Statistical Interference Assignment Coursera

Michaela Spiegel

21 August 2018

Part 1: Exponential Distribution & Central Limit Theorem

Overview

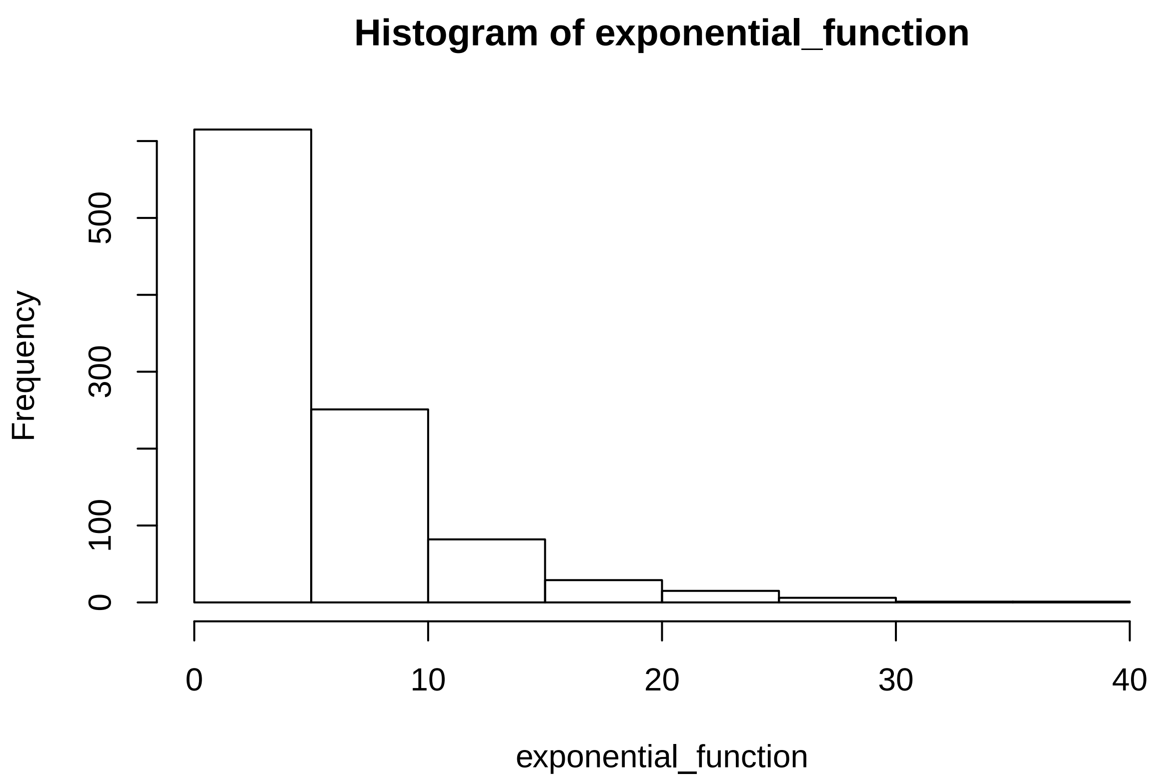
This part is about the difference between the distribution of large collection of random exponentials vs. the distribution of large collections of averages of random exponentials. In this simulation a large collection contains 1,000 values and the averages of exponentials are calculated as average over 40 exponentials.

The relation between these two distributions is given by the central limit theorem, which states when independent random variables (in some situations) are added, their properly normalized sum tends toward a normal distribution even if the original variables themselves are not normally distributed.

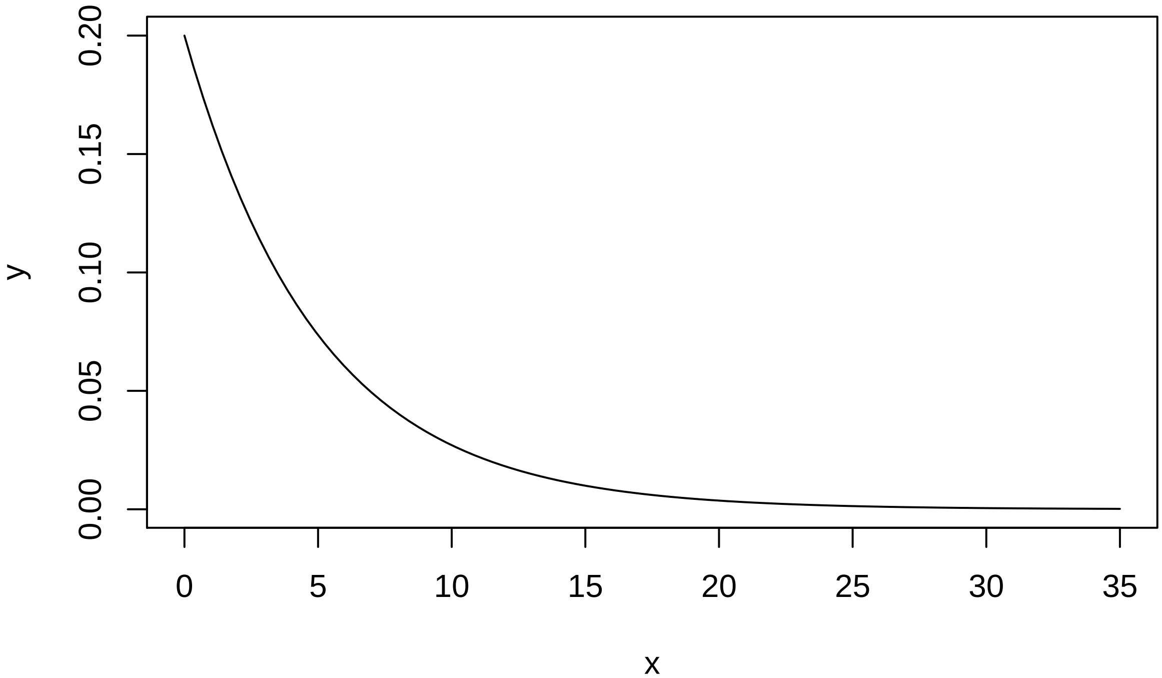
Exponential distribution

The exponential function $\lambda * \exp(-\lambda x)$ in R is given by "rexp()" function. If we pick randomly from this distribution (1,000 times), we get a sample distribution which looks pretty much like the original exponential distribution.

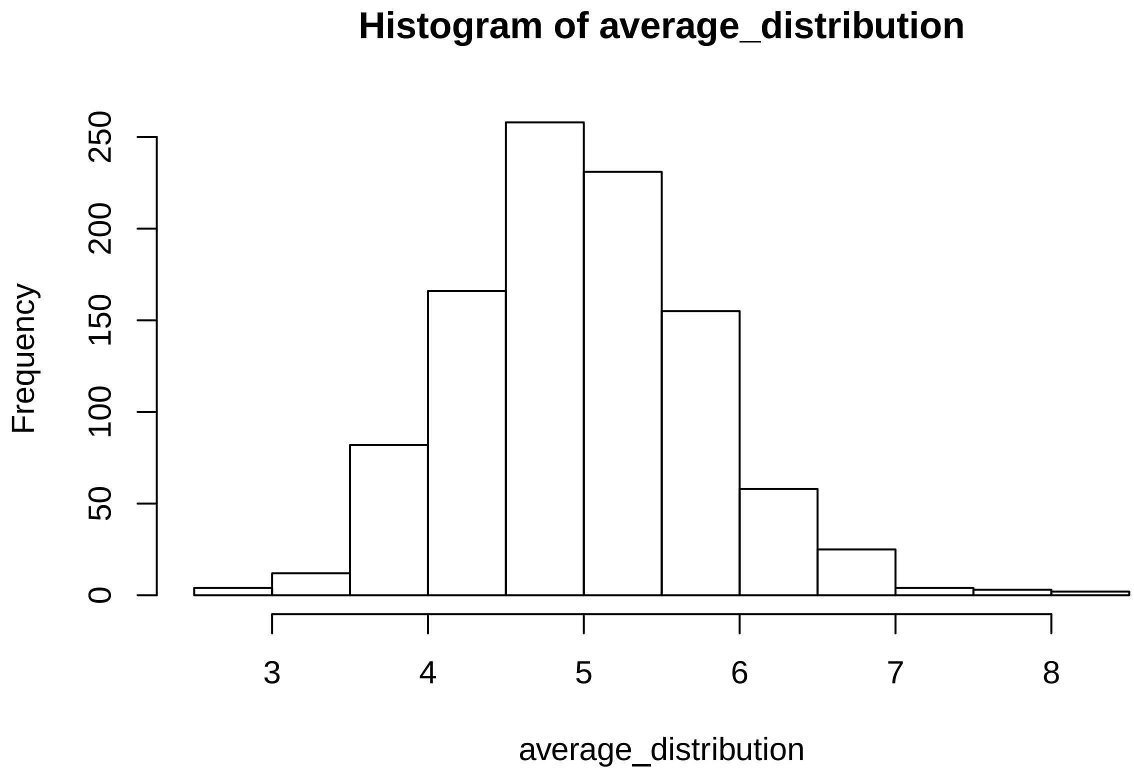
```
lambda = 0.2
n=1000
exponential_function <- rexp(n, lambda)
hist(exponential_function)
```



```r
eq = function(x){lambda*exp(-lambda*x)}
curve(eq, from=0, to=35, xlab="x", ylab="y")
```




```r
#set seed(123456)
lambda = 0.2
n=40
average_distribution = c()
for (i in 1:1000)
  average_distribution[i] <- mean(rexp(n, lambda))
hist(average_distribution)
```



approximately normal distribution ... t distribution?

The mean and standard deviation of the normal distribution are $1/\lambda$, with $\lambda = 0.2$ follows $mean = \sigma = \frac{1}{\lambda} = \frac{1}{0.2} = 5$.

Aufbau

Title (give an appropriate title) and Author Name

Overview: In a few (2-3) sentences explain what is going to be reported on.

Simulations: Include English explanations of the simulations you ran, with the accompanying R code. Your explanations should make clear what the R code accomplishes.

Sample Mean versus Theoretical Mean: Include figures with titles. In the figures, highlight the means you are comparing. Include text that explains the figures and what is shown on them, and provides appropriate numbers.

Sample Variance versus Theoretical Variance: Include figures (output from R) with titles. Highlight the variances you are comparing. Include text that explains your understanding of the differences of the variances.

Distribution: Via figures and text, explain how one can tell the distribution is approximately normal.

Part 2: Tooth Growth Data Set

Description

The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice or ascorbic acid (a form of vitamin C and coded as VC).

[,1] len numeric Tooth length [,2] supp factor Supplement type (VC or OJ). [,3] dose numeric Dose in milligrams/day

Source

C. I. Bliss (1952) The Statistics of Bioassay. Academic Press.

exploratory data analyses

```
library(datasets)
data(ToothGrowth)
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
```

```
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

Provide a basic summary of the data.

Hypothesis Testing

Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering) State your conclusions and the assumptions needed for your conclusions.