Statistical Inference Project 1 Report
Exponential Distribution and the Central Limit Theorem

Kari Ross

April 26th, 2015

Contents

1. Overview

This report details simulations performed in R and subsequent calculations utilized to investigate the Central Limit Theorem, using the exponential distribution for the simulations

2. Simulations

In order to explore the exponential distribution and the central limit theorem, XXX number of simulations were performed.

a. Simulation 1: Sample Mean of Exponential Distribution

- 1,000 sets of 40 random exponential numbers were created using the rexp function in R, with lamba = 0.2
- The cumulative means of the 1,000 experiments were stored, and plotted.
- The mean of the population was calculated and stored in mean_simulation1

The R-code for Simulation 1 is as follows:

```
#Project1
lambda          <- 0.2       ## set parameters
num_sims        <- 1000      ## set number of simulations to run
num_randoms  <- 40          ## set number of random exponential numbers created
mns                 <- NULL     ## set initial variable to NULL
## Perform 1,000 simulations, in each simulation, calculate the mean and store
for (i in 1 : num_sims) mns <-c(mns,mean(rexp(num_randoms,lambda)))
hist(mns)                                         ## create a histogram of the simulations
mean_simulation1 <- mean(mns)  ## calculate the value of the mean of the population of
```

simulations

b. Simulation 2: Sample Variance of Exponential distribution

- 1,000 sets of 40 random exponential numbers were created using the rexp function in R, with lamba = 0.2
- The cumulative means of the 1,000 experiments were stored, and plotted.
- The mean of the population was calculated and stored in mean_simulation1

The R-code for Simulation 1 is as follows:

```
#Project1
lambda          <- 0.2       ## set parameters
num_sims        <- 1000      ## set number of simulations to run
num_randoms  <- 40        ## set number of random exponential numbers created
mns             <- NULL     ## set initial variable to NULL
## Perform 1,000 simulations, in each simulation, calculate the standard deviation and store
for (i in 1 : num_sims) mns <-c(mns, sd(rexp(num_randoms,lambda)))
hist(mns)                               ## create a histogram of the simulations
mean_simulation1 <- mean(mns)  ## calculate the value of the mean of the population of
                                        ## simulations
```

3. Results

c. Sample Mean versus Theoretical Mean

The Law of Large Numbers states that the average limits to what it's estimating.
The mean of an exponential distribution is given by

Equation 3.1: $\mu = \frac{1}{lambda}$

Substituting in lambda = 0.2 into Equation 3.1, we get

$$\mu = \frac{1}{0.2} = 5 \text{ for the theoretical value of the mean when lambda is 0.2}$$

By running the code in Section 2b, a histogram was obtained which shows the mean is centered around 5. See Appendix AFigure 1, page 4 for the histogram, which shows that the center of mass is around 5.

Additionally, the mean was calculated by using the R-code shown in Section2a.

The results for the standard deviation of the populations was: $u_{1000\_simulation} = 4.9833$

Table 3.1 (below) shows: the estimated value of μ for various numbers of simulations, as well as the delta from the known mean.

Number of Simulations	Estimated value of s	Delta from Known mean
10	5.241151	0.07520765
100	4.98721	0.05078465
1000	4.983299	0.01670110

As the central limit theorem states, as the number of simulations increases, the delta from the known mean decreases and the population approaches the true mean.

d. Results: Sample Variance versus Theoretical Variance

The Law of Large Numbers states that the average limits to what it's estimating.
The standard deviation of an exponential distribution is given by

Equation 3.2: $\sigma = \frac{1}{lambda}$

Substituting in lambda = 0.2 into Equation 3.2, we get

$$\sigma = \frac{1}{0.2} = 5 \text{ for the theoretical value of the mean when lambda is 0.2}$$

By Running the code in Section 2c, a histogram was obtained which shows the mean of the standard deviation is centered around 5. See Appendix A, Figure 2, page 5 for the histogram.
Additionally, the mean of the standard deviation was calculated by using the R-code shown in Section2b.

$$\text{The results for the mean of the populations was: } s_{1000\_simulation} = 4.94394$$

Table 3.2 (below) shows: the estimated value of μ for various numbers of simulations, as well as the delta from the known mean. As the central limit theorem states, as the number of simulations increases, the delta from the known mean decreases and the population approaches the true mean.

Number of Simulations	Estimated value of s	Delta from Known σ
10	4.9355	0.0645
100	4.93567	0.0643
1000	4.94394	0.0561

As the central limit theorem states, as the number of simulations increases, the delta from the known mean of deviations decreases and the population approaches the true mean.

4. Distribution

The Central Limit Theorem is that Xn is that the average is approximately Normally distributed with a mean given by the population mean and the variance given by the standard error of the mean.
The standard error is the standard deviation of the sample-mean estimate of a population mean. This is of the mean is given by:

$Equation\ 4.1\ \ SE_x = \frac{s}{\sqrt{n}}$ where:

s is the sample standard deviation
n is the size (number of observations) of the sample.

This estimate may be compared with the formula for the true standard deviation of the sample mean:

Equation 4.2 $SD_x = \frac{\sigma}{\sqrt{n}}$

Plugging in values from Section 3a and 3b, we can compare the estimate with the Theoretical

Estimate: $SE_x = \frac{4.94394}{\sqrt{40}} = 0.7812$ theoretical: $SD_x = \frac{5}{\sqrt{40}} = 0.79$

As we can see, the simulations of the standard error, the mean, and the standard deviation are all good approximates of the theoretical standard error, the mean, and the standard deviation

5. Appendix – Charts

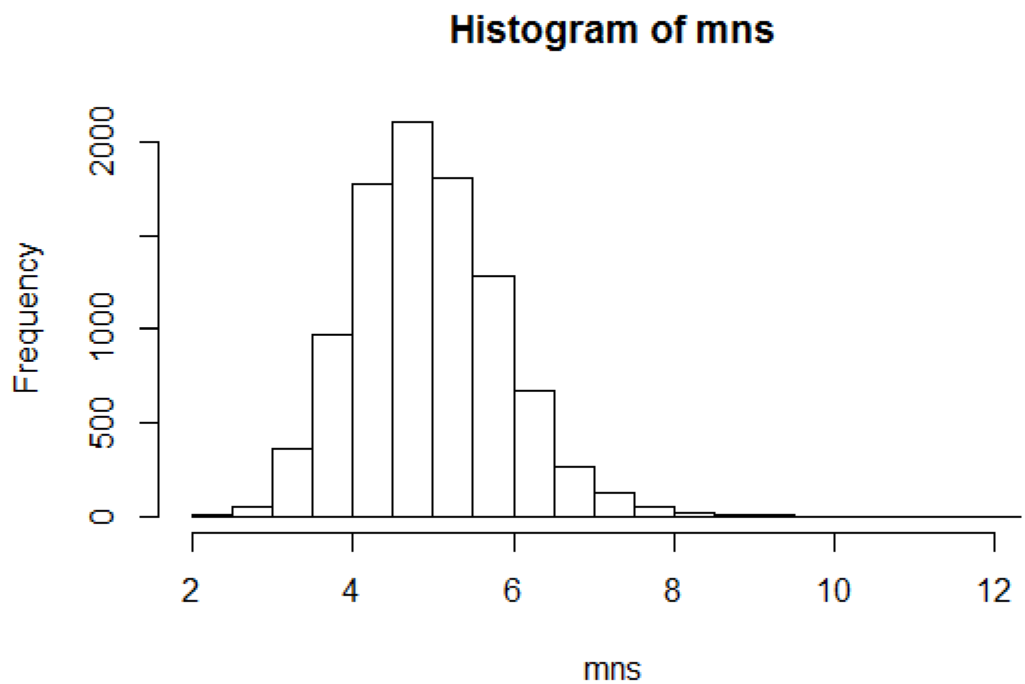

Figure 1 – Histogram of the population with lambda = 0.2 , of 1,000 simulations for the Mean of the population of 40 random exponents

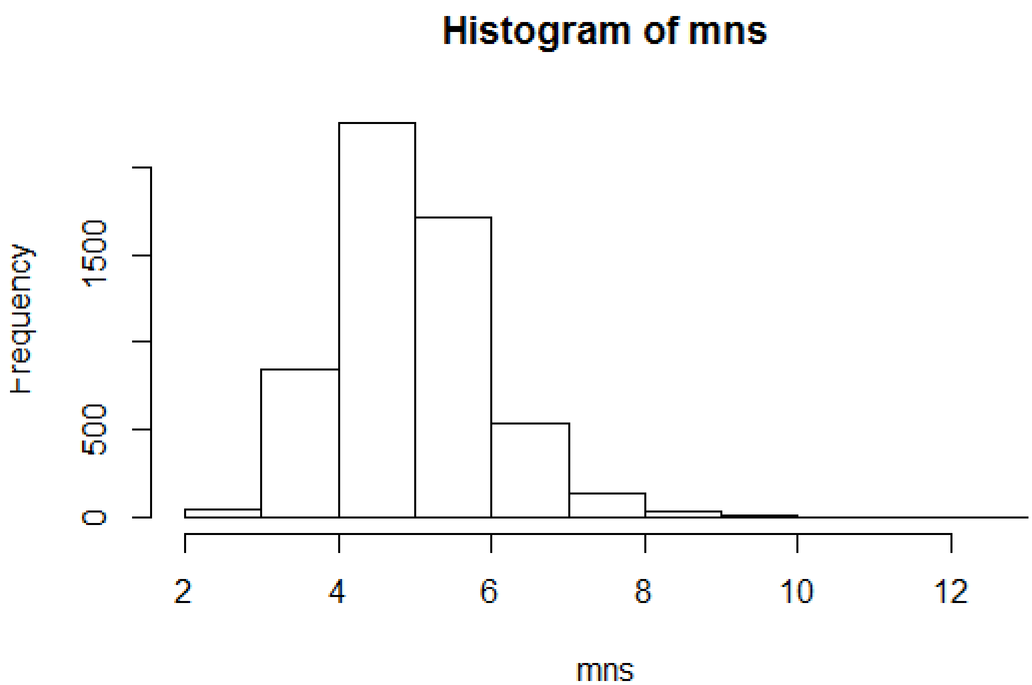

Figure 2 – Histogram of the population with lambda = 0.2 , of 1,000 simulations for the Mean of the standard deviation of population of 40 random exponents