Part 1: Investigation of the exponential distribution in R and comparison with the Central Limit Theorem

Elena Fedorova

11 august 2018

1. Overview

The key aim of this report is to investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution will 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. For this purpose a thousand of exponential simulations will be performed with lambda = 0.2 for averages of 40 exponentials.

2. Simulations

Code simulation with the described parameters:

```
set.seed(800)
lambda = 0.2
exponentials = 40
nsim = 1:1000
```

Run the required number of simulations:

```
simMeans <- data.frame(x = sapply(nsim, function(x) {mean(rexp(exponentials,
lambda))}))
head(simMeans)

##          x
## 1 6.191990
## 2 4.841509
## 3 5.949288
## 4 4.454205
## 5 6.233570
## 6 5.049980
```

3. Sample Mean versus Theoretical Mean

Sample Mean

Calculating the mean from the simulations with give the sample mean.

```
mean(simMeans$x)

## [1] 5.013284

sd(simMeans$x)
```

```
## [1] 0.7857452
```

Theoretical Mean

The theoretical mean of an exponential distribution is 1/lambda.

```
(1/lambda)

## [1] 5
```

Comparison

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

```
abs(mean(simMeans$x)-(1/lambda))

## [1] 0.01328389
```

4. Sample Variance versus Theoretical Variance

Sample Variance

Calculating the variance from the simulation means with give the sample variance.

```
var(simMeans$x)

## [1] 0.6173955
```

Theoretical Variance

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

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

## [1] 0.625
```

Comparison

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

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

## [1] 0.007604453
```

5. Distribution

Here we'll plot a density histogram for the 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.

```
library(ggplot2)
ggplot(simMeans, aes(x=x)) +
  geom_histogram(aes(y=..density..), binwidth=0.25,
fill="steelblue",color="darkblue") +
stat_function(fun = dnorm, args = list(mean = (1/lambda),
sd=(lambda*sqrt(exponentials))^-1), size = 1, color = "midnightblue") +
        labs(title="Plot of the Simulations", x="SimMean")
```