

Basic Inference Analysis Part 1/2

Ameen AboDabash

May 29, 2016

Setting the Scene

This Analysis come into two parts, the first part will be covered in this document and it about Exponential Distribution vs CLT (Central Limit Theorem), the second part ,we're going to analyze the ToothGrowth data in the R datasets package, each analysis covered in separate report.

Part 1

Exponential Distribution Model Data Sample Dist vs CLT

Background: The Exponential Distribution models “time until failure” of, say, lightbulbs. It is parametric by a constant parameter λ called the failure rate (or rate parameter as declared here) that is the average rate of lightbulb burnouts.

Objectives:

- 1-Show the sample mean and compare it to the theoretical mean of the distribution.
- 2-Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
- 3-Show that the distribution is approximately normal.

Setting up simulated data

Giving $\lambda = 0.2$, we'll simulate a total of 40 exponentiation 1000 times ($n = 40$, simulations = 1000) using R expression `rexp(n, λ)` as following:

```
## setting up known vars
lambda = .2
n = 40
## numeric vector to store simulated data
mns = NULL

for (i in 1 : 1000) mns <- c(mns, mean(rexp(40,lambda)))
```

Sample Mean vs Theoretical Mean

Strait forward question, lets calculate μ of above sample of then using R “mean” then will calculate it theoretically giving $1/\lambda$ it should be same or too close:

```
## [1] "The sample means converged to: 5.014947"
```

```
## [1] "The Thearotical sample means converged to: 5.000000"
```

So theirs no big deference between the theory result and real simulated data sample mean, So lets visualize whatever we have in hand :

```

hist(mns,probability = TRUE,col="gray", #breaks = c(1,40),
     main= expression(paste("Sample ",mu, ' Distribution')),
     xlab = expression(paste("1000 Sample ",mu, ' of 40 Exp Distribution'))
)

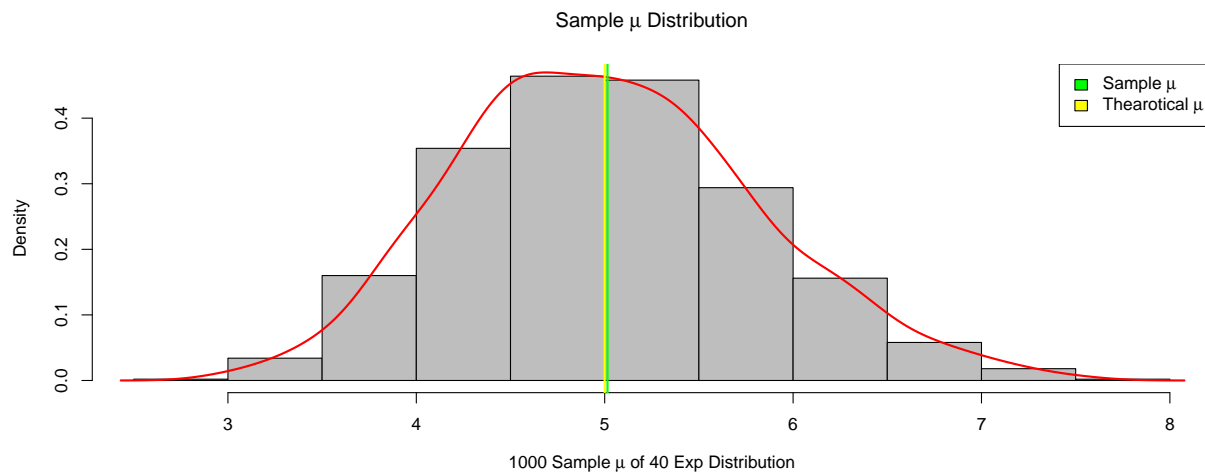
## density plot
lines(density(mns), lwd = 2, col = "red")

## Draw the means
abline(v=smpl_mean,col="green",lwd=2)

## Thearotical mean
abline(v=theory_mean,col="yellow",lwd=2)

legend("topright", c( expression(paste("Sample ",mu)), expression(paste("Thearotical ",mu))), fill=c("green","yellow"))

```



Sample Variance vs Theoretical Variance

So lets discover the variability of our sample, using R ,The standard diviation of the sample means calculated by `sd(mns)` function is :

```
## [1] 0.799569
```

accordingly , the variance is calculated sameway using `var(mns)` function is:

```
## [1] 0.6393107
```

Theoritically,the standard diviation $\frac{1}{\lambda \cdot \sqrt{n}} = \frac{1}{0.2 \cdot \sqrt{40}} =$

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

accordingly, the theoretical variance is

```
##Theoretical standard deviation
theory_var <- ((1/lambda)/sqrt(40))^2
print(theory_var)
```

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

Normal Distribution:

According to CLT (central limit theorem), the averages of samples follow normal distribution.

The figure above also shows the density computed using the histogram and the normal density plotted with theoretical mean and variance values. Also, the q-q plot below suggests the normality. The theoretical quantiles again match closely with the actual quantiles. These four methods of comparison prove that the distribution is approximately normal.

```
qqnorm(mns); qqline(mns)
```

