Statistical Inference: Exponential Distribution

Sanjay Somraj

March 22, 2017

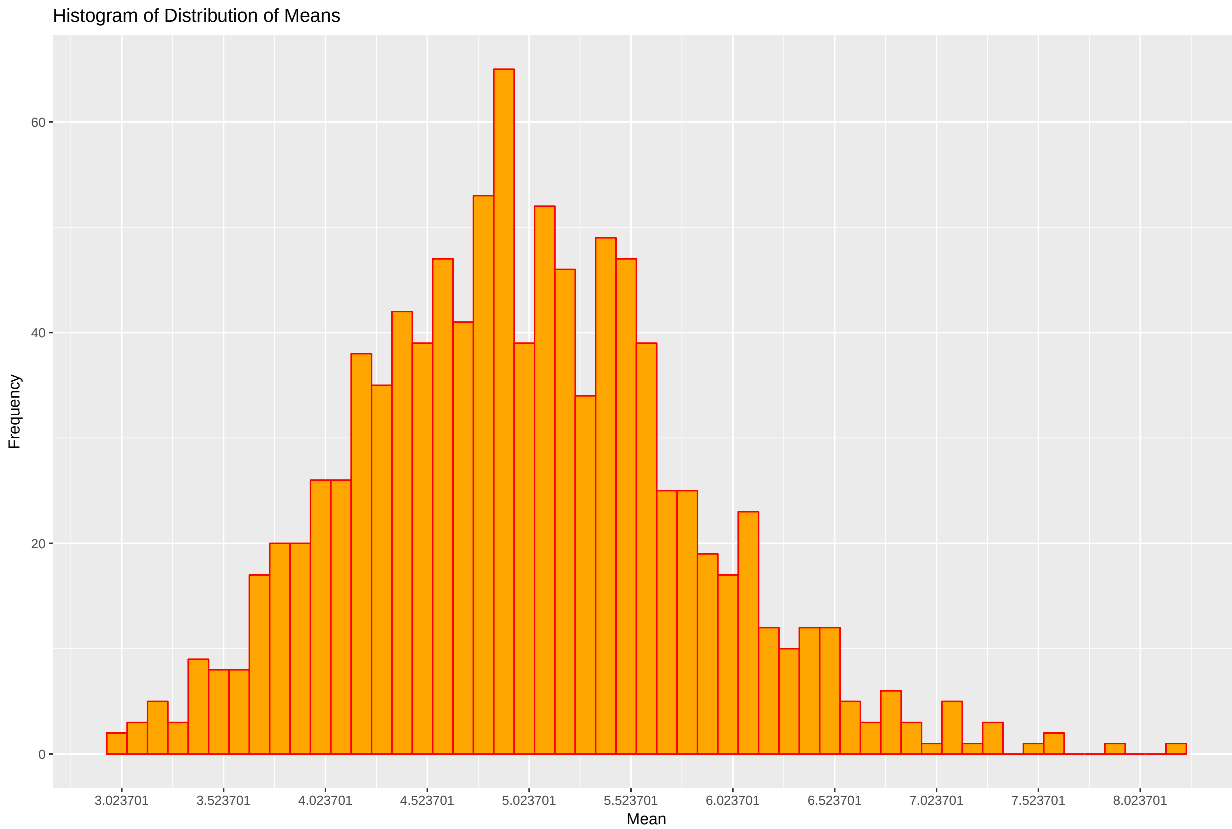
Statistical Inference Course Project - PART 1

Overview

1 This project investigates the exponential distribution in R and compare it with the Central Limit Theorem.
2 This project will investigate the distribution of averages of 40 exponentials over 1000 test simulations.
3 The Rate parameter, lambda = 0.2 for all of the simulations.
4 The exponential distribution can be simulated in R with rexp(n, lambda).
5 The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda.

Load libraries for data processing and plotting

```
library(ggplot2)
# INITIALIZE CONSTANTS
     # LAMBDA for rexp
          lambda <- 0.2
     # Number of EXPONENTIALS
          n <- 40
     # Number of TEST SIMULATIONS
          numberOfSimulations <- 1000
     # set the SEED for reproducability
          set.seed(1000)

# EXECUTE TEST PRODUCING 40 X 1000 MATRIX
exponentialDistributions <-
     matrix(data=rexp(n * numberOfSimulations, lambda),
            nrow=numberOfSimulations)

exponentialDistributionMeans <-
     data.frame(means=apply(exponentialDistributions, 1, mean))

minExpValue <- min(exponentialDistributionMeans$means)
maxExpValue <- max(exponentialDistributionMeans$means)

# PLOT THE DISTRIBUTION MEANS
ggplot(data = exponentialDistributionMeans,
       aes(x = means)) +
     geom_histogram(col="red", fill = "orange", binwidth = 0.1)+
     labs(title="Histogram of Distribution of Means",
          x = "Mean", y = "Frequency") +
     scale_x_continuous(breaks=seq(minExpValue,maxExpValue,by=0.5))
```



Sample Mean versus Theoretical Mean

First, let us compute the mean of sample means of the 1000 simulations of 40 randomly sampled exponential distributions. We will compute Theoretical Mean is the 1/lambda i.e.

```
sampleMean <- mean(exponentialDistributionMeans$means)
theoryMean <- 1/lambda
```

Theoretical Mean: 5
Sample Mean: 4.9869634

We can see that SAMPLE MEAN (4.987) AND THEORETICAL MEAN (5) are very close.

Sample Variance versus Theoretical Variance

Theoretical **STANDARD DEVIATION** is computed as mu/square-root(n) i.e 1/lambda/sqrt(n)
And the **VARIANCE** is Square(Standard Deviation)

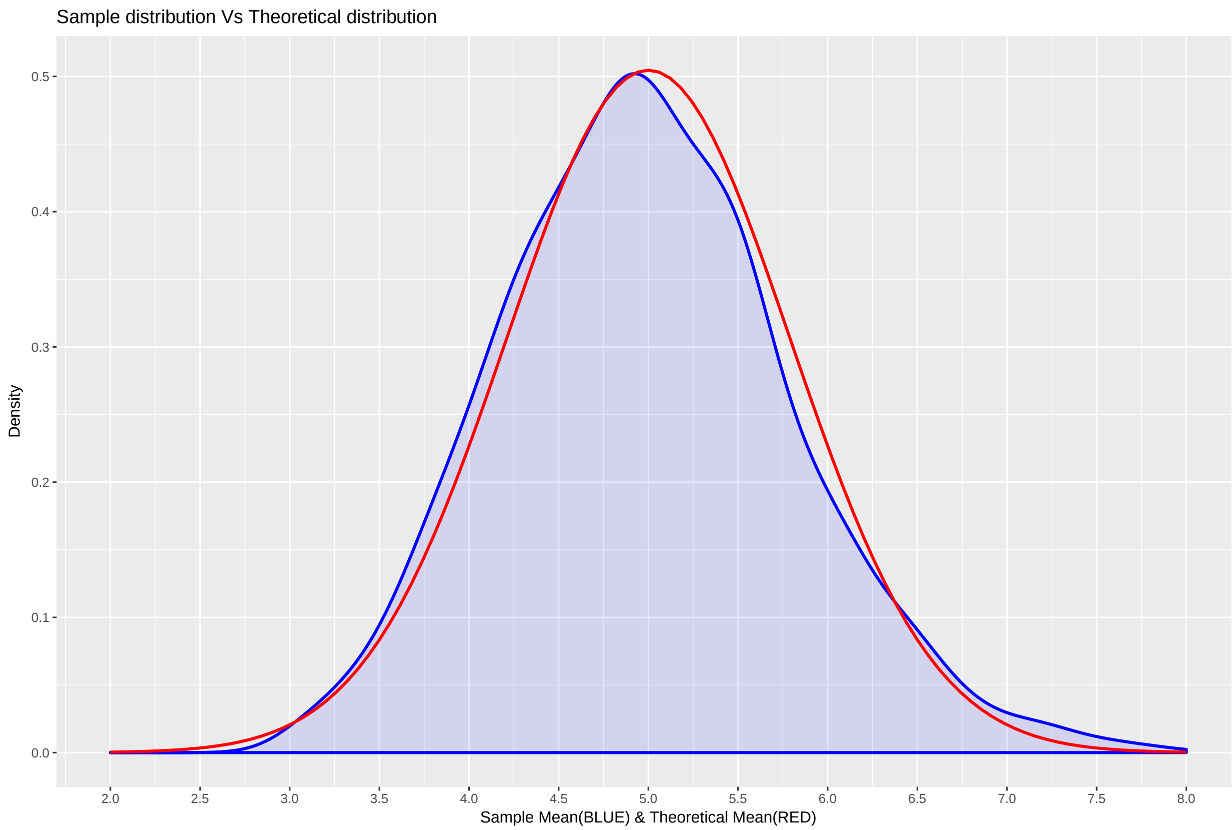
```
stdDev <- theoryMean/sqrt(40)
varnc <- stdDev^2

sampleStdDev <- sd(exponentialDistributionMeans$means)
sampleVarnc <- var(exponentialDistributionMeans$means)
```

- **Theoretical Variance:** 0.625 **Sample Variance:** 0.6583551
  We observe that the Sample Variance is very close to Theoretical Variance.

Distribution

The plot shows the normal distribution and sample data distribution.

```
ggplot(data = exponentialDistributionMeans, aes(x = means)) +
     geom_density(colour="blue", fill="blue", alpha=0.1, size=1) +
     stat_function(fun = dnorm, args = list(mean = theoryMean, sd = stdDev),
                   colour = "red", size=1) +
     scale_x_continuous(breaks=seq(theoryMean-3, theoryMean+3,0.5),
                        limits=c(theoryMean-3,theoryMean+3))+
     xlab('Sample Mean(BLUE) & Theoretical Mean(RED)') + ylab('Density') +
     ggtitle('Sample distribution Vs Theoretical distribution')
```



Summary

For the given values of LAMBDA, N and the 1000 simulations, we observe that the computed Sample Mean, Variance and Standard Deviation are very close to respective Theoretical values.

The distribution of the sample and theoretical values are overlayed on each other to help us with a clear view on the extent of overlap. The RED line represents the Normal distribution while the BLUE line denotes the sample data distribution. The calculated distribution of means of random sampled exponential data overlaps with the normal distribution.