Statistical Inference Project Part 1

Umair Ahmed Khan

24 July 2016

Overview

The purpose of this document is to analyse the exponential distribution and compare it with the Central Limit Theorem. For this purpose, we set lambda = 0.2 for all the simulations. This analysis will compare the distribution of averages of 40 exponentials over 1000 simulations.

First we set the variables for simulation lambda, exponentials, and seed (for replicating results).

```
ECHO=TRUE
set.seed(1337)
lambda = 0.2
exponentials = 40
```

Running simulations with defined variables

```
sim_means = NULL
for (i in 1 : 1000) sim_means = c(sim_means, mean(rexp(exponentials,
lambda)))
```

Sample Mean VS Theoretical Mean

Sample Mean

Mean from the simulations will give the sample mean.

```
mean(sim_means)

## [1] 5.055995
```

Theoretical Mean

The theoretical mean of an exponential distribution is defined to be as lambda^-1.

```
lambda^-1

## [1] 5
```

Comparison

There is only a marginal difference between the simulations sample mean and the exponential distribution theoretical mean.

```r
abs(mean(sim_means)-lambda^-1)

## [1] 0.05599526
```

Sample Variance VS Theoretical Variance

Sample Variance

The variance from the simulation means will give the sample variance.

```r
var(sim_means)

## [1] 0.6543703
```

Theoretical Variance

The theoretical variance of an exponential distribution is (lambda * sqrt(n))^-2.

```r
(lambda * sqrt(exponentials))^-2

## [1] 0.625
```

Comparison

There is only a marginal difference between the simulations sample variance and the exponential distribution theoretical variance.

```r
abs(var(sim_means)-(lambda * sqrt(exponentials))^-2)

## [1] 0.0293703
```

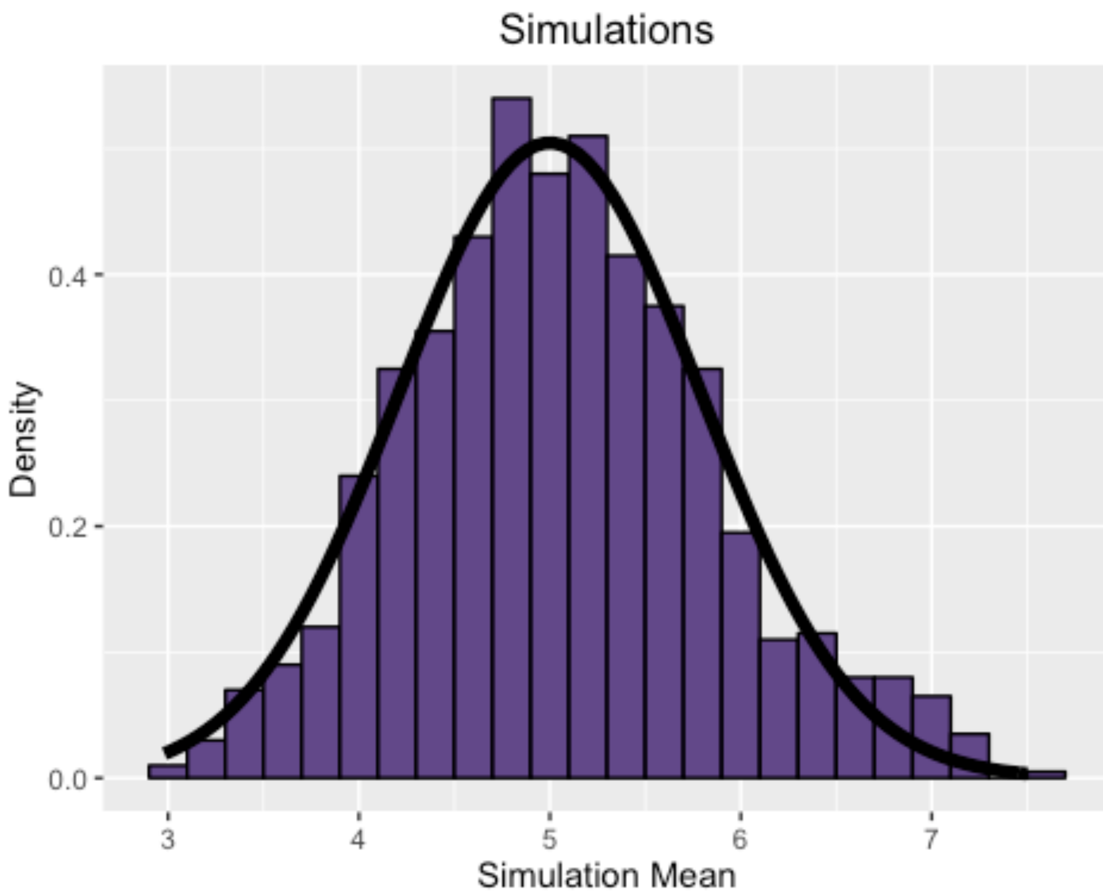
Distribution

This following is a density histogram of 1000 simulations. There is an overlay with a normal distribution that has a mean of lambda^-1 and standard deviation of (lambda*sqrt(n))^-1, the theoretical normal distribution for the simulations.

```r
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.2.4

ggplot(data.frame(y=sim_means), aes(x=y)) +
  geom_histogram(aes(y=..density..), binwidth=0.2, fill="#624889",
                 color="black") +
  stat_function(fun=dnorm, args=list(mean=lambda^-1,
                                     sd=(lambda*sqrt(exponentials))^-1),
                size=2) +
  labs(title="Simulations", x="Simulation Mean", y="Density")
```


Simulations
Density
0.0
0.2
0.4
3
4
5
6
7
Simulation Mean