Investigation of the Exponential Distribution in R and its Comparison with the Central Limit Theorem

Shayan (Sean) Taheri

June 08, 2019

Overview

In this data analysis project, the goal is investigating the exponential distribution and compare it to the Central Limit Theorem. Through this analysis, the lambda will be set to 0.2 for all the simulation. This investigation will compare compare the distribution of aversages of 40 exponential over 1000 simulation.

Simulations

Set the simulation variables for simulation: lambda, exponentials, and seed.

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

Let's run the simulations with the variables.

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

Sample Mean versus Theoretical Mean

Sample Mean

Calculation of the mean from the simualtions with the given sample mean value.

```
mean(simMeans)
```

```
## [1] 5.055995
```

Theoretical Mean

We define the theoretical mean of an exponential distribution as lambda to the power of -1.

```
lambda^-1
```

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

Comparison

We have only a small difference between the sample mean from simulations and the theoretical mean from the exponential distribution.

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

```
## [1] 0.05599526
```

Sample Variance versus Theoretical Variance

Sample Variance

Calculation of the variance from the simulation means and the given sample variance.

```r
var(simMeans)
```

```
## [1] 0.6543703
```

Theoretical Variance

We can define the theoretical variance of an exponential distribution as multiplication of lambda and the squared-root of "n", all to the power of -2.

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

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

Comparison

We have a small difference between the simulations sample variance and the theoretical variance from the exponential distribution.

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

```
## [1] 0.0293703
```

Distribution

A density histogram of 1000 simulations. We have an overlay with a normal ditribution with a mean of lambda to the power of -1.

The standard deviation is defined as multiplication of lamda and the squared-root of "n" to the power of -1. This is the theoretical normal distribution for the simulations.

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

```r
ggplot(data.frame(y=simMeans), aes(x=y)) +
  geom_histogram(aes(y=..density..), binwidth=0.2, fill="#0072B2",
                 color="black") +
  stat_function(fun=dnorm, arg=list(mean=lambda^-1,
                                    sd=(lambda*sqrt(exponentials))^-1),
                size=2) +
  labs(title="Plot of the Simulations", x="Simulation Mean")
```

```
## Warning: Ignoring unknown parameters: arg
```