Statistical Inference Course Project

Course : Exploratory Data Analysis
Document : Statistical Inference Course Project
Author : Marcelo Dominguez

Overview

In this project I 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.
I will set lambda = 0.2 for all of the simulations and I will investigate the distribution of averages of 40 exponentials.

Note that I will need to do a thousand simulations.

Simulations

This point includes explanations of the simulations I ran, with the accompanying R code.

```
# 1. Loading required libraries.
library(ggplot2)

# 2. Setting up simulation constants.
const_lambda = 0.2
const_exponentials = 40
const_simulations = 1000

# 3. Setting up seed for reproducability.
set.seed(40020015)

# 4. Running defined simulations.
simulated_exponentials = replicate(const_simulations, rexp(const_exponentials, con
st_lambda))

# 5. Calculating mean for defined exponentials.
sample_distribution_means <- apply(simulated_exponentials, 2, mean)
```

Sample Mean versus Theoretical Mean

This point shows where the distribution is centered at and a comparison with the theoretical center of the distribution.

```
# 6. Calculating mean of distribution.
sample_distribiution_mean = mean(sample_distribution_means)
sprintf("Current mean for exponential distribution with range %.1f is = %1f",const
_lambda,sample_distribiution_mean)
```

```
## [1] "Current mean for exponential distribution with range 0.2 is = 4.981854"
```

The Theoretical mean (**theorical_mean**) for a exponential distribution with rate **const_lambda** is: 1 / const_lambda

```
# 7. Calculating Theoretical Mean.
theorical_distribution_mean = 1 / const_lambda
sprintf("Theorical mean for exponential distribution with range %.1f is = %1f",con
st_lambda,theorical_distribution_mean)
```

```
## [1] "Theorical mean for exponential distribution with range 0.2 is = 5.000000"
```

The center of distribution of averages of **const_exponentials** (default=40) exponentials is close to the theoretical center of the distribution.

Sample Variance versus Theoretical Variance

This point shows where sample distribution variance is located at and a comparison with the theoretical variance of exponential distribution.

First, we´re calculating standard deviation of sample distribution.

```
# 8. Calculating standard deviation of sample distribution.
sample_distribution_variance = sd(sample_distribution_means)
sprintf("Current variance for sample distributions with range %.1f is = %1f",const
_lambda,sample_distribution_variance)
```

```
## [1] "Current variance for sample distributions with range 0.2 is = 0.773110"
```

Now I´m getting standard deviation of theorical exponential distribution with rate **const_lambda** is: (1 / const_lambda) / sqrt(const_exponentials)

```
# 8. Calculating theoretical variance of exponential distribution and given consta
nts.
theorical_distribution_variance = (1/const_lambda) / sqrt(const_exponentials)
sprintf("Theorical variance for exponential distributions with range %.1f is = %1f
",const_lambda,theorical_distribution_variance)
```

```
## [1] "Theorical variance for exponential distributions with range 0.2 is = 0.790
569"
```

The standard deviations are very close.
Since variance is the square of the standard deviations minor diffrences will we enhanced but the´re still so close.

Distribution

By cheking next plot we can see that the distribution of averages of 40 exponentials is very close to a normal distribution.