Statistical Inference Project - Part 1 - Simulation Exercise

Tim Westran

Dec 27, 2018

Part 1: Simulation Exercise Instructions

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

Load Libraries

```
## Let's load some libraries we need.
library("data.table")
library("ggplot2")
```

This project aims to investigate the exponential distribution in R, which can be simulated (in R) with `rexp(n, lambda)`, where `lambda` λ is the rate parameter. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. We will use $\lambda = 0.2$ for all simulations, investigating the distribution of averages of 40 samples from the exponential distribution, with $\lambda = 0.2$.

We will perform 1000 simuations, taking the averages of 40 samples from the exponential distribution

Question 1 - Sample mean vs theoretical mean

First, we will compare the sample mean with the theoretical mean.

```
## First, set the seed to ensure repeatability.  I picked 444 just because.
set.seed(444)

## Set lambda equal to .2, as specified within the problem.
lambda <- 0.2

## Set the number of simulations run equal to 1000, as specified within the problem.
simulations <- 1000

## Set the sample size equal to 40, as specified within the problem.
sampleSize <- 40

## Let's perform the simulations.
simulatedRuns <- replicate(simulations, rexp(sampleSize, lambda))

## Now, let's calculate the mean for each set of 40 samples, one mean for each simulation.
## simulatedMean will contain 1000 means.
simulatedMean <- apply(simulatedRuns, 2, mean)
```

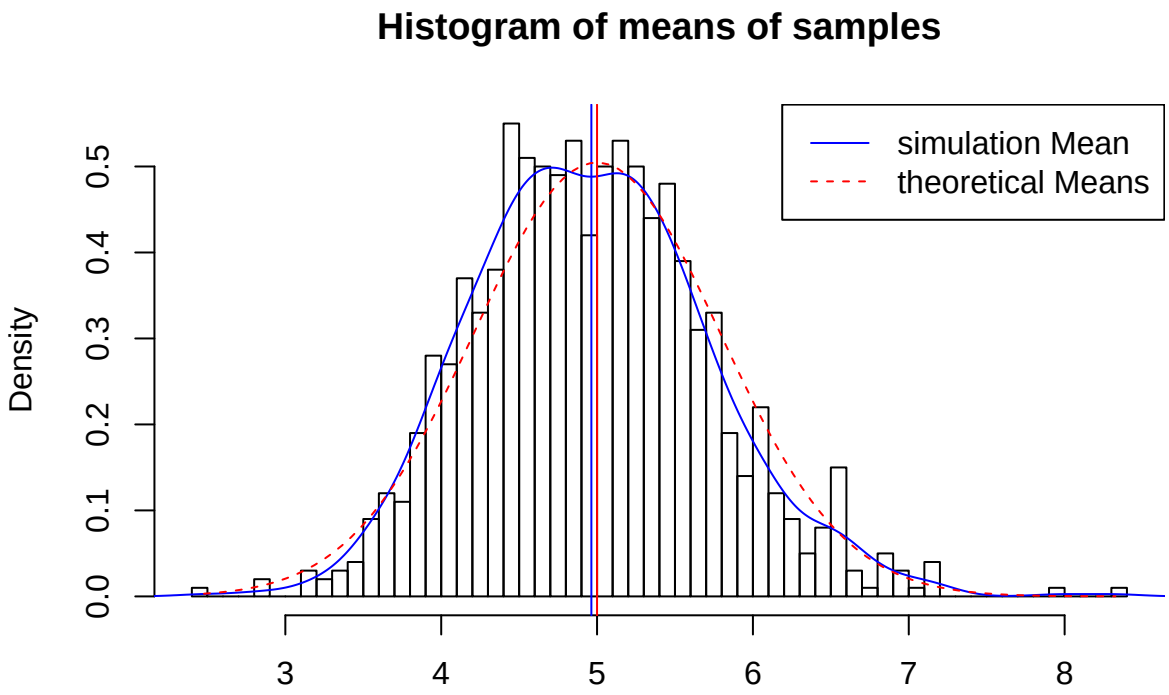
Now, let's look at the distribution of the simulated means.

```
## Take the mean of all of the simulated means
singleSimMean <- mean(simulatedMean)
singleSimMean
```

```
## [1] 4.963546
```

```
## Calculate the theoretical mean.
theoreticalmean <- 1/lambda
theoreticalmean
```

```
## [1] 5
```



The simulated means distribution is centered at 4.9635464 and the the theoretical mean is $1 / \lambda = 5$. We can see that the simulated mean is very close to the theoretical mean, and the simulated distribution is very close to the normal distribution.

Question 2 - Sample variance vs theoretical variance

```
## standard deviation of simulated distribution
simulatedSD <- sd(simulatedMean)
simulatedSD
```

```
## [1] 0.7582372
```

```
## Standard deviation of theoretical distribution
theoreticalSD <- (1/lambda)/sqrt(sampleSize)
theoreticalSD
```

```
## [1] 0.7905694
```

```r
## Variance of simulated distribution
simulatedVariance <- simulatedSD^2
simulatedVariance
```

```
## [1] 0.5749236
```

```r
## variance of theoretical distribution
theoreticalVariance <- ((1/lambda)*(1/sqrt(sampleSize)))^2
theoreticalVariance
```

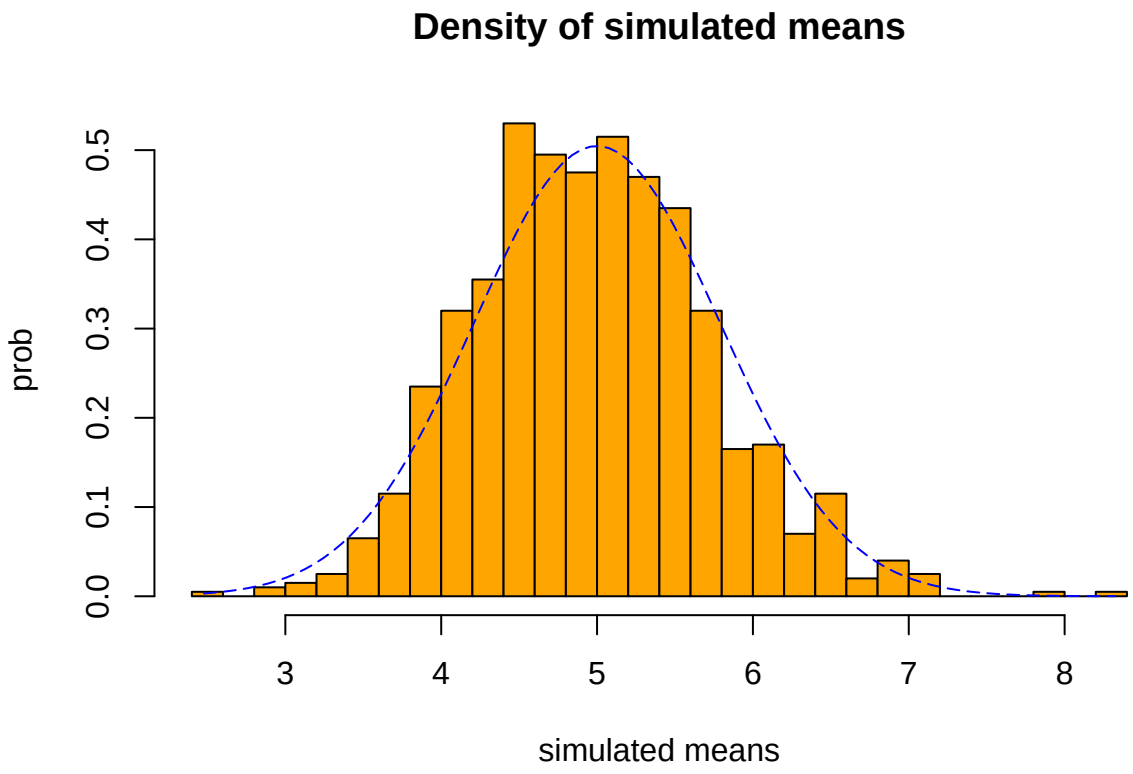
```
## [1] 0.625
```

The variance of the sample means is 0.5749236 and the variance of the theoretical distribution is 0.625.

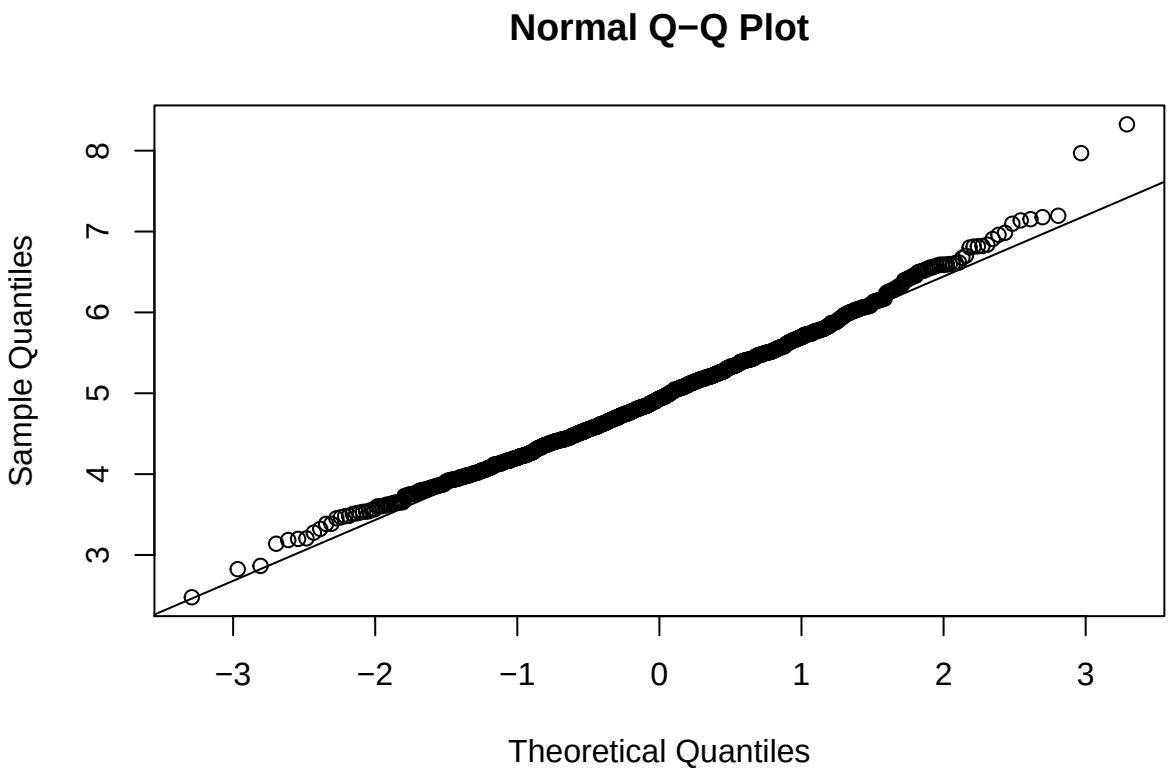
Due to the large number of samples, the distribution of simulated means approaches the normal distribution. The simulated variance and standard deviation are both lower than the theoretical variance and standard deviation.

Question 3 - sample distribution vs theoretical distribution

Show that the distribution is approximately normal.



We can see from the distribution above that the sample distribution does approach the theoretical (normal) distribution.



From the above QQ plot, we see that most of the means lie on the theoretical line. This tells us that the simulated distribution is approximately normal.