

Repeated Exponential Distribution vs Central Limit Theorem.

Vladimir Iglovikov

02/13/2015

Introduction

In this project I will compare theoretical results for the Central limit theorem http://en.wikipedia.org/wiki/Central_limit_theorem with repeated numerical simulations.

Set up:

- Distribution that I work with - exponential http://en.wikipedia.org/wiki/Exponential_distribution.
- Mean $\lambda = 1/\mu = 1/0.2 = 5$
- Standard deviation $sd = 1/\mu = 1/0.2 = 5$
- Number of simulations $N = 1000$.
- Average will be over $n = 40$ exponents.

Numerical simulations

```
ld = 0.2 #lambda - parameter of the distribution
N = 1000 #Number of simulations
n = 40
mu = 1 / ld #theoretical value of the mean

mns = NULL
for (i in 1:N) {
  mns = c(mns, mean(rexp(n, ld)))
}

standard.error <- sd(mns) / sqrt(length(mns)) # standard error of the sampled data
```

Now I will compare mean, variance of the simulated data and theoretical predictions from the CLT.

```
print(c("Mean of the exponentials = ", toString(mean(mns))))
```

```
## [1] "Mean of the exponentials = " "4.97390366644904"
```

```
print(c("Theoretical mean = ", toString(mu)))
```

```
## [1] "Theoretical mean = " "5"
```

```
print(c("Difference between sampled and theoretical mean = ", toString(mean(mns) - mu)))
```

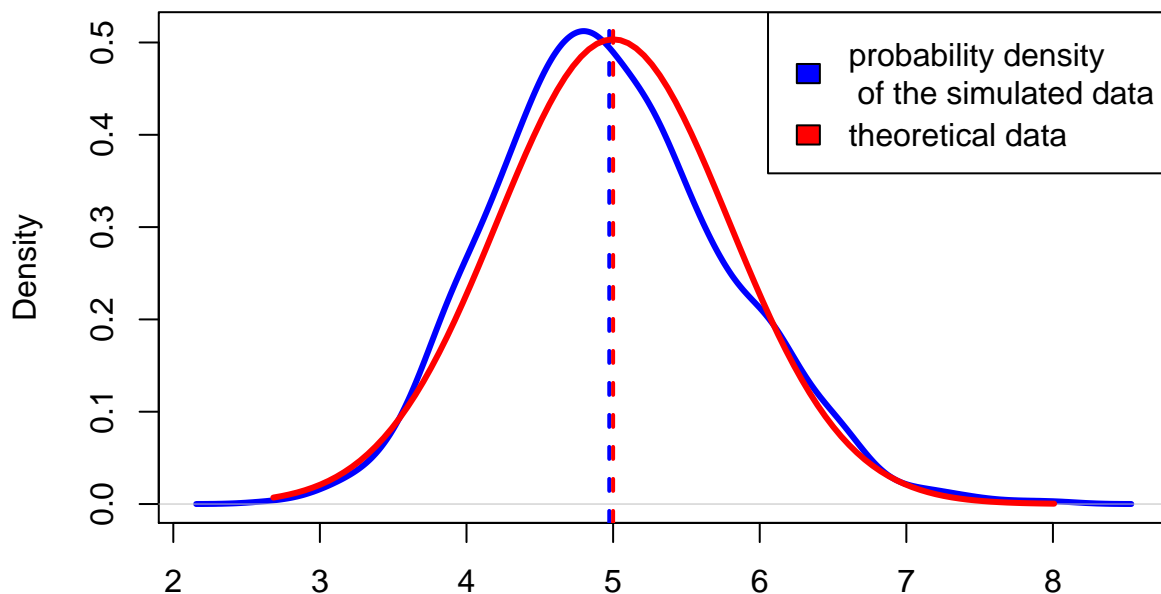
```
## [1] "Difference between sampled and theoretical mean = "  
## [2] "-0.0260963335509583"
```

```
print(c("Standard error of the sampled data = ", toString(standard.error)))
```

```
## [1] "Standard error of the sampled data = "  
## [2] "0.0250575967272042"
```

```
x <- mns  
y.list = density(x)  
plot(y.list, col="blue", main="Comparison of the theoretical and simulated data", lwd=3)  
abline(v=mean(mns), lty=2, col="blue", lwd=2)  
  
xfit <- seq(min(mns), max(mns), length=100)  
yfit <- dnorm(xfit, mean = 1 / ld, sd=sd(x))  
  
lines(xfit, yfit, col="red", lwd=3)  
abline(v=1/ld, lty=2, col="red", lwd=2)  
  
legend("topright", c("probability density\n of the simulated data",  
                     "theoretical data"), fill=c("blue", "red"))
```

Comparison of the theoretical and simulated data



N = 1000 Bandwidth = 0.1754

Plot of the histogram of the sampled data + gaussian curve that is predicted by the Central Limit Theorem. Dashed vertical lines correspond to the corresponding means.

Conclusion

As we can see from the plot above, probability density of the generated data is very close to the normal data. Mean and standard deviation are almost identical. It can be shown from the numerical comparison and from the figure.