Exponential Distribution and the CLT

Bankbintje

24th May 2015

Overview

The Central Limit Theorem states that the mean of a sufficiently large number of iterates of iid variables will be approximately normally distributed, *regardless* of the underlying distribution. This simulation excercise runs 1000 simulations of exponential distributions to investigate this. This paper compares the theoretical average and variance with the observed/simluated values *(questions 1 and 2)* and whether the resulting distribution is approximately normal *(question 3)*.

Simulations

First, set some basic variables and call needed libraries.

Next, run the 1000 simulations, using lambda = 0.2, nbr.of.samples = 40 and nbr.of.simulations = 1000.

```
for (i in 1:nbr.of.simulations) {
        simulation.mean <- c(simulation.mean, mean(rexp(nbr.of.samples,lambda)))
        simulation.variance <- c(simulation.variance, var(rexp(nbr.of.samples,lamb
da)))}
```

Sample Mean versus Theoretical Mean

Question 1 - Show the sample mean and compare it to the theoretical mean of the distribution.
The **theoretical mean** is calculated as: 1/lambda. In this case 1/0.2 = 5

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

The **sample mean** is calculated as the mean of the 1000 simulated means.

```
## [1] 5.006928
```

The difference between theoretical and sample mean is very small.

```
## [1] 0.006928149
```

Sample Variance versus Theoretical Variance

Question 2 - Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
There appear to exist two interpretations of this question. I decided to include both as it illustrates the effect of choosing/interpreting your estimator and estimand.

Interpretation 1: In this scenario is the *Estimator* the mean of the 1000 simulated variances and the *Estimand* the variance of the (exponentially distributed) population.

The **theoretical variance** is calculated as: 1/lambda^2. In this case (1/0.2)^2 = 5^2 = 25

```
## [1] 25
```

The **sample variance** is calculated as the mean of the 1000 simulated variances.

```
## [1] 24.44498
```

The **difference** between theoretical and sample variance is small.

```
## [1] 0.5550208
```

Interpretation 2: In this interpretation is the *Estimator* the variance of the 1000 simulated means. The *Estimand* is the variance of the (normally distributed) population of simulations.

The **theoretical standard deviation** calculated as: (1/lambda)/sqrt(n).

```
## [1] 0.7905694
```

The **theoretical variance** calculated as: ((1/lambda)/sqrt(n))^2.

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

The **sample standard deviation** is taken from the simulation data.

```
## [1] 0.7691207
```

The **sample variance** is taken from the simulation data.

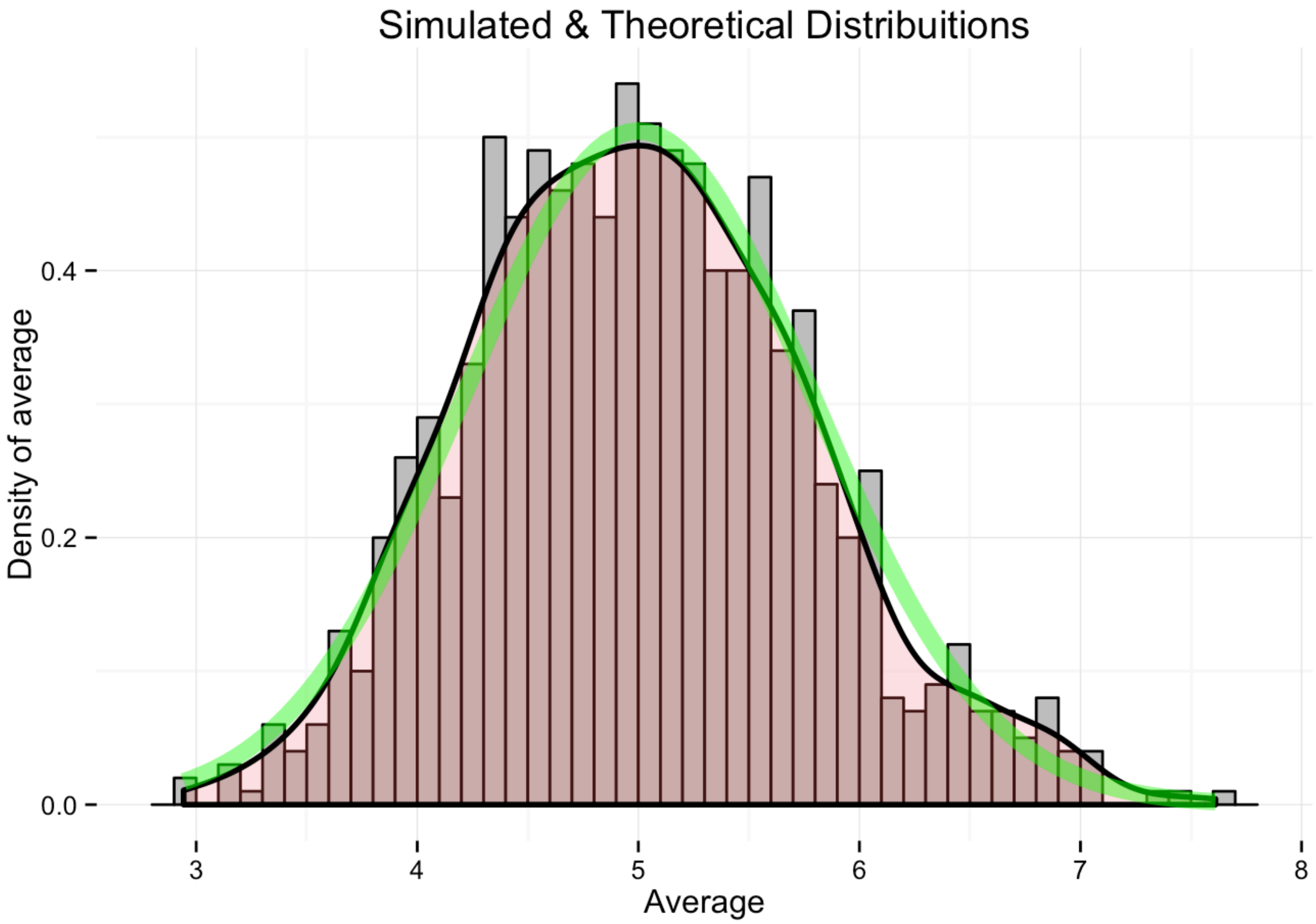
```
## [1] 0.5915466
```

Again, the **difference** between theoretical and sample variance is very small.

```
## [1] 0.03345337
```

Distribution

Question 3 - Show that the distribution is approximately normal.
The following plot shows the simulated distribution (bars & black line), combined with the theoretical distribution (the green line).



The distribution of simulated means approximates a normal distribution. This is more or less as the Central Limit Theorem predicted. A large sample does provide a good approximation of the population mean - regardless of the underlying distribution!

Appendix

The code used to create the plot:

```
# calculate the mean of the 1000 simulated means of 40 samples
mean.simulation.mean<-mean(simulation.mean)
# put the results in a data frame
dfsm = as.data.frame(simulation.mean)
# draw the plot using ggplot2
ggplot (dfsm) + aes(x=dfsm$simulation.mean)+ geom_histogram(fill="grey",binwidth=.
1, color="black", aes (y=..density..)) + ylab ("Density of average") + xlab("Avera
ge") + theme_minimal()  + scale_x_continuous(breaks=c(2:8)) + ggtitle("Simulated &
Theoretical Distribuitions") +  geom_density(size = 1, alpha=.2, fill="#FF6666") +
stat_function(fun=dnorm,args=list(mean=1/lambda, sd=(1/lambda)/sqrt(40)),color = "
green", size = 3.0, alpha=.50)
```