Statistical Inference Assignement

Bruno Berrehuel

November 13, 2016

Contents

Part I
Exponantial Distribution

1 Overview

In this part I'll do some investigations of the exponential distribution, with the R function rexp(n, λ) and more precisely the distribution of averages of 40 exponentials distributions. In accordance with the Central Limit Theorem, I'll investigate that the sample mean μ and the sample variance s^2 are :

$$\mu = \frac{1}{\lambda} \quad \text{and} \quad s^2 = \frac{\sigma^2}{n}$$

Then I'll look at the distribution of 10 000 means of 40 exponential distributions, and verify that they follow a normal distribution $\mathcal{N}(\mu, s^2)$, as expected with the CLT.

For numerical purposes, λ is choosen as $\lambda = 0.2$, so $\mu = \sigma = \frac{1}{\lambda} = 5$.

2 Simulations

Calculate 10 000 means of 40 exponantials distributions with $\lambda = \frac{1}{5}$: for each i, calculate the mean of 40 exponantials distributions, then add it to the variables rexpMean and rexpVar. The rexpMean and rexpVar variables will contain 10 000 values each.

```
set.seed(12345)
rexpMean=NULL
rexpVar=NULL
for (i in 1:10000) {
    rexpMean = c(rexpMean, mean(rexp(40,0.2)))
    rexpVar = c(rexpVar, var(rexp(40,0.2)))
}
```

3 Sample mean and variance vs. theoricals

Calculate the means of the mean and the variance of the sample with R, and compare with mean and variance expected. Recall that the expected sample mean μ and the expected sample variance s are :

$$\mu = \frac{1}{\lambda} = 5 \quad \text{and} \quad s^2 = \frac{\sigma^2}{n} = \frac{5^2}{40} \simeq 0.625$$

The expected distribution variance is $\sigma^2 = 25$.

```
data.frame(mean=mean(rexpMean), sampleVariance=var(rexpMean),
           variance=mean(rexpVar))

##       mean sampleVariance variance
## 1 5.003857      0.6083614 24.92046
```

The values are very close to the expected values :

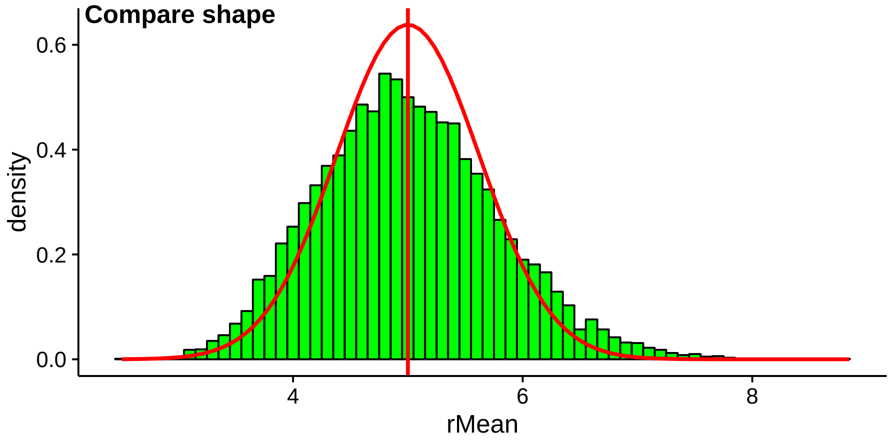
- 0.08% for the sample mean,
- 2.66% for the sample variance, and
- 0.32% for the distribution variance.

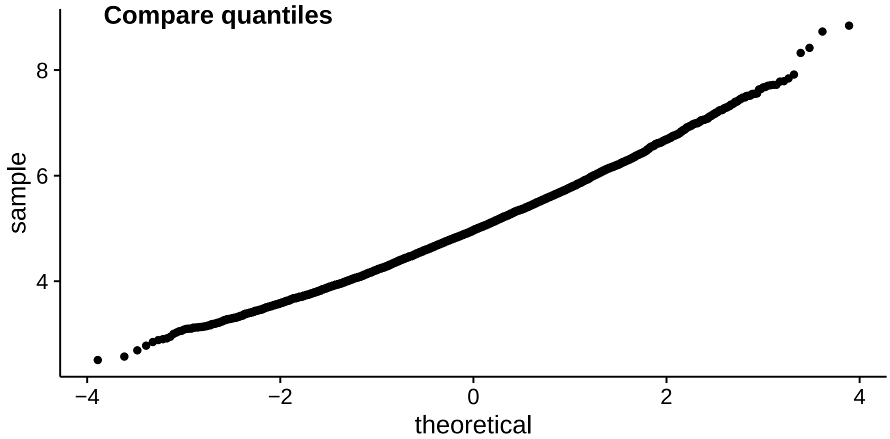
4 Distribution

Plot an histogram with the 10 000 previous calculation, and compare the shape with a normal distribution $\mathcal{N}(5, 0.625)$, then compare the distributions using a quantile-quantile diagram :

```
library(ggplot2)
library(cowplot)
dfRexp  <- data.frame(rMean=rexpMean, rVar=rexpVar)
histoMean <- ggplot(dfRexp, aes(x=rMean)) +
    geom_histogram(aes(y=..density..), color="black",
                   fill="green", binwidth=0.1) +
    geom_vline(aes(xintercept=5), color="red", size=1) +
    stat_function(fun=dnorm, color="red", size=1,
                  args=list(mean=5, sd=0.625))
qqMean <- ggplot(dfRexp, aes(sample=rMean)) + stat_qq()

plot_grid(histoMean, qqMean, ncol=1, nrow=2,
          labels=c("Compare shape","Compare quantiles"))
```





5 Conclusion

As points are aligned in the qqnorm diagram and with the shape of the histogram, the distribution of the mean follow the normal distribution $\mathcal{N}(5{,}0.625)$, as expected with the CLT.

Part II

ToothGrowth Data Analysis

1 Data exploratory

The datas are about the length of odontoblast (cells responsible for tooth growth) for 60 pigs after an experimental threatment in vitamin C. Each animal received a dose of vitamin C, from 0.5 to 2 mg/day, by orange juice (OJ) or ascorbic acid (VC)[1]. Each experiment concern 10 pigs.

First I'll take some informations about the means and the standard deviation of each dose with the dplyr package :

```
library(datasets)
library(dplyr)
ToothGrowth %>% group_by(supp,dose) %>% summarise(mean(len), round(sd(len),3))

## Source: local data frame [6 x 4]
## Groups: supp [?]
##
##     supp  dose `mean(len)` `round(sd(len), 3)`
##   <fctr> <dbl>       <dbl>               <dbl>
## 1     OJ   0.5       13.23               4.460
## 2     OJ   1.0       22.70               3.911
## 3     OJ   2.0       26.06               2.655
## 4     VC   0.5        7.98               2.747
## 5     VC   1.0       16.77               2.515
## 6     VC   2.0       26.14               4.798
```

2 Probability tests

3 Conclusions

[1] Source : https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html