Exponential Distribution and CLT

Part 1 Exponetial and CLT

Overview

In this project, I will investigate the exponential distribution in R and compare it with the Central Limit Theorem: the distribution of averages of 40 exponentials with a thousand simulations.

Show the sample mean and compare it to the theoretical mean of the distribution.

Theortical Mean

```
lambda <- 0.2
T_mean <- 1/lambda
print(T_mean)
```

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

Sample Mean

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

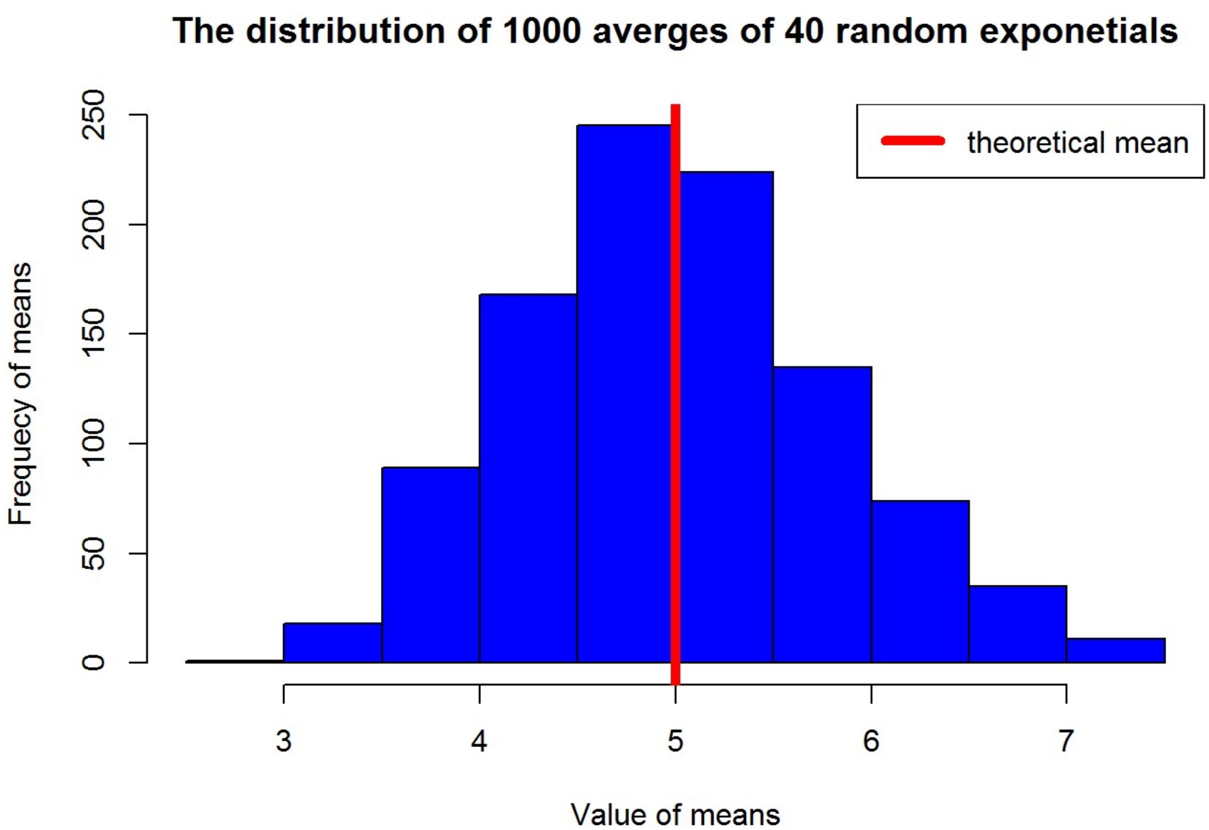
```
## [1] 5.00369
```

```
head(mns)
```

```
## [1] 5.777610 3.914035 4.163126 5.392891 5.437866 6.291835
```

compare distribution

```
hist(mns, main="The distribution of 1000 averges of 40 random exponetials", xla
b="Value of means", ylab="Frequecy of means", col="blue")
abline(v=T_mean, lty=1, lwd=5, col="red")
legend("topright", lty=1, lwd=5, col="red", legend="theoretical mean")
```



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

Theortical variance

```
T_variance <- (1/lambda)^2
print(T_variance)
```

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

Sample variance

```
vns = NULL
for (i in 1 : 1000) vns = c(vns, var(rexp(40,lambda)))
S_variance <- var(vns)
print(S_variance)
```

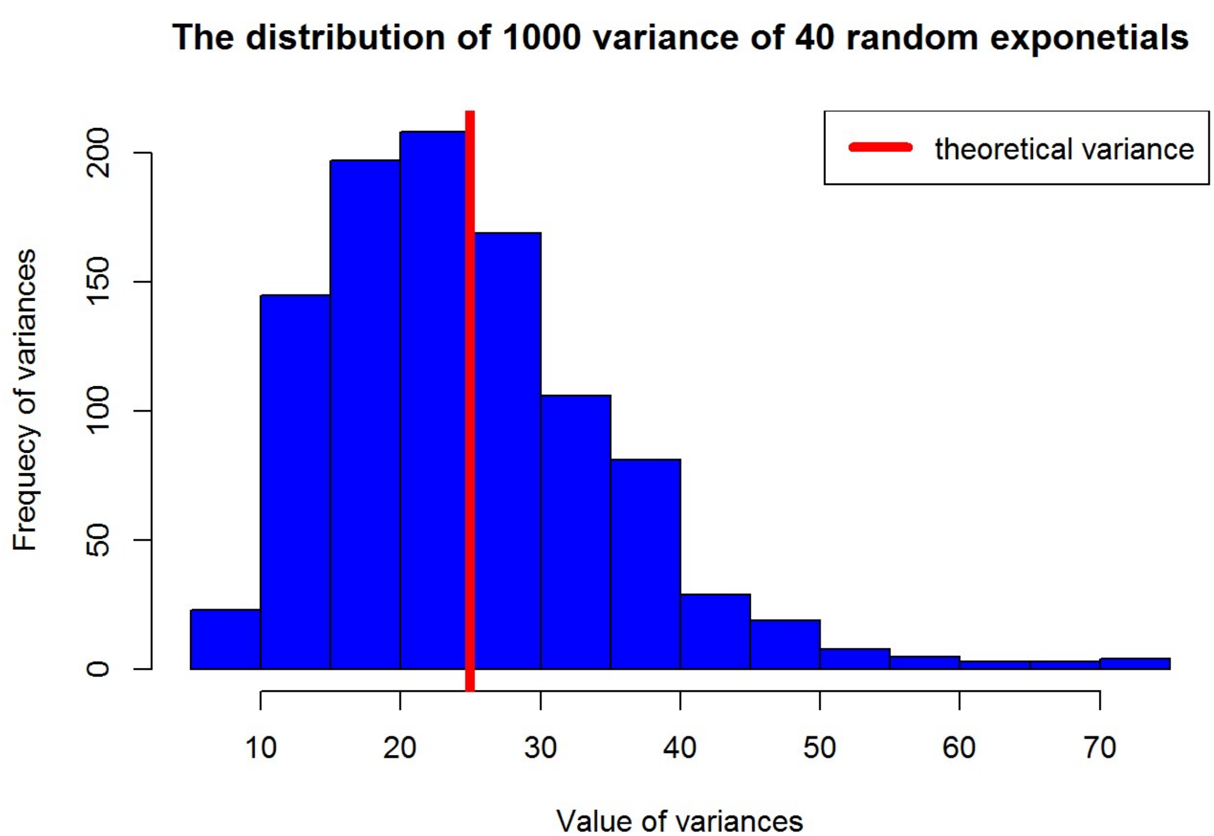
```
## [1] 108.3528
```

```
head(vns)
```

```
## [1] 27.11311 21.68268 28.17645 14.08804 22.24334 26.13239
```

compare distribution

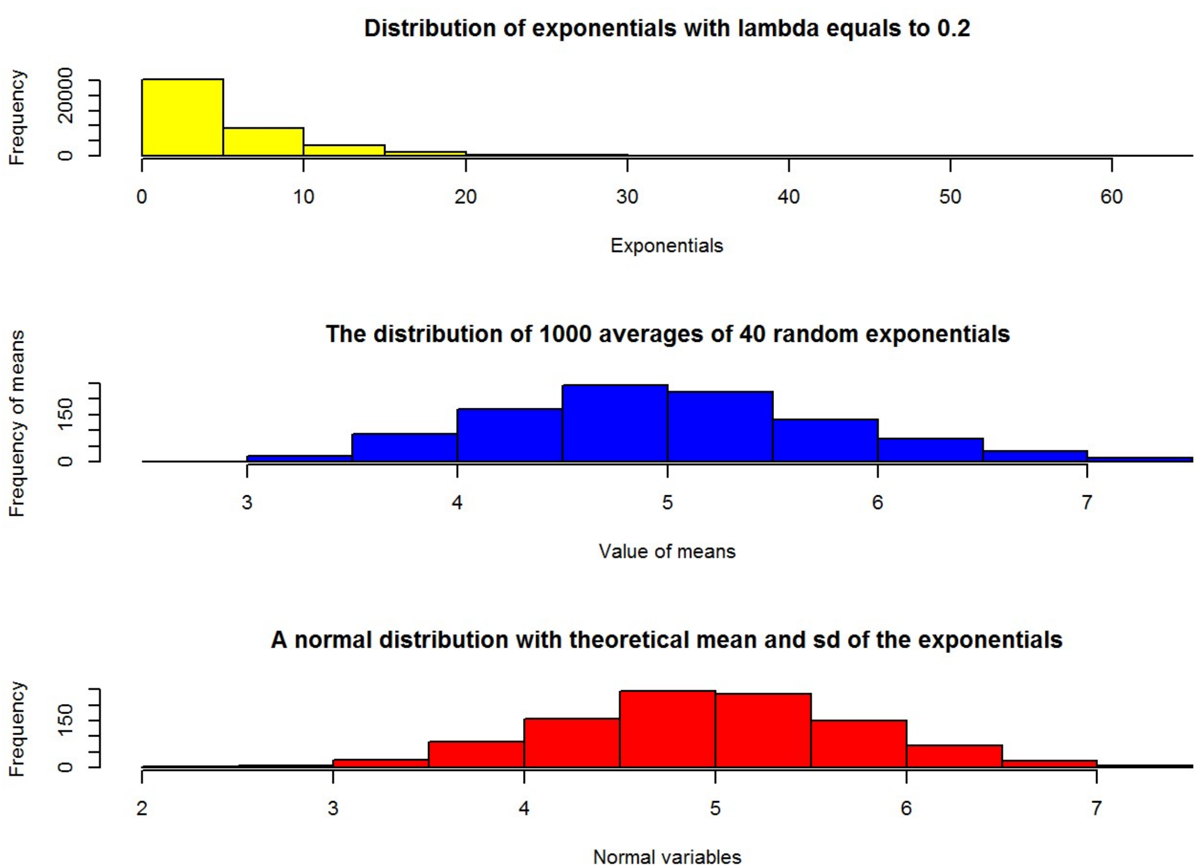
```
hist(vns, main="The distribution of 1000 variance of 40 random exponetials", xl
ab="Value of variances", ylab="Frequecy of variances", col="blue")
abline(v=T_variance, lty=1, lwd=5, col="red")
legend("topright", lty=1, lwd=5, col="red", legend="theoretical variance")
```



Show that the distribution is approximately normal:focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

```
ns = NULL
for (i in 1 : 1000) ns = c(ns, rexp(40,lambda))

par(mfrow = c(3, 1))
hist(ns, main = "Distribution of exponentials with lambda equals to 0.2", xlab
= "Exponentials", col = "yellow")
hist(mns, main = "The distribution of 1000 averages of 40 random exponential
s", xlab = "Value of means", ylab = "Frequency of means", col = "blue")
Norm <- rnorm(1000, mean = mean(mns), sd = sd(mns))
hist(Norm, main = "A normal distribution with theoretical mean and sd of the ex
ponentials", xlab = "Normal variables", col = "red")
```



summary: as a result of the central limit theorem, comparing the second and the third histogram, we can see the distribution of the means is similar to a real normal distribution with the same mean and standard deviation.

Part 2 ToothGrowth Analysis

Overview : analyze tooth growth data and have some conclusions

load data/libarary, basic summary and explore data

```
library(ggplot2)
data("ToothGrowth")
dim(ToothGrowth)
```

```
## [1] 60  3
```

```
head(ToothGrowth)
```

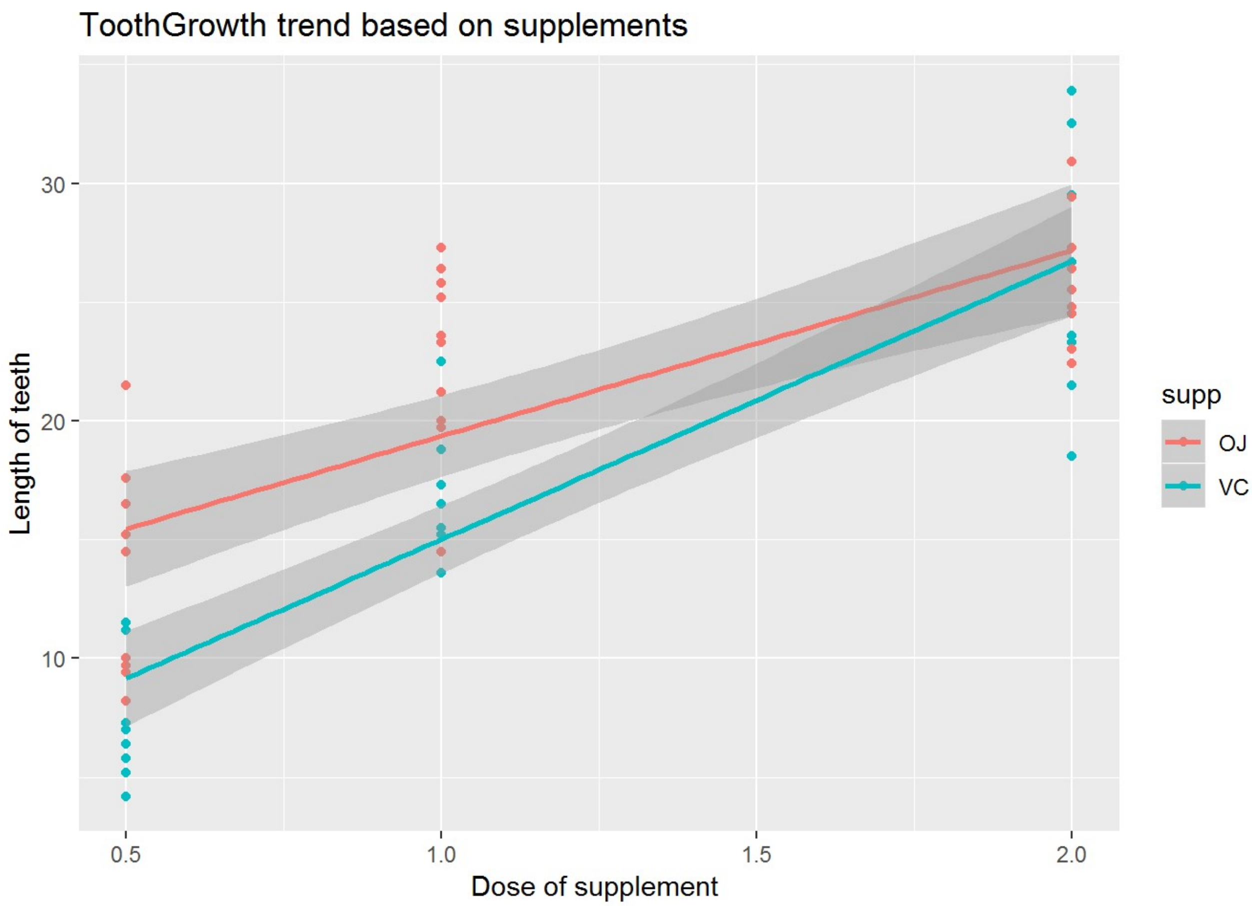
```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
summary(ToothGrowth)
```

```
##       len        supp         dose      
##  Min.   : 4.20   OJ:30   Min.   :0.500  
##  1st Qu.:13.07   VC:30   1st Qu.:0.500  
##  Median :19.25           Median :1.000  
##  Mean   :18.81           Mean   :1.167  
##  3rd Qu.:25.27           3rd Qu.:2.000  
##  Max.   :33.90           Max.   :2.000  
```

plot and see the trend

```
ggplot(ToothGrowth, aes(x = dose, y=len, color=supp)) +
  geom_point() + geom_smooth(method = "lm")+
  labs(title = "ToothGrowth trend based on supplements", x = "Dose of supplemen
t ", y = "Length of teeth")
```



initial conclusions:

- supplment and tooth growth are linear relationship
- same dose, OJ has bigger effect compared with VC
- VC has larger gradient of increase compared with OJ

Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose

assumption – normal distribution, null hypothesis – no difference between mean of 2 kinds of supplements or dose

```
t.test(x = ToothGrowth$len, data = ToothGrowth, paired = FALSE, conf.level = 0.
95)$conf.
```

```
## [1] 16.83731 20.78936
## attr(,"conf.level")
## [1] 0.95
```

calculate confidence level and p value based on kind of supplement

```
t.test(ToothGrowth[ToothGrowth$supp == "OJ", ]$len)$p.value
```

```
## [1] 1.037346e-16
```

```
t.test(ToothGrowth[ToothGrowth$supp == "OJ", ]$len)$conf.int
```

```
## [1] 18.19678 23.12989
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(ToothGrowth[ToothGrowth$supp == "VC", ]$len)$p.value
```

```
## [1] 4.362782e-12
```

```
t.test(ToothGrowth[ToothGrowth$supp == "VC", ]$len)$conf.int
```

```
## [1] 13.87675 20.04992
## attr(,"conf.level")
## [1] 0.95
```

calculate confidence level and p value based on dose

```
t.test(ToothGrowth[ToothGrowth$dose == 0.5, ]$len)$p.value
```

```
## [1] 2.241441e-09
```

```
t.test(ToothGrowth[ToothGrowth$dose == 0.5, ]$len)$conf.int
```

```
## [1]  8.499046 12.710954
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(ToothGrowth[ToothGrowth$dose == 1, ]$len)$p.value
```

```
## [1] 3.218468e-14
```

```
t.test(ToothGrowth[ToothGrowth$dose == 1, ]$len)$conf.int
```

```
## [1] 17.66851 21.80149
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(ToothGrowth[ToothGrowth$dose == 2, ]$len)$p.value
```

```
## [1] 1.026427e-17
```

```
t.test(ToothGrowth[ToothGrowth$dose == 2, ]$len)$conf.int
```

```
## [1] 24.33364 27.86636
## attr(,"conf.level")
## [1] 0.95
```

calculate mean based on dose for each supplement

```
dose05 <-ToothGrowth[ToothGrowth$dose == 0.5, ]
mean(dose05[dose05$supp=="OJ",]$len)
```

```
## [1] 13.23
```

```
mean(dose05[dose05$supp=="VC",]$len)
```

```
## [1] 7.98
```

```
dose1 <-ToothGrowth[ToothGrowth$dose == 1, ]
mean(dose1[dose1$supp=="OJ",]$len)
```

```
## [1] 22.7
```

```
mean(dose1[dose1$supp=="VC",]$len)
```

```
## [1] 16.77
```

```
dose2 <-ToothGrowth[ToothGrowth$dose == 2, ]
mean(dose2[dose1$supp=="OJ",]$len)
```

```
## [1] 26.06
```

```
mean(dose2[dose1$supp=="VC",]$len)
```

```
## [1] 26.14
```

OJ delivers more tooth growht than VC for dose 0.5 & 1. OJ and VC deliver same tooth growth for dose 2.