Course Project Statistical Inference, Part 1

Author: Hendrik L.

August 2015

Overview

This project part investigates the exponential distribution in R and compares it with the Central Limit Theorem. The exponential distribution can be simulated with rexp(n, lambda) where lambda is set to `lambda = 0.2` for all simulations. Focus is to investigate the distribution of averages of 40 exponentials by means of 1,000 simulations.

Simulation

Exponential distributions can be simulated in R with `rexp(n, lambda)` where `lambda` is the rate parameter. Both the mean and the standard deviation of the exponential distribution are: `1/lambda`

For all simulations, lambda is set to `lambda = 0.2`. The following 1,000 simulations in R are run in order to investigate the distribution of averages of 40 exponentials:

```
set.seed(164720)
lambda = 0.2
exponentials = 40
simulationMeans = vector()
for (i in 1:1000) { simulationMeans = c(simulationMeans, mean(rexp(exponentials, lambda))) }
```

1. Sample mean versus theoretical mean

1.1. Sample mean

Calculation of the mean from the simulations which is the sample mean:

```
mean(simulationMeans)
```

```
## [1] 5.02089
```

1.2. Theoretical mean

The theoretical mean of an exponential distribution is: `lambda^-1`

```
lambda^-1
```

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

1.3. Conclusion

There is only a small difference between the mean of the simulation samples and the theoretical mean of the exponential distribution.

```
abs(mean(simulationMeans) - lambda^-1)
```

```
## [1] 0.02089017
```

2. Sample variance versus theoretical Variance

2.1. Sample variance

```
var(simulationMeans)
```

```
## [1] 0.6544698
```

2.2. Theoretical variance

The theoretical variance of the exponential distribution is: `(lambda * sqrt(n))^-2`

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

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

2.3. Conclusion

There is only a small difference between the variance of the simulation samples and the theoretical variance of the exponential distribution.

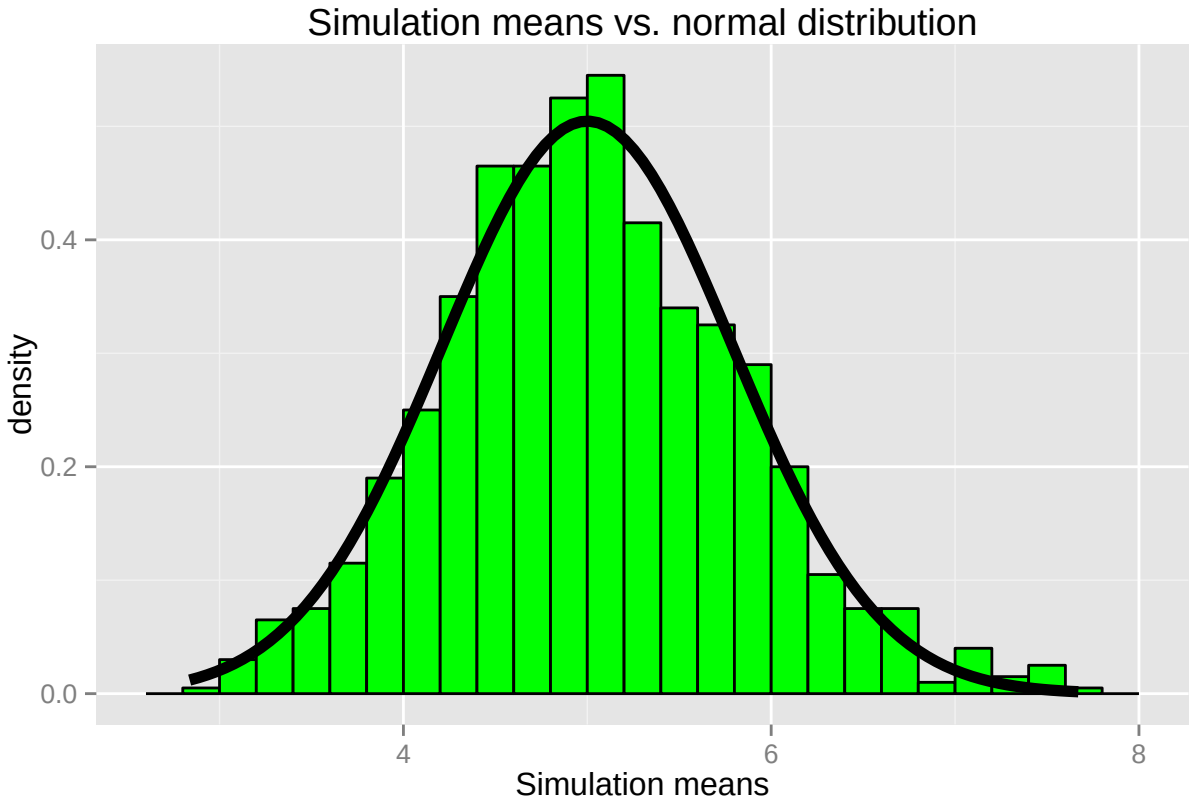
```
abs(var(simulationMeans) - (lambda * sqrt(exponentials))^-2)
```

```
## [1] 0.0294698
```

3. Distribution

The following plot shows a histogram of the 1,000 simulations. In an overlay, also the normal distribution is shown that has a mean of `lambda^-1` and a standard deviation of `(lambda * sqrt(n))^-1` which is the theoretical normal distribution for the simulations. As expected according to the Central Limit Theorem, the 1,000 simulations of averages of 40 exponentials are approximately normally distributed.

```
ggplot(data.frame(y = simulationMeans), aes(x = y)) +
    geom_histogram(aes(y = ..density..), binwidth = 0.2,
        fill = "green", color = "black") +
    stat_function(fun = dnorm, arg = list(mean = lambda^-1,
    sd = (lambda * sqrt(exponentials))^-1), size = 2) +
    labs(title = "Simulation means vs. normal distribution", x = "Simulation means")
```


Simulation means vs. normal distribution
density
0.4
0.2
0.0
4
6
8
Simulation means