Statistical Inference Course Project - Part 1

Anthony Macey

17 November 2017

Simulationulation Exercise

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulationulated 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 simulationulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

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.

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.

```
library(ggplot2)

set.seed(63456)                     #Set Seed to allow reproducibility
lambda                  <- 0.2      #Set lambda to 0.2
number.simulations      <- 1000     #Number of simulations
number.exponentials     <- 40       #Number of Samples

## Creates a 40x1000 matrix of the simulation
simulation <-
  replicate(number.simulations, rexp(number.exponentials, lambda))
```

Results

The sample mean is close to that of the theoretical mean and this tends to be the case in samples > 30 as the data tends to become more normal. According to the CLT, the sample mean is approximately normal with mean μ and standard deviation σ. This can be observed in the following plot and tables.

The theoretical standard deviation σ of a exponential distribution of rate λ is

$\sigma = \frac{1/\lambda}{\sqrt{n}}$

```
theoretical.sd <- (1/lambda)*(1/sqrt(number.exponentials))
```

The theoretical mean μ of a exponential distribution of rate λ is

$\mu = \frac{1}{\lambda}$

```
theoretical.mean <- 1/lambda
```

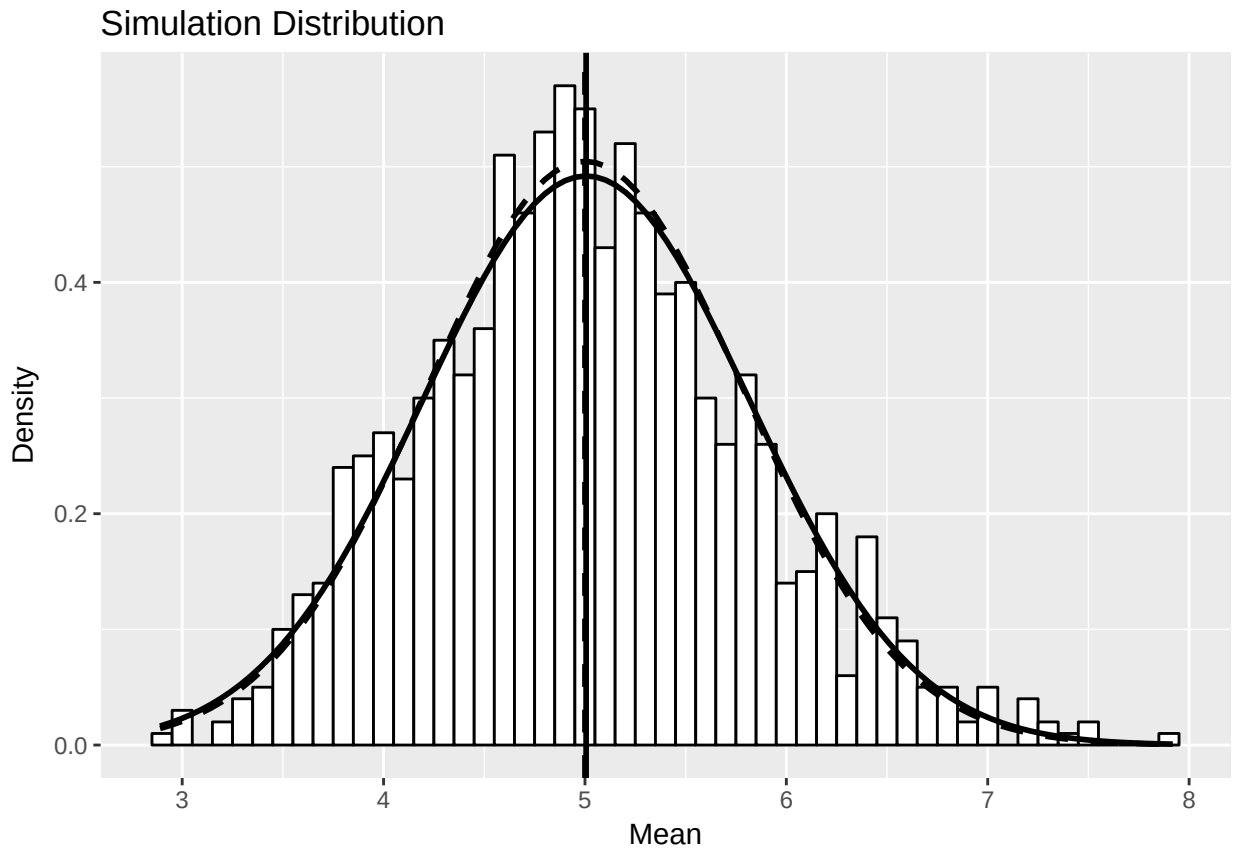
The theoretical variance Var of standard deviation σ is

```
theoretical.var <- theoretical.sd^2
```

Calculate the actual distributions for the simulation

```
simulation.mean <- apply(simulation,2,mean)
sample.mean <- mean(simulation.mean)
sample.sd <- sd(simulation.mean)
sample.var <- var(simulation.mean)
```

Description	Value
Sample Mean	5.0057656
Theoretical Mean	5
Sample Varience	0.6574716
Theoretical Varience	0.625
Sample SD	0.8108462
Theoretical SD	0.7905694



The theoretical mean and distribution are shown as dotted lines on the plot vs solid lines for the simulation.