Statistical Inference - Course Project (Part 1)

Overview

This is the project for the statistical inference class. In it, you will use simulation to explore inference and do some simple inferential data analysis. The project consists of two parts:

1. A simulation exercise.
2. Basic inferential data analysis.

Instructions

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials.

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.

Simulations

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated 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.

```r
# Set lambda to 0.2
lambda = 0.2

# 40 samples
samples = 40

# 1000 simulations
sim = 1000

# Set seed
set.seed(100)

# Running the simulations
simulation = matrix(rexp(sim*samples, rate=lambda), sim, samples)

# Getting the means
means = rowMeans(simulation)
```

Analysis

1. Sample Mean versus Theoretical Mean Show the sample mean and compare it to the theoretical mean of the distribution.

```
# Sample Mean
mean(means)
```

```
## [1] 4.999702
```

```
# Theoretical Mean
1/lambda
```

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

The mean of the sample means from the simulations is 4.999702 and very close to the theoretical mean of 5.0.

2. Sample Variance versus Theoretical Varianc Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
# Sample Variance
var(means)
```

```
## [1] 0.6335302
```

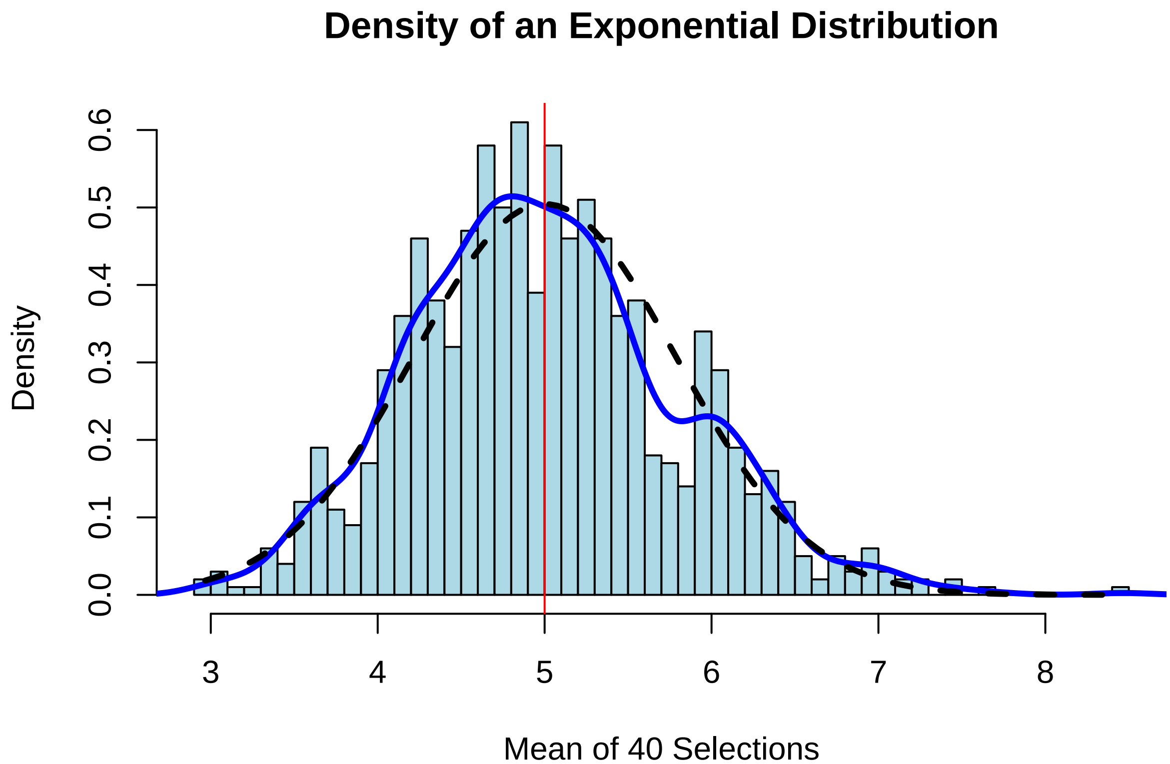
```
# Theoretical Varianc
((1/lambda)/sqrt(samples))^2
```

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

The variance of the sample is 0.6335302 and therefore also very close to the theoretical variance of 0.625.

3. Distribution Show that the distribution is approximately normal.

```
hist(means, col = "lightblue", breaks = 40, freq=FALSE, main="Density of an Exponential Distribution",xl
lines(density(means),lwd=3, col="blue")
# Theoretical Center
abline(v=1/lambda, col="red", lwd=1)
# Density of the Averages
x <- seq(min(means), max(means), length=100)
y <- dnorm(x, mean=1/lambda, sd=(1/lambda/sqrt(samples)))
lines(x,y, pch=19, col="black", lwd=3, lty=2)
```



Due to the central limit theorem, the averages of samples follow a normal distribution. The plot above shows the density of the simulation (in blue) and the normal density plotted with theoretical mean and variance values (in black).