Coursera Statistical Inference Project - Part 1

Brian Altman

December 25, 2015

Overview

This project explores R's exponential distribution and compares it with the Central Limit Theorem. The exponential distribution is simulated in R with rexp(n, lambda). The project's objectives:

1. Create a simulation using rexp(n, lambda)
2. Compare the simulation mean to the theoretical mean of the distribution.
3. Compare the simulation variance to the theoretical variance of the distribution
4. Show the sample distribution is approximately normal.

This file and related code can be found here: https://github.com/beaisis/Statistical-Inference

Simulation R's rexp(n, lambda), where lamda is the rate parameter, was used to create a matix of 1000x40 random exponential numbers. With simulation parameters given by the project requirements, the means, variance and confidence intervals are determined for both the simulation and the theoretical models.

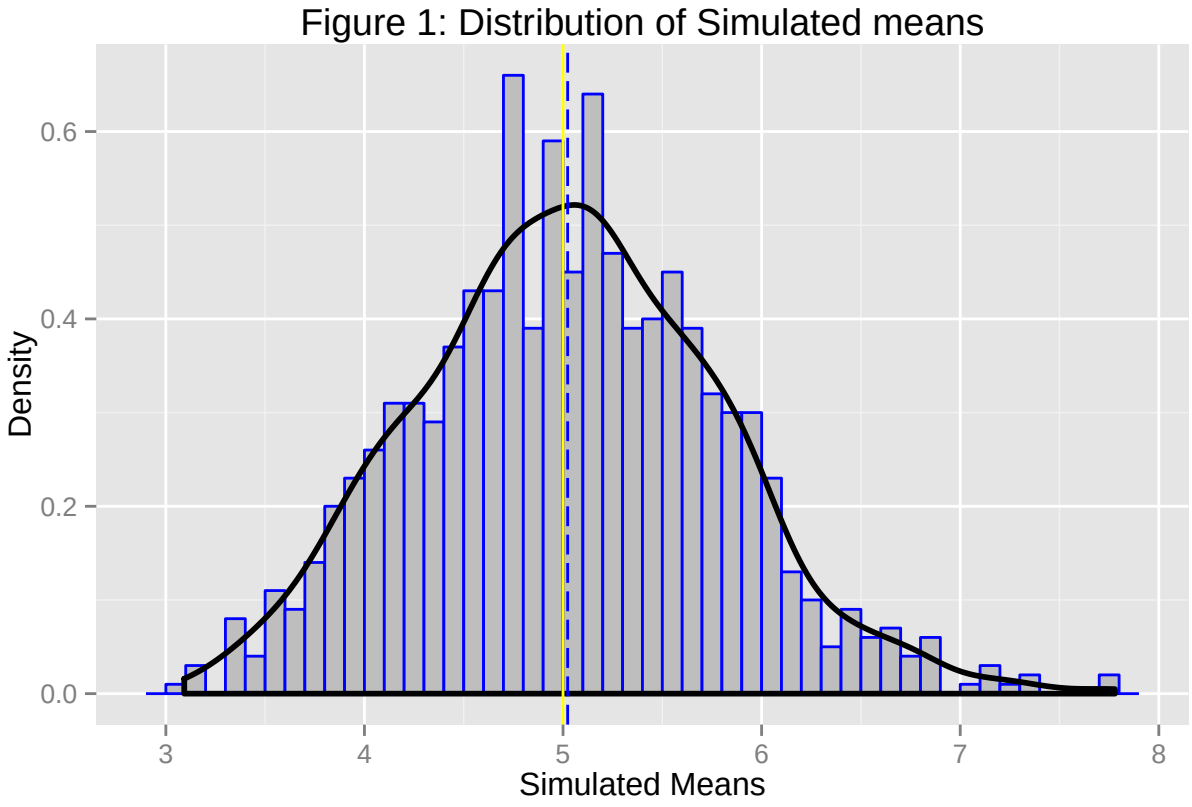
```
library(ggplot2)
set.seed(314)
simulations   <- 1000  #Given by the project requirements.
lambda        <- 0.2   #Given by the project requirements.
n             <- 40    #Given by the project requirements.

simulated_means <- apply(matrix(rexp(simulations*n, lambda), simulations, n), 1, mean)

simulated_mean <- mean(simulated_means)
simulated_var  <- var(simulated_means)
simulated_std  <- sd(simulated_means)
simulated_ci   <- simulated_mean + c(-1,1)*1.96*sd(simulated_means)/sqrt(n)

theoretical_mean <- 1/lambda   #Given by the project requirements.
theoretical_var  <- 1/(lambda^2*n)
theoretical_std  <- (1/lambda)/sqrt(n)
theoretical_ci   <- theoretical_mean + c(-1,1)*1.96*theoretical_std/n
```

Compare the sample mean to the theoretical mean of the distribution. The following figure shows the density for the simulation, the mean of the simulation (dashed blue line) and the theoretical mean (yellow).



Refering to above Figure 1, the simulation and theoretical mean are visually the same. The values indicate less than 1% difference:

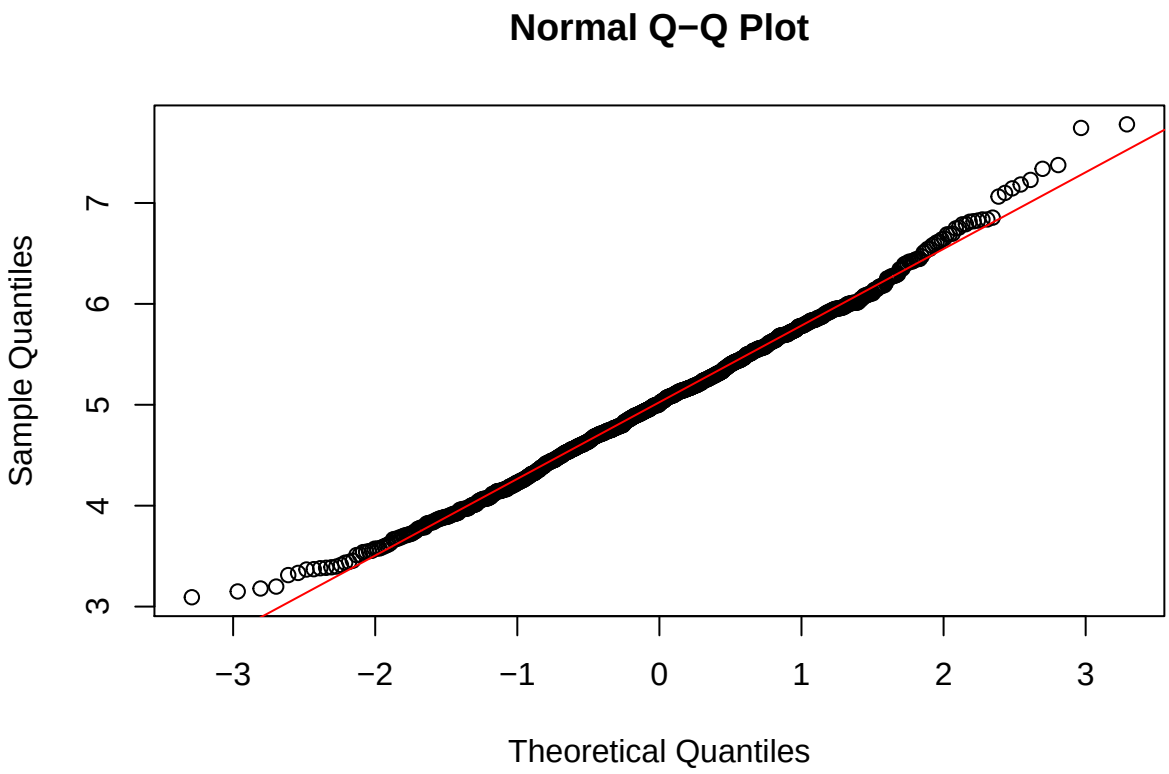
```
Simulated mean  : 5.0233404
Theoretical mean: 5
```

Compare the simulated and theoretical variance. The simulated and theoretical standard deviation and variance are as follows:

```
Standard Deviation
      Simulated std  : 0.760382
      Theoretical std: 0.7905694
Variance
      Simulated var  : 0.5781807
      Theoretical var: 0.625
```

Confirm the sample distribution is approximately normal In addition to Figure 1, the following shows the distribution compared to the normal distribution and evaluates the coverage of the confidence interval for 1/lambda: $\bar{X} \pm 1.96S/n$.

```
qqnorm(simulated_means)
qqline(simulated_means, col = 2)
```



The confidence intervals for the simulation and the theoretical are:

```
Simulated CI  : 4.7876956, 5.2589853
Theoretical CI: 4.9612621, 5.0387379
```

Conclusion The distribution is very close to a normal distribution, supporting the Central Limit Theorem