

Statistical Inference Week 4 Project

Grejell Segura

8/13/2017

Statistical Inference - Week 4 Assignment: Course Project Part 1

Overview

In this report I will investigate and compare the theoretical values of the exponential distribution parameters to the actual observed values. In addition, it will also show the visual comparison comparison of the values.

The project consists of two parts:

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

Load necessary libraries

```
library(ggplot2)
set.seed(1234321)
```

Assign `n` as the number of exponential distributions and `lambda` as the rate parameter. Moreover, the number of simulations is set to 5000.

```
n <- 40
lambda <- 0.2
simulations <- 5000 ## number of simulations ##
```

Create the data set.

```
data.1 <- matrix(rexp(n*simulations, lambda), simulations)
```

Identify the theoretical mean and actual calculated mean and compare to each other.

```
## compare the theoretical parameters vs the actual observed parameters ##
```

```
## 1. mean
```

```
Row.Mean <- apply(data.1, 1, mean)
```

```
observed.Mean <- mean(Row.Mean)
```

```
theory.Mean <- 1/lambda
```

```
print(observed.Mean) ## the calculated mean
```

```
## [1] 4.995456
```

```
print(theory.Mean) ## the theoretical mean
```

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

As observed, both mean values are close to each other.

Next is the theoretical standard deviation and variance compared to its actual calculated values.

```
## 2. variance
```

```
observed.Var <- var(Row.Mean)
```

```
observed.Std <- sd(Row.Mean)
```

```

theory.Std <- ((1/lambda) * (1/sqrt(n)))
theory.Var <- theory.Std^2

print(observed.Std) ## the calculated standard deviation

## [1] 0.7856688
print(theory.Std) ## the theoretical standard deviation

## [1] 0.7905694
print(observed.Var) ## the calculated variance

## [1] 0.6172755
print(theory.Var) ## the theoretical variance

## [1] 0.625

```

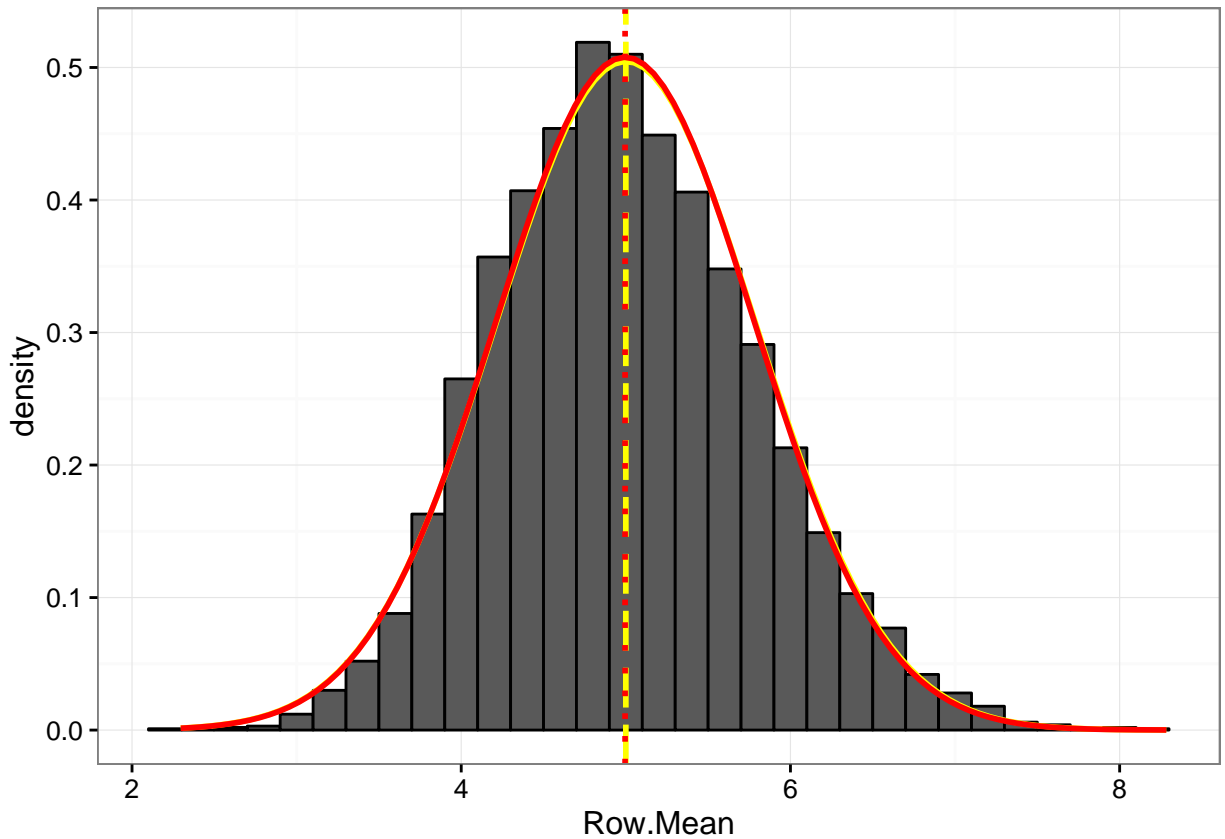
Clearly, the values are very close to each other as expected.

To make it more convincing, we try to visualize the distribution and plot the theoretical and calculated parameters below.

```

## visualization
Row.Mean <- as.data.frame(Row.Mean)
plot1 <- ggplot(Row.Mean, aes(x = Row.Mean)) + geom_histogram(binwidth = lambda, color = "black", aes(y
plot1 <- plot1 + geom_vline(xintercept = theory.Mean, size=1.0, color = "yellow", linetype = "longdash")
plot1 <- plot1 + geom_vline(xintercept = observed.Mean, size=1.0, color = "red", linetype = "dotted")
plot1 <- plot1 + stat_function(fun = dnorm, args = list(mean = theory.Mean, sd = theory.Std), color = "
plot1 <- plot1 + stat_function(fun = dnorm, args = list(mean = observed.Mean, sd = observed.Std), color
plot1 <- plot1 + theme_bw()
plot1

```



The histogram shows the distribution of the mean of the 40 exponential distributions simulated 5000 times. The shape is representative of a bell curve which is a proof that the distribution is normally distributed and follow the Central Limit Theorem. The red line and bell curve represent the plot for the actual observed mean and standard deviation while the yellow line and bell curve represent the theoretical values. The difference as we see is very minimal, which is an evidence that the values are very much similar.

Conclusion

This report shows the difference between the actual calculated and theoretical parameters of an exponential distribution. The difference is very small which shows that the values are similar. It also shows that the distribution of the parameters follow a normal distribution which supports the Central Limit Theorem.