

Comparing Exponential Distribution with CLT

Tiago V R da Costa

14/06/2015

Overview

It will be shown comparisons between Exponential distribution and Central Limit Theorem (CLT). Investigation is about assumption that under a large number of simulations, according to Law of Large Numbers, Exponential distribution converges to Gaussian distribution. Therefore, analysis will compare sample mean and theoretical mean, then sample variance and theoretical variance. At last, generate plots comparing data.

This simulation is part of Statistical Inference Coursera Project, part I.

Simulations

First, let's simulate an exponential distribution with $n = 40$ and $\lambda = 0.2$ and repeat simulation in 1000 times:

```
#Reproducible seed
set.seed(37)

#Simulation parameters
lambda <- 0.2
n = 40
rep <- 1000

#exponential distribution replicated 1000 times
exp <- replicate(rep, rexp(n, lambda))

#Resultant mean
means_exp <- apply(exp, 2, mean)
head(means_exp)
```

```
## [1] 6.080768 5.116062 5.517271 4.049645 5.396198 5.521925
```

Sample Mean versus Theoretical Mean

Question 1 states: Show the sample mean and compare it to the theoretical mean of the distribution. After simulation, we compare sample mean and theoretical mean:

```
sample_mean <- mean(means_exp)
sample_mean
```

```
## [1] 5.036936
```

```
theoretical_mean <- 1/lambda
theoretical_mean
```

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

Answer: generated sample mean and theoretical mean are very close.

Sample Variance versus Theoretical Variance

Question 2 states: Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
sample_variance <- var(means_exp)
sample_variance
```

```
## [1] 0.6478333
```

```
theoretical_variance <- ((1/lambda)/sqrt(n))^2
theoretical_variance
```

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

Answer: Sample variance and theoretical variance also are very close.

Approximately Normal Distribution

Finally, question 3 states: Show that the distribution is approximately normal. It will be generated exponential density and normal density line and compare them.

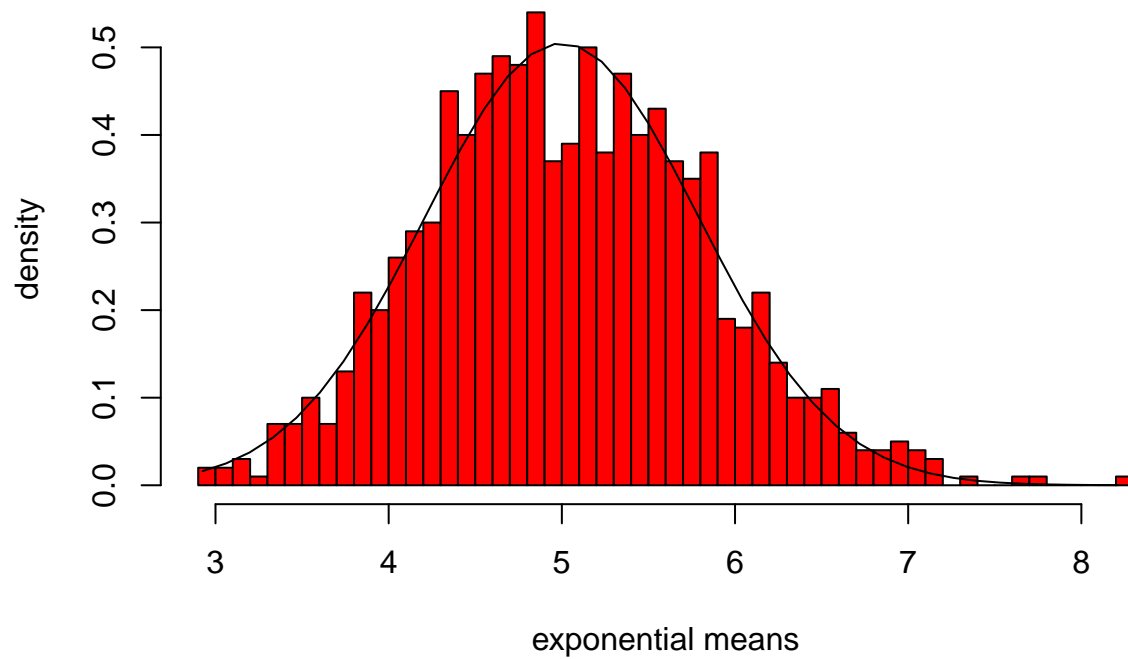
```
#X-Axis: Generating 40 x points using exponential means
x <- seq(min(means_exp),max(means_exp), length = n)

#Y-Axis: Using theoretical mean and variance as Gaussian parameters
y <- dnorm(x,mean = theoretical_mean, sd = sqrt(theoretical_variance) )

#Exponential means histogram
hist(means_exp, breaks = n, main = "Exponential means distribution"
     , col="red", xlab = "exponential means", ylab = "density", freq=FALSE)

#Normal Distribution line
lines(x, y, pch=22, col="black", lty=1)
```

Exponential means distribution



Answer: Replicating 1000 times 40 exponentials will approximate Gaussian distribution.

Summary

This work is just about validate that exponential distribution converges to a normal distribution over a large number of simulations, according to the Law of Large Numbers.