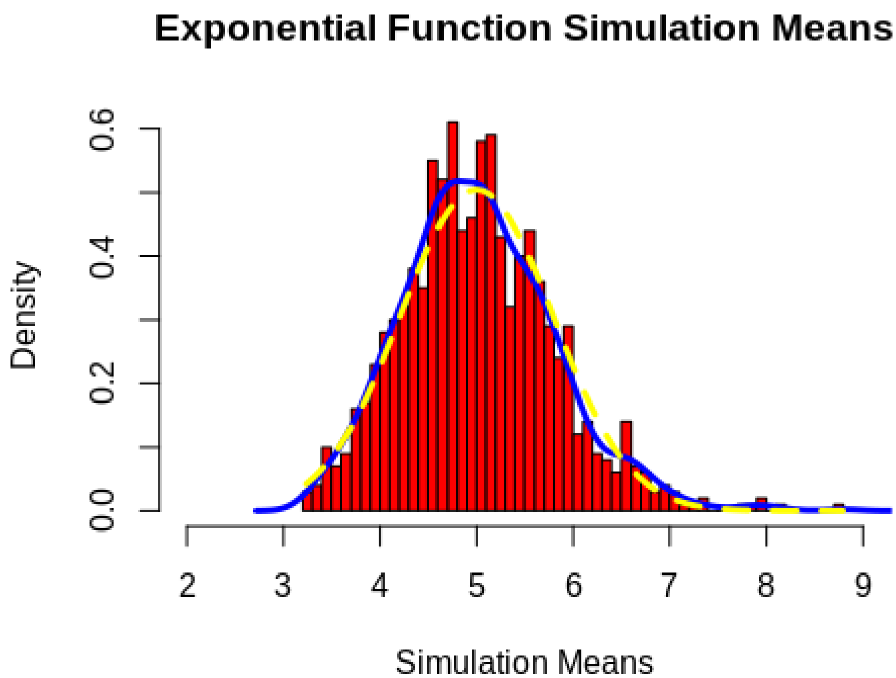
```
## [1] 0.6261326
theo_var  <- (1 / lambda)^2 / (n)
print("Theo var:")
## [1] "Theo var:"
theo_var
## [1] 0.625
sample_sd <- sd(means_exp)
print("Sampl sd:")
## [1] "Sampl sd:"
sample_sd
## [1] 0.7912854
theo_sd  <- 1/(lambda * sqrt(n))
print("Theo sd:")
## [1] "Theo sd:"
theo_sd
## [1] 0.7905694
```

As shown in the output above, the simulated variance of the means and the theoretical variance are close, further validating CLT.

Sample Distribution - Approximately Normal?

Finally, we'll investigate whether the exponential distribution is approximately normal. According to the Central Limit Theorem, the means of the sample simulations should follow a normal distribution.

```
#General Plot with ditribution curve drawn
hist(means_exp, prob=TRUE, col="red", main="Exponential Function
Simulation Means", breaks=40, xlim=c(2,9), xlab = "Simulation
Means")
lines(density(means_exp), lwd=3, col="blue")

# Normal distribution line creation
x <- seq(min(means_exp), max(means_exp), length=2*n)
y <- dnorm(x, mean=1/lambda, sd=sqrt(((1/lambda)/sqrt(n))^2))
lines(x, y, pch=22, col="yellow", lwd=3, lty = 2)
```



As shown in the graph, the calculated distribution of means of random sampled exponantial distributions overlaps with the normal distribution, due to the Central Limit Theorem. The more samples we would get (now 1000), the closer will the density distribution be to the normal distribution bell curve.

References

The following references were used in this report:

- The Central Limit Theorem Applied to the Exponential Distribution in R by Calin Uioreanu
- Coursera: Statistical Inference Course Project 2 - Part 1 by Byron Raco
- Statistical Inference Course Project Part 1 by Zhikang Xu
- Statistical Inference - Course Project Part 1 by Clint Lehman