Course Notes for FSML Part II*

Tobías Chavarría

DSTI | DSBD2-001

Contents

*Replication files are available on the author's Github account (https://github.com/tobias-chc/FSML_II).

Introduction

Statistics notation:

1. If $X_1, ..., X_n$ are random variables (r.v).
2. $x_1, ..., x_n$ are observations.
3. If we write $i.i.d$ means that the r.v are independent and identically distributed.

First aim: To propose a model for a random variable.

Generalization to multi-dimensional case:

- Y: response variable.
- $X^{(1)}, ..., X^{(p)}$: explanatory variables.

Aim: To find a functional link between Y and the explanatory variables.

To find this functional link , the method to apply depends on the nature of the r.v's.

Y	Model
Numeric	Linear model
Qualitative (labels)	Classification

Linear model

A linear model is given by:

$Y_i = \beta_0 + \beta_1 X_i^1 + ... + \beta_p X_i^p + \varepsilon_i$

where:

- $\beta_0, ..., \beta_p$ are unknown *fixed* parameters that can be estimated by two methods:
- Point estimation
- Confidence interval
- ε is the noise and also a random variable.

Chapter 1: Estimation for one parameter

Previous Knowledge

- Random Variable:
- The notion of distribution.
- The expectation and variance
- The distribution function
- The classical distributions (in particular the Gaussian)
- The Law of Large numbers and the Central Limit theorem

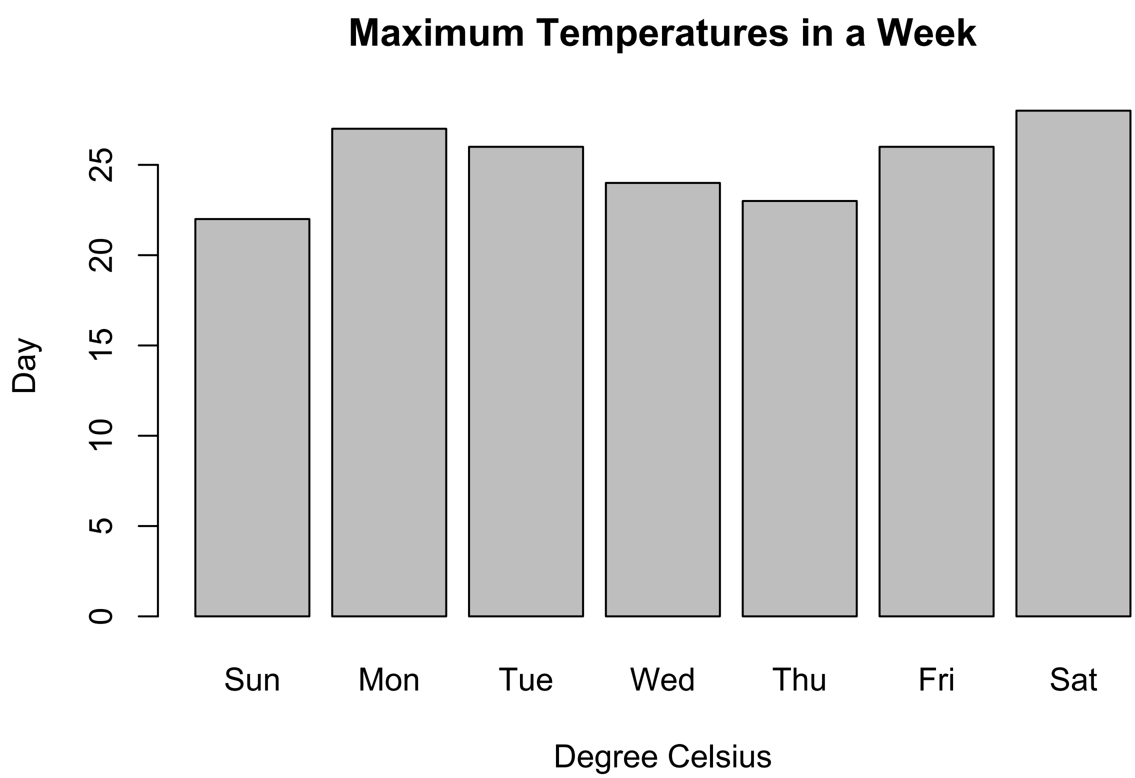
Introduction

Given $x_1, ..., x_n$ numeric observations, to try to find a correct parametric model, we can use 2 graphs:

Plot type	Variable type	Density
Bar plot	Discrete	count/n
Histogram	Continuous	count for a bin/ n x (length of the bin)

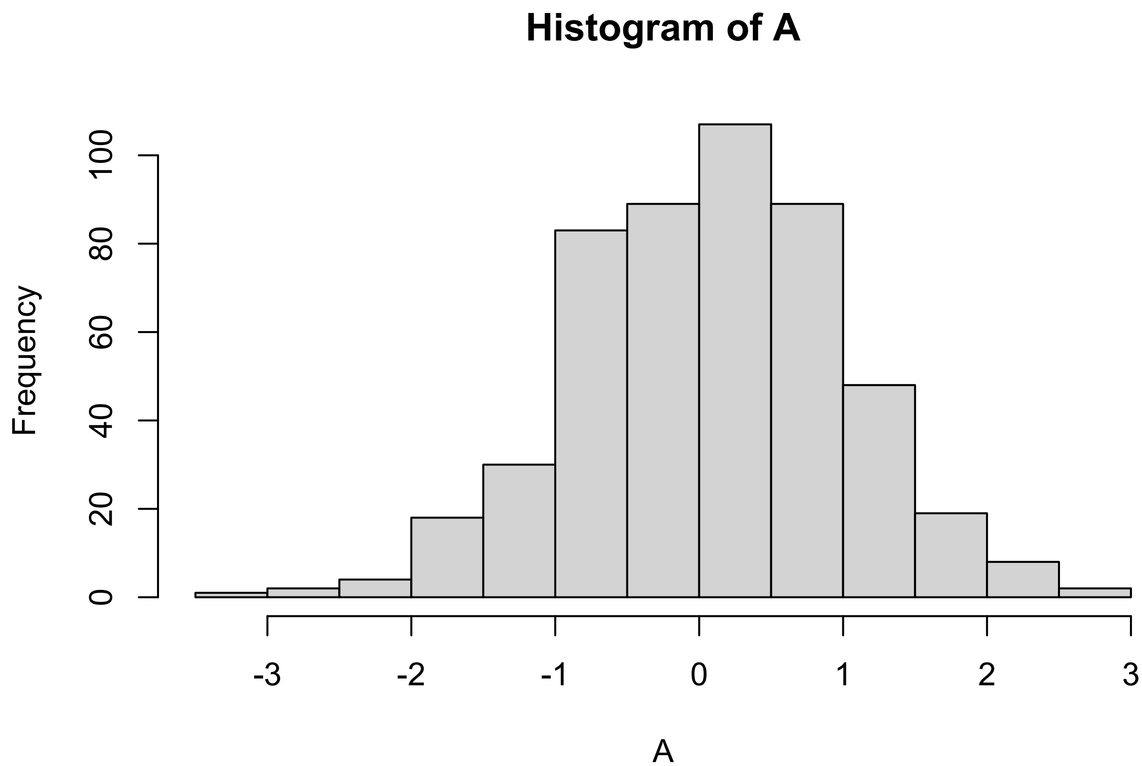
Barplot for discrete variables

```
max.temp <- c(22, 27, 26, 24, 23, 26, 28)

barplot(max.temp,
        main = "Maximum Temperatures in a Week",
        xlab = "Degree Celsius",
        ylab = "Day",
        names.arg = c("Sun", "Mon", "Tue", "Wed", "Thu", "Fri", "Sat"))
```

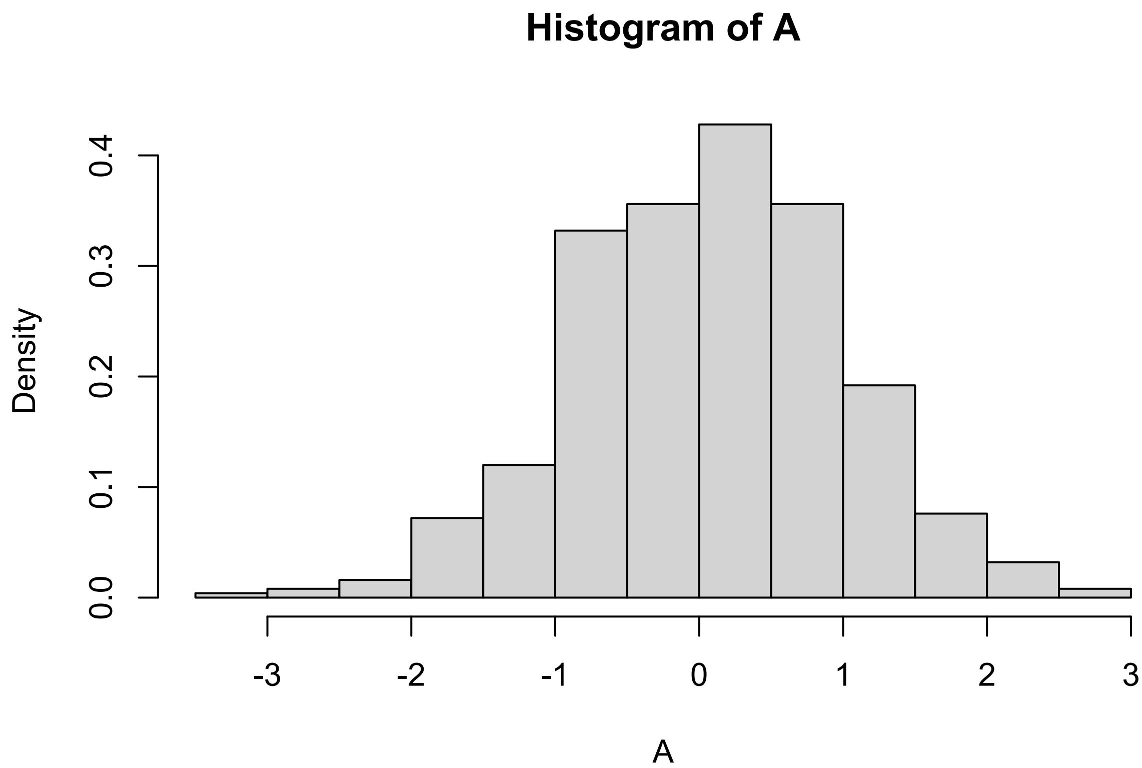


Histogram for continuous variables

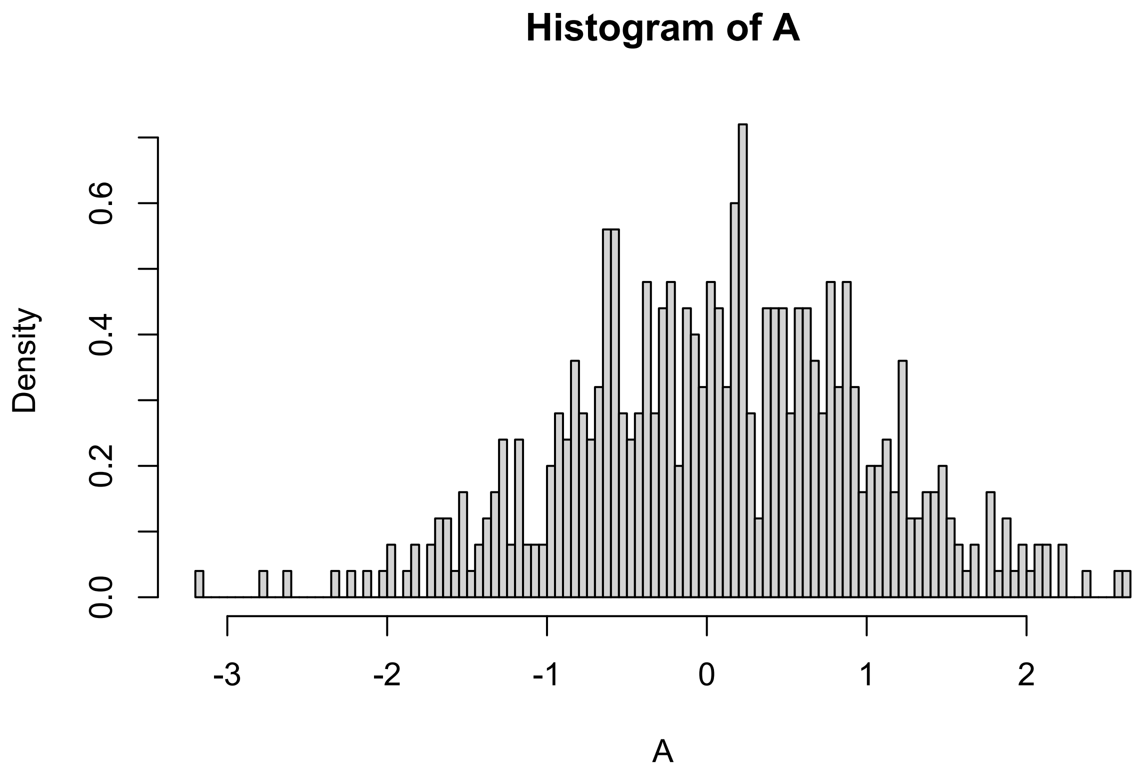
```
A <- rnorm(500, 0, 1)

## The default execution of this function doesn't generate a density:
hist(A)
```



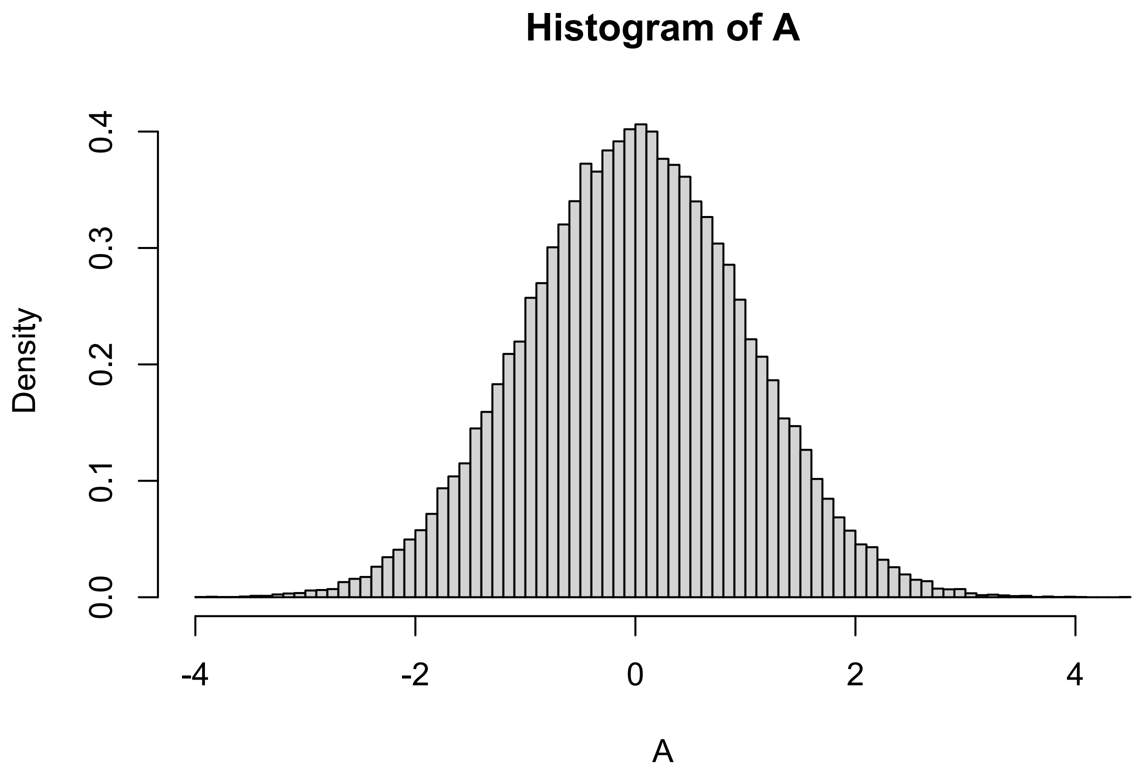

```r
## You need to set freq = FALSE:
hist(A, freq = FALSE)
```




```
## You can set the numbers of bins that you want to use:
hist(A, freq = FALSE, breaks = 100)
```




```r
# But in order to create more breaks, you need to increase the
# numbers of observations:


A <- rnorm(50000, 0, 1)
hist(A, freq = FALSE, breaks = 100)
```



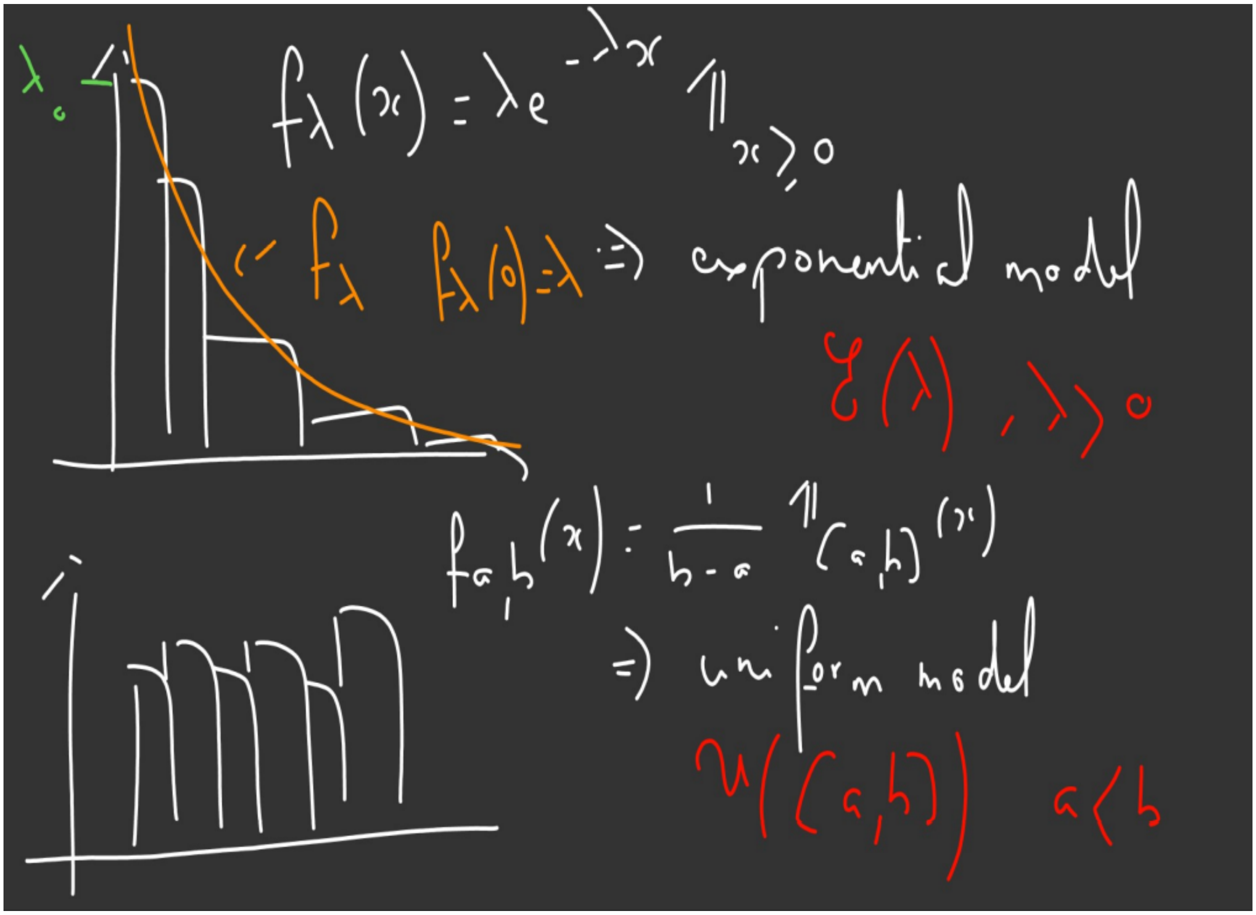
To propose a parametric model:

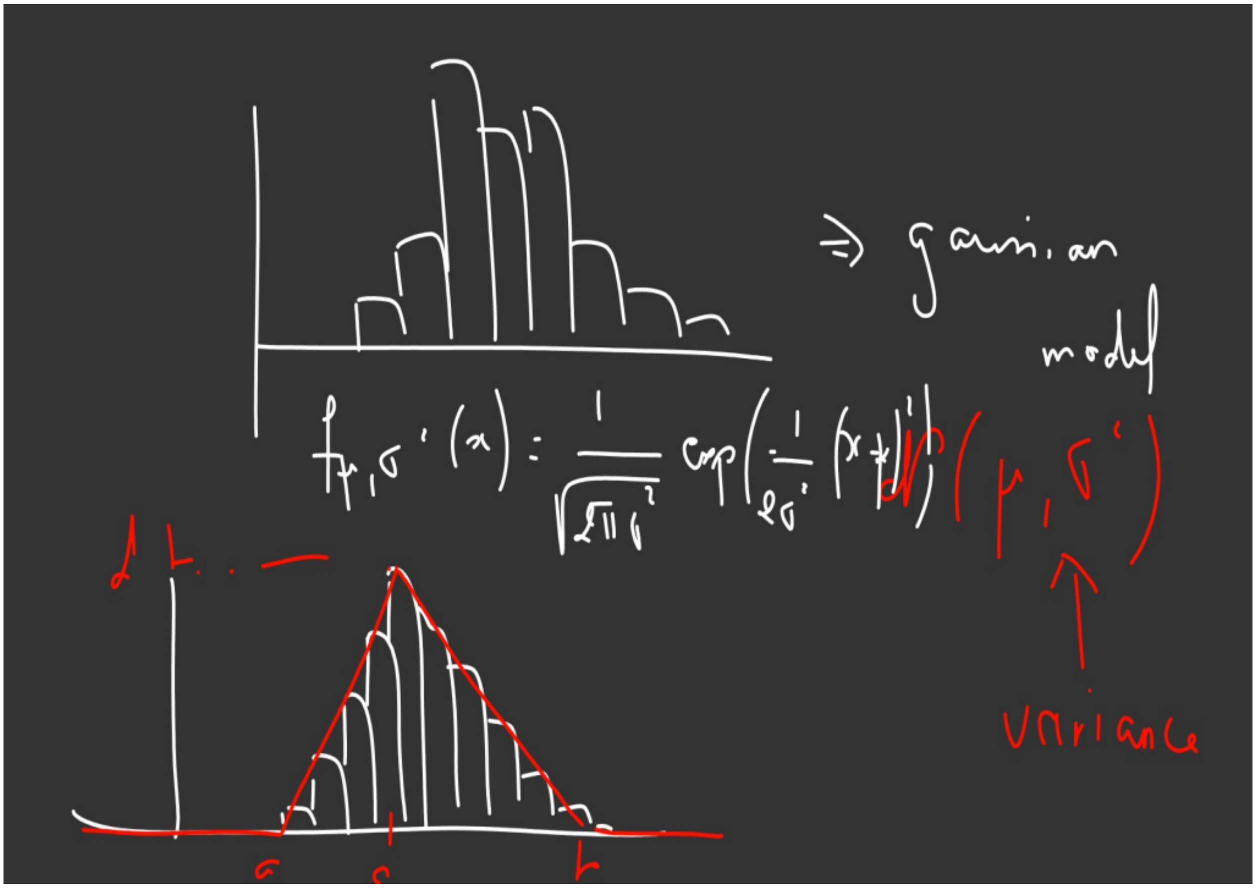
1. Make a graphical representation of the observations.
2. Guess a theoretical model by looking the previous graphic.

Examples:

Pasarlos a R!!

Example 1:


$f_\lambda(x) = \lambda e^{-\lambda x} \mathbb{1}_{x \geq 0}$
λ
$\leftarrow f_\lambda$ $f_\lambda(0) = \lambda$ $\Rightarrow$ exponential model
$\mathcal{E}(\lambda)$, $\lambda > 0$
$f_{a,b}(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x)$
$\Rightarrow$ uniform model
$\mathcal{U}([a,b])$ $a < b$

Example 2: 

Question: with a representation, we can guess a parametric family of models, denoted by $\{P_\Theta, \theta \in \Theta\}$. How to guess a correct value for θ thanks to the observations?

Answer: *Estimation.*

Point estimation

Let x_i an observation of a r.v X_i we assume that $X_1, ..., X_n$ are *i.i.d* with common distribution P_θ.

Estimator **Definition:** An estimator of Θ is just a function of $X_1, ...X_n$ that **does not depend onto others unknown parameters.**

Remark: An estimator is a random variable!

Estimation **Definition:** An estimation is the value of an estimator computed thanks to the observations.

Example

Consider $X_1, ..., X_n$ exponential distributed and *i.i.d*, an *estimator* of λ is $\hat{\lambda}_n = \dfrac{n}{\sum X_i}$ an *estimation* is $\hat{\lambda}_n = \dfrac{n}{\sum x_i}$.

Distribution	Parameter	Estimator	Estimation
Exponential $\xi(\lambda)$	λ	$\frac{n}{\sum X_i}$	$\frac{n}{\sum x_i}$

Bias (for univariate parameter) **Definition:** Let consider $\hat{\theta}_n$ an estimator of θ.

The bias of $\hat{\theta}_n$ is defined by:

$$b(\hat{\theta}_n) := \mathbb{E}(\hat{\theta}_n) - \theta$$

- We say that $\hat{\theta}_n$ is an unbiased estimator if

$$\forall n \in \mathbb{N}^+ \quad b(\hat{\theta}_n) = 0$$

- We say that $\hat{\theta}_n$ is *asymptotic unbiased* estimator if:

$$b(\hat{\theta}_n) \to 0 n \to +\infty$$

How to construct estimator?

- Method of moments
- less computations
- based on the Law of large numbers
- Maximum likelihood

Method of moments Let θ a parameter to estimated, parameter which is associate to $X_1, ..., X_n$ *i.i.d* r.v.

Let consider $k \in \mathbb{N}^*$:

- the moment of order k : $\mathbb{E}[X^k]$
- the centered moment of order k: $\mathbb{E}[X - \mathbb{E}[X]]^k$

If there exist a value k such that:

- (a) $\mathbb{E}[X^k] = g(\theta)$
- (b) $\mathbb{E}[[X - \mathbb{E}[X]]^k] = h(\theta)$

Then the estimator $\hat{\theta}_n$ of θ is solution of:

- (a) $g(\hat{\theta}_n) = \frac{1}{n} \sum X_i^k$
- (b) $h(\hat{\theta}_n) = \frac{1}{n} \sum (X_i - \overline{X}_n)^k$

where

$$\overline{X}_n = \frac{1}{n} \sum X_i$$

is the **empirical mean**.

Remark: Exponential Distribution

If Let $X \sim \xi(\lambda)$, then:

- $f_\lambda(x) = \lambda \exp(-\lambda x) \mathbf{1}_{x \geq 0}$
- $\mathbb{E}[X] = \int_{\mathbb{R}} x \cdot f_\lambda(x) dx$
- $V[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

Transfer formula:

$$\mathbb{E}[l(X)] = \int_{\mathbb{R}} l(x) \cdot f_\lambda(x) dx$$

Applications

1. Let consider $X_1, ..., X_n$ exponential distributed and $i.i.d$ compute an estimator of λ using the methods of moments.

Solution:

Let $X \sim \xi(\lambda)$, so:

$$\mathbb{E}[X] = \frac{1}{\lambda} \quad \text{and} \quad V[X] = \frac{1}{\lambda^2}$$

By applying the method of moments ($k = 1$) we get:

$\frac{1}{\hat{\lambda}_{n,1}} = \frac{1}{n} \sum X_i$

Thus:

$$\hat{\lambda}_{n,1} = \frac{n}{\sum X_i}$$

in the same way but using the variance ($k = 2$), we get:

$\frac{1}{\hat{\lambda}^2_{n,2}} = \frac{1}{n} \sum (X_i - \overline{X}_n)^2$

Thus:

$$\hat{\lambda}_{n,2} = \frac{\sqrt{n}}{\sqrt{\sum (X_i - \overline{X}_n)^2}}$$

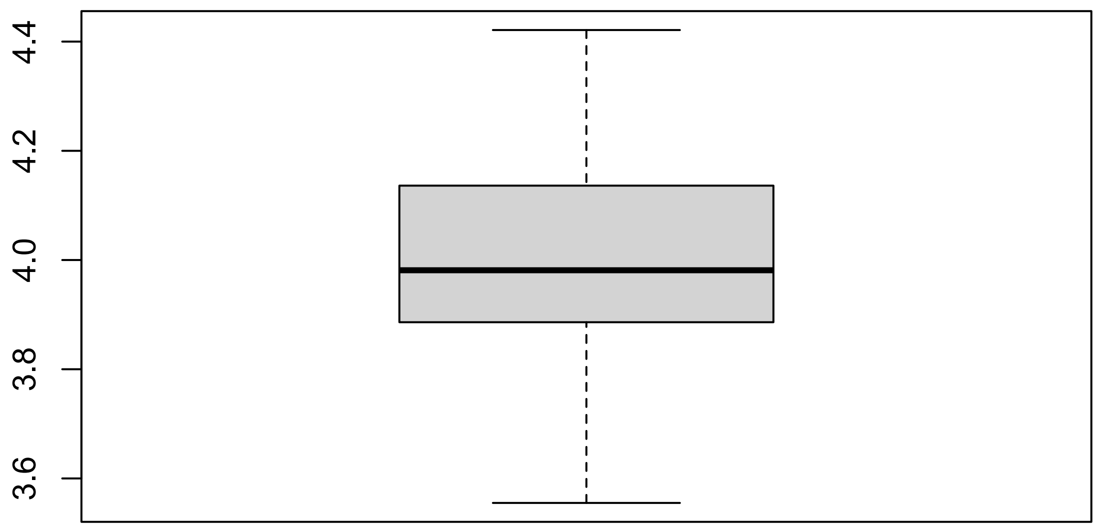
```
A = rexp(500, 4)

1/mean(A)
```

```
## [1] 3.841935
```

```r
m = c()

for (i in 1:50) {
  A = rexp(500, 4)
  m[i] <- 1/mean(A)
}


mean(m)
```

```
## [1] 3.995975
```

```r
boxplot(m)
```

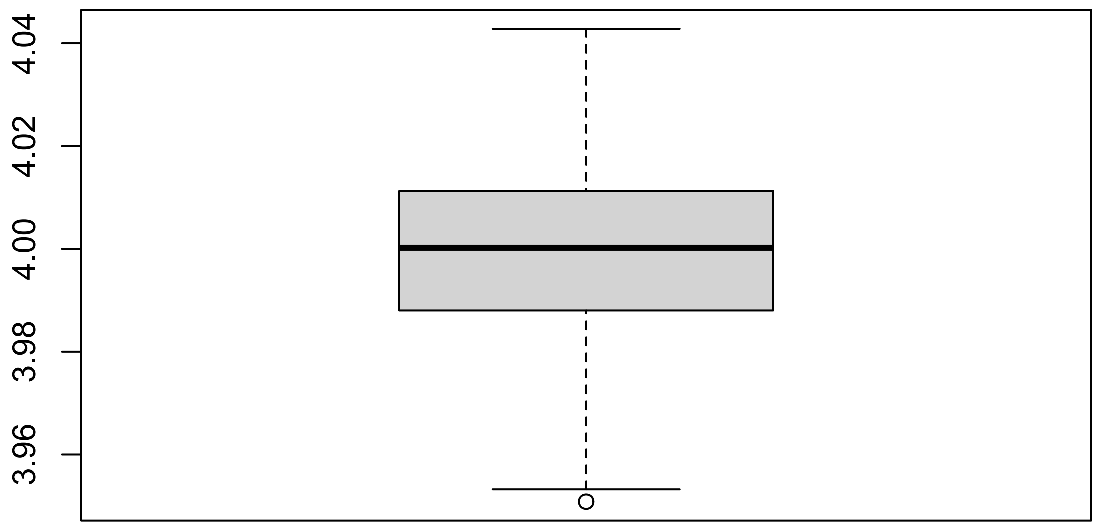



```r
## With more observations we got less variation (500 -> 5000)

## Law of large numbers

m = c()

for (i in 1:50) {
  A = rexp(50000, 4)
```

```r
  m[i] <- 1/mean(A)
}


mean(m)
```

```
## [1] 3.997964
```

```r
boxplot(m)
```



2. Let consider $X_1, ..., X_n$ $i.i.d$ $\mathcal{U}([0,\sigma])$.

- Determine an estimator of σ using the methods of moments.
- Let denote $\hat{\sigma}_n = 2\overline{X}_n$. Is $\hat{\sigma}_n$ an unbiased estimator?

Solution:

Let $X \sim \mathcal{U}([0,\sigma])$.

$$\mathbb{E}(X) = \int_{\mathbb{R}} x \cdot f_X dx = \int_{\mathbb{R}} x \cdot \frac{1}{\sigma} \cdot \mathbf{1}_{[0,\sigma]} dx = \frac{1}{\sigma} \int_0^\sigma x dx = \frac{1}{\sigma} \frac{x^2}{2} \Big|_0^\sigma = \frac{\sigma}{2}$$

By the method of moments, we get that an estimator $\hat{\sigma}_n$ is solution of:

$\frac{\hat{\sigma}_n}{2} = \frac{1}{n}\sum X_i = \overline{X}_n$

Thus:

$$\hat{\sigma}_n = 2\overline{X}_n$$

since $\hat{\sigma}_n$ it is a function of $X_1, ..., X_n$ it's an estimator.

Let's compute the bias for this estimator, let $n \in \mathbb{N}^*$.

$b(\hat{\sigma}_n) = \mathbb{E}(\hat{\sigma}_n) - \sigma$

First, compute the expected value of $\hat{\sigma}_n$:

$$\begin{aligned}
\mathbb{E}(\hat{\sigma}_n) &= \mathbb{E}(2\overline{X}_n) \\
&= \mathbb{E}\left[\frac{2}{n}\sum X_i\right] \\
&= \frac{2}{n}\sum \mathbb{E}[X_i] \quad \text{(by linearity)} \\
&= \frac{2}{n} \cdot n \cdot \mathbb{E}[X_1] \quad \text{(because are identically distributed)} \\
&= \sigma
\end{aligned}$$

Therefore $b(\hat{\sigma}_n) = 0$ and then we conclude that $\hat{\sigma}_n = \overline{X}_n$ is an unbiased estimator for σ.

3. Let consider $X_1, ..., X_n$ $i.i.d$ $\mathcal{U}([-\sigma, \sigma])$.

Since $\mathbb{E}[X_1] = 0 \quad$ (not a function of σ)

we can't use the first moment, then we use the second moment:

$V[X_1] = \frac{\sigma^2}{3}$

then we have:

$$\hat{\sigma}_n = \sqrt{\frac{3}{n}\sum(X_i - \overline{X}_n)^2}$$

The Maximum Likelihood

Likelihood Definition: Let $X_1, ..., X_n$ independent random variables, whose distributions are all depending on the same parameter θ.

Let $x_1, ..., x_n$ observations of those r.v

$$\mathcal{L}(x_1, ..., x_n, \theta) = \begin{cases} \prod_{i=1}^{n} P_\theta(X_i = x_i) & \text{(discrete framework)} \\ \prod_{i=1}^{n} f_{X_i,\theta}(X_i = x_i) & \text{(continuous framework)} \end{cases}$$

where $f_{X_i,\theta}$ are the density function of X_i.

Estimator thanks to the maximum likelihood **Definition:** $\hat{\theta}_n$, an estimator for θ, due to the maximum likelihood, is solution of:

$$\mathcal{L}(x_1, ..., x_n, \theta) = \max_{\theta} \mathcal{L}(x_1, ..., x_n, \theta)$$

Applications

1. Let consider $X_1, ..., X_n$ $\xi(\lambda)$ *i.i.d.* Compute the maximum likelihood estimator.

Solution:

Let $x_1, ..., x_n \in \mathbb{R}$.

$$\begin{aligned}
\mathcal{L}(x_1, ..., x_n, \theta) &= \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \cdot \mathbf{1}_{x_i \geq 0} \\
&= \lambda^n e^{-\lambda \sum x_i} \cdot \mathbf{1}_{\min{(x_i)} \geq 0}
\end{aligned}$$

we need to maximize with respect to λ, since $\mathbf{1}_{\min(x_i)}$ does not depend on λ we can forget it, and consider:

$$h(\lambda) = \lambda^n e^{-\lambda \sum x_i}$$

taking log in both sides, with get a better expression to deal with.

$$g(\lambda) = \log(h(\lambda)) = \log(\lambda^n e^{-\lambda \sum x_i}) = n \log(\lambda) - \lambda \sum x_i$$

Since

$g'(\lambda) = \dfrac{n}{\lambda} - \sum x_i$ then $\lambda = \dfrac{n}{\sum x_i} = \dfrac{1}{\overline{X}_n}$ is critical point.

and $g''(\lambda) = \dfrac{-n}{\lambda^2} < 0$ then this critical point correspond to a maximum.

So, $\lambda = \dfrac{1}{\overline{X}_n}$ is solution of the maximization problem and therefore the *Maximum likelihood estimator* is:

$$\hat{\lambda}_n = \frac{n}{\sum X_i}$$

2. Let consider $X_1, ..., X_n$ *i.i.d* $\mathcal{U}([0, \theta])$

- Compute the maximum likelihood estimator.
- Compute the bias of this estimator and transform it into an unbiased one.

Solution:

Compute the maximum likelihood estimator:

$$\begin{aligned}
\mathcal{L}(x_1, ..., x_n, \theta) &= \prod_{i=1}^{n} f_\theta(x_i) \\
&= \prod_{i=1}^{n} \frac{1}{\theta} \cdot \mathbf{1}_{[0,\theta]}(x_i) \\
&= \frac{1}{\theta^n} \cdot \mathbf{1}_{\min(x_i) \geq 0} \cdot \mathbf{1}_{\max(x_i) \leq \theta}
\end{aligned}$$

Since $\mathbf{1}_{\min(x_i)\geq 0}$ does not depend on λ we can forget it, and consider:

$$g(\lambda) = \frac{1}{\theta^n} \cdot \mathbf{1}_{\max(x_i)\leq\theta} = \frac{1}{\theta^n} \cdot \mathbf{1}_{[\max(x_i),+\infty[}(\theta)$$

we can't compute g' since this function it's not derivable in all points.

But since the function $\frac{1}{t^n}$ it's decreasing we can conclude that the maximum estimator likelihood is given by:

$$\hat{\theta}_n = \max(X_i)$$

Compute the bias of this estimator and transform it into an unbiased one.

In order to compute the bias of this estimator, we have to compute: $\mathbb{E}(\hat{\theta}_n) = \mathbb{E}(\max(X_i))$.

To do this we need to determine the density function of this new random variable $\max(X_i)$.

How to compute a density?

1. First step: Computation of the distribution function.

Let $t \in \mathbb{R}$.

$$\begin{aligned} F_{\hat{\theta}_n}(t) &= P(\hat{\theta}_n \leq t) \\ &= P(\max(X_i) \leq t) \\ \iff & \\ & P(X_1 \leq t, ..., X_n \leq t) \\ &= \prod_{i=1}^{n} P(X_i \leq t) \quad \text{(by independency of the r.v)} \\ &= (P(X_1 \leq t))^n \quad \text{(because they are identically distributed.)} \end{aligned}$$

Since

$$P(X_1 \leq t) = \int_{-\infty}^{t} \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x)dx = \begin{cases} 0 & t < 0 \\ \frac{t}{\theta} & t \in [0,\theta] \\ 1 & t > 1 \end{cases}$$

We conclude:

$$F_{\hat{\theta}_n}(t) = \begin{cases} 0 & t < 0 \\ \frac{t^n}{\theta^n} & t \in [0,\theta] \\ 1 & t > 1 \end{cases}$$

2. Second step: Computation of the density function.

Using the relation between the distribution and density function we have:

$$f_{\hat{\theta}_n}(t) = F'_{\hat{\theta}_n}(t) = n \cdot \frac{t^{n-1}}{\theta^n} \cdot \mathbf{1}_{[0,\theta]}(t)$$

3. Third step: Computation of $\mathbb{E}(\hat{\theta}_n)$

$$\mathbb{E}[\hat{\theta}_n] = \int_{\mathbb{R}} x \cdot f_{\hat{\theta}_n}(x)dx = \frac{n}{n+1}\theta$$

Now we have:

$$b(\hat{\theta}_n) = \mathbb{E}[\hat{\theta}_n] - \theta = -\frac{\theta}{n+1} \neq 0$$

Therefore $\hat{\theta}_n$ is not an unbiased estimator for θ. But since $\lim_{n\to\infty} b(\hat{\theta}_n) = 0$ it is an asymptotically unbiased estimator for θ.

How to obtain an unbiased estimator?

- First attempt:

We have $\mathbb{E}[\hat{\theta}_n] - \theta = -\frac{\theta}{n+1}$, then:

$$\begin{gathered} \mathbb{E}[\hat{\theta}_n] - \theta + \frac{\theta}{n+1} = 0 \\ \Longleftrightarrow \\ \mathbb{E}\left[\hat{\theta}_n + \frac{\theta}{n+1}\right] - \theta = 0 \end{gathered}$$

It is ok to consider $\tilde{\theta}_n = \hat{\theta}_n + \frac{\theta}{n+1}$ as unbiased estimator of θ?

The answer is **no**, $\tilde{\theta}_n$ depends on an unkonown parameter (θ in this case), so by definition it is not an estimator!

- Second attempt:

We have:

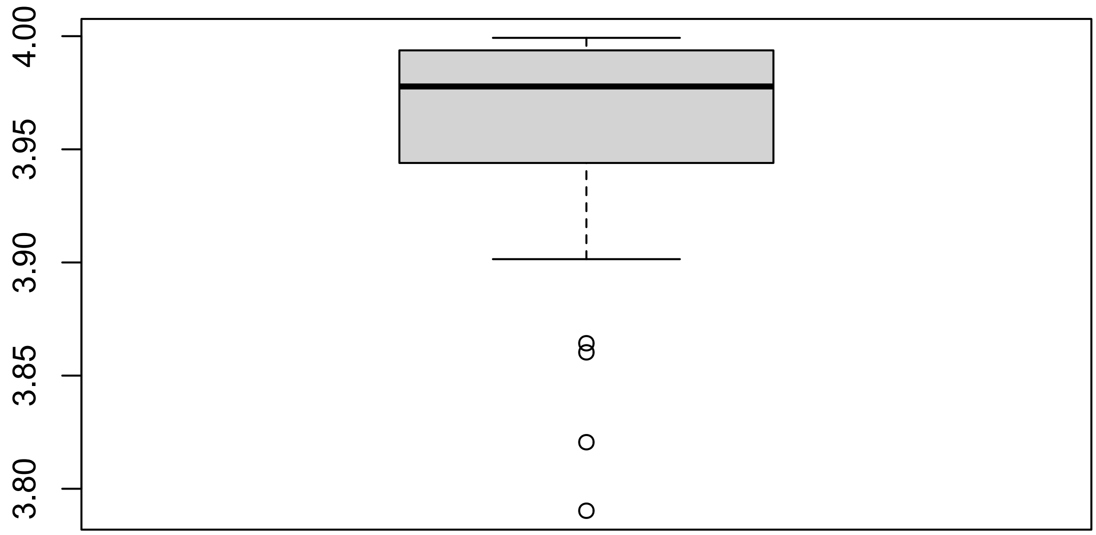
$$
\begin{gathered}
\mathbb{E}[\hat{\theta}_n] - \theta = -\frac{\theta}{n+1} = 0 \\
\Longleftrightarrow \\
\mathbb{E}[\hat{\theta}_n] = \theta - \frac{\theta}{n+1} \\
\Longleftrightarrow \\
\mathbb{E}[\hat{\theta}_n] = \frac{n}{n+1}\theta \\
\Longleftrightarrow \\
\mathbb{E}\left[\frac{n+1}{n}\hat{\theta}_n\right] = \theta \\
\Longleftrightarrow \\
\mathbb{E}\left[\frac{n+1}{n}\hat{\theta}_n\right] - \theta = 0
\end{gathered}
$$

Thus, $\frac{n+1}{n}\hat{\theta}_n$ is an unbiased estimator for θ.

```r
## Comparing the estimators for the uniform distribution.

n= 100

U = runif(n, 0, 4)

theta = max(U)


for (i in 1:50) {
  U = runif(n, 0, 4)
  theta = c(theta, max(U))
}


mean(theta)
```

```
## [1] 3.961274
```

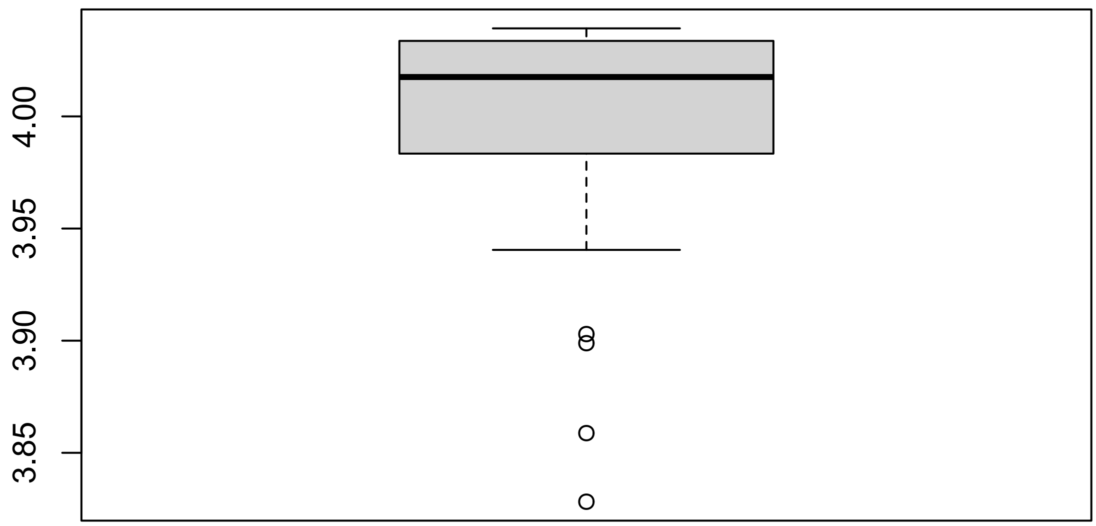
```r
boxplot(theta)
```




```r
## Adjust to make the estimator unbiased

thetab = (n+1)/n*theta

mean(thetab)
```

```
## [1] 4.000887
```

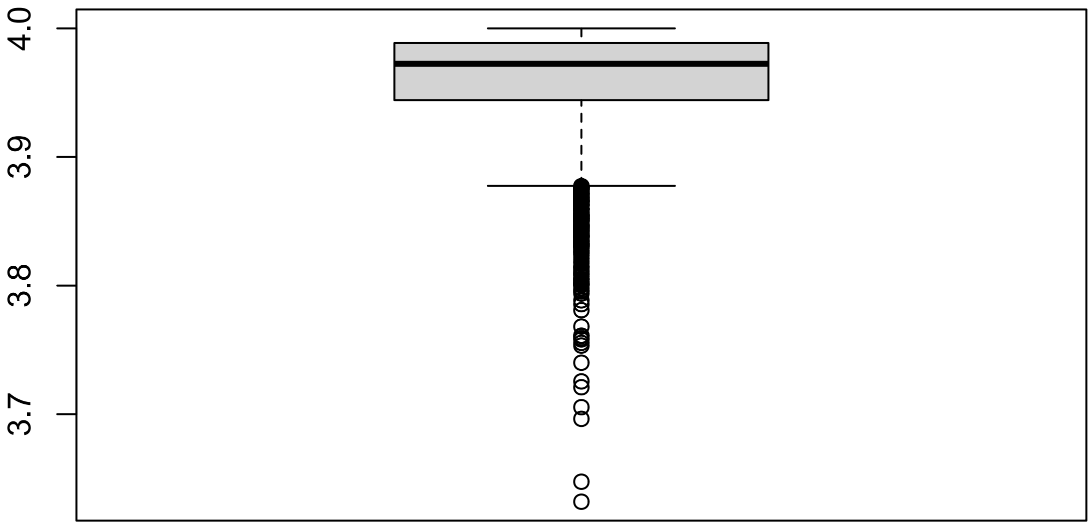
```r
boxplot(thetab)
```




```r
## With more observations

n= 100

U = runif(n, 0, 4)

theta = max(U)


for (i in 1:5000) {
  U = runif(n, 0, 4)
  theta = c(theta, max(U))
}


mean(theta)
```

```
## [1] 3.960553
```

```r
boxplot(theta)
```



Property Let $X_1, ..., X_n$ $i.i.d$ random variables.

- Let $\mu = \mathbb{E}[X_1]$ (unknown)
- Let $\sigma^2 = V[X_1]$ (unknown)

A classical estimator for:

- μ is

$$\hat{\mu}_n = \overline{X}_n$$

- σ^2 is

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum (X_i - \overline{X}_n)^2$$

Exercise

Show that:

1. $\hat{\mu}$ is unbiased.
2. $\hat{\sigma}_n^2$ is biased and that $\frac{n}{n-1}\hat{\sigma}_n^2$ is unbiased.

Solution:

1.

$$
\begin{aligned}
\mathbb{E}[\hat{\mu}_n] &= \mathbb{E}\left[\frac{1}{n}\sum X_i\right] \\
&= \frac{1}{n}\sum \mathbb{E}[X_i] \quad \text{(by linearity)} \\
&= \frac{1}{n} n \cdot \mathbb{E}[X_1] \quad \text{(as they are indentically distributed)} \\
&= \mathbb{E}[X_1]
\end{aligned}
$$

Therefore: $\hat{\mu}_n = \overline{X}_n$ is an unbiased estimator for μ.

2.

First note that we can rewrite $\hat{\sigma}_n^2$ as follows:

$$
\begin{aligned}
\hat{\sigma}_n^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(X_i^2 - 2X_i\overline{X}_n + \overline{X}_n^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\overline{X}_n\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) + \overline{X}_n^2 \quad \text{(check that factors that depends on i)} \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\overline{X}_n^2 + \overline{X}_n^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2
\end{aligned}
$$

Now use this expression to compute the expectation:

$$
\begin{aligned}
\mathbb{E}[\hat{\sigma}_n^2] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i^2] - \mathbb{E}\left[\overline{X}_n^2\right] \quad \text{(by linearity)} \\
&= \frac{1}{n}\sum_{i=1}^{n}(V[X_i] + \mathbb{E}[X_i]^2) - (V[\overline{X}_n] + \mathbb{E}[\overline{X}_n]^2) \quad \text{(by using the variance formula)} \\
&= \sigma^2 + \mu^2 - V[\overline{X}_n] - \mu^2 \quad \text{(by substituting the expressions of the variance and expectation)} \\
&= \sigma^2 - V[\overline{X}_n]
\end{aligned}
$$

Now lets compute $V[\overline{X}_n]$:

$$
\begin{aligned}
V[\overline{X}_n] &= V\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n^2} V\left[\sum_{i=1}^{n} X_i\right] \quad \text{(by linearity)} \\
&= \frac{1}{n^2}\sum_{i=1}^{n} V[X_i] \quad \text{(because they are indepent)} \\
&= \frac{\sigma^2}{n}
\end{aligned}
$$

Putting all together we get:

$$
b(\hat{\sigma}_n^2) = \mathbb{E}[\hat{\sigma}_n^2] - \sigma^2 = -V[\overline{X}_n] = -\frac{\sigma^2}{n}
$$

so $\hat{\sigma}_n^2$ is an biased estimator.

But we can modify it to obtain an unbiased version by noticed that:

$$
\mathbb{E}\left[\frac{n}{n-1}\hat{\sigma}_n^2\right] = \sigma^2
$$

Therefore:

$$
\hat{\sigma}_{n-1}^2 = \frac{n}{n-1}\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2
$$

is an unbiased estimator for σ^2.

Theorem Let $X_1, ..., X_n$ $i.i.d$ with $\mu = \mathbb{E}[X_1]$ and $\sigma^2 = V[X_1]$.

- An unbiased estimator for μ is:

$$
\hat{\mu}_n = \overline{X}_n \quad \text{and it's variance is} \quad \frac{\sigma^2}{n}
$$

- An unbiased estimator for σ^2 is:

$$
\hat{\sigma}_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2
$$

Quality of an estimator

Mean Quadratic Error **Definition:** Let θ an unknown parameter, let $\hat{\theta}_n$ an estimator of θ, the mean quadratic error (MQE) is given by:

$$MQE(\hat{\sigma}_n) = \mathbb{E}[(\hat{\sigma}_n - \sigma)^2]$$

Property:

$$MQE = V[\hat{\theta}_n] + ((b(\hat{\theta}_n)))^2$$

Proof:

Lets rewrite the MQE definition.

$$\begin{aligned} MQE(\hat{\theta}_n) &= \mathbb{E}[(\hat{\sigma}_n - \sigma)^2] \\ &= \mathbb{E}[(\hat{\sigma}_n - \mathbb{E}[\hat{\sigma}_n] + \mathbb{E}[\hat{\sigma}_n] - \sigma)^2] \quad \text{(by adding and substracting yhe same thing)} \\ &= \underbrace{\mathbb{E}[(\hat{\sigma}_n - \mathbb{E}[\hat{\sigma}_n])^2]}_{V[\hat{\sigma}_n]} + \underbrace{\mathbb{E}[(\mathbb{E}[\hat{\sigma}_n] - \sigma)^2]}_{\mathbb{E}[b(\hat{\sigma}_n)]^2} + 2\underbrace{\mathbb{E}[(\hat{\sigma}_n - \mathbb{E}[\hat{\sigma}_n])(\mathbb{E}[\hat{\sigma}_n] - \sigma)]}_{A_n} \end{aligned}$$

So to proof our result we just need to proof that $A_n = 0$.

$$\begin{aligned} A_n &= \mathbb{E}[\underbrace{(\mathbb{E}[\hat{\sigma}_n] - \sigma)}_{b(\hat{\sigma}_n)} \cdot (\hat{\sigma}_n - \mathbb{E}[\hat{\sigma}_n])] \\ &= \underbrace{b(\hat{\sigma}_n)}_{\text{deterministic term}} \cdot \mathbb{E}[\underbrace{(\hat{\sigma}_n - \mathbb{E}[\hat{\sigma}_n])}_{\text{this is a centeret r.v.}}] \quad \text{(the expectation of a centered r.v is zero.)} \\ &= 0 \end{aligned}$$

Comparison of estimators **Definition:** Let θ an unknown parameter, $\hat{\theta}_{n,1}$ and $\hat{\theta}_{n,2}$ for θ.

$\hat{\theta}_{n,1}$ is better than $\hat{\theta}_{n,2}$ if

$$\forall n \geq 1, \quad MQE(\hat{\theta}_{n,1}) \leq MQE(\hat{\theta}_{n,2})$$

Application Let $X_1, ..., X_n$ $i.i.d$ $\mathcal{U}([0,\theta])$, compare the following estimators to see if one is better than the others.

- $\hat{\theta}_{n,1} = 2\overline{X}_n$
- $\hat{\theta}_{n,2} = \max(X_n)$
- $\hat{\theta}_{n,3} = \frac{n+1}{n}\overline{X}_n$

Remember:

$f_{\hat{\theta}_{n,2}}(t) = n\dfrac{t^{n-1}}{\theta^n}\mathbf{1}_{[0,\theta]}(t)$

Solution:

Lets compute the MQE of each estimator.

1. Since $\hat{\theta}_{n,1}$ is an unbiased estimator for θ, we have:

$$
\begin{aligned}
MQE(\hat{\theta}_{n,1}) &= V[\hat{\theta}_{n,1}] \\
&= V[2\overline{X}_n] \\
&= 4V[\overline{X}_n] \\
&= 4\sum_{n=1}^{n} V[X_1] \quad \text{(by using i.i.d)} \\
&= 4\frac{\sigma^2}{12n} \quad \text{(check the variance of an uniform distribution)}
\end{aligned}
$$

Thus:

$$
MQE(\hat{\theta}_{n,1}) = \frac{\sigma^2}{3n}
$$

2. We already know that: $b(\hat{\theta}_{n,2}) = -\frac{\theta}{n+1}$ and $\mathbb{E}[\hat{\theta}_{n,2}] = \frac{n}{n+1}\theta$.

We have to compute $V[\hat{\theta}_{n,2}]$, since

$$
V[\hat{\theta}_{n,2}] = \mathbb{E}[(\hat{\theta}_{n,2})^2] - \mathbb{E}[\hat{\theta}_{n,2}]^2
$$

and

$$
\mathbb{E}[(\hat{\theta}_{n,2})^2] = \int_{\mathbb{R}} n\frac{t^{n-1}}{\theta^n}\mathbf{1}_{[0,\theta]}(t)dt = \frac{n}{n+2}\theta^2
$$

Therefore:

$$
\begin{aligned}
MQE(\hat{\theta}_{n,2}) &= V[\hat{\theta}_{n,2}] - b(\hat{\theta}_{n,2})^2 \\
&= \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2 + \frac{1}{(n+1)^2}\theta^2 \\
&= \frac{2}{(n+1)(n+2)}\theta^2
\end{aligned}
$$

3. Since $\hat{\theta}_{n,3}$ is an unbiased estimator for θ, we have:

$$
\begin{aligned}
MQE(\hat{\theta}_{n,3}) &= V[\hat{\theta}_{n,3}] \\
&= V\left[\frac{n+1}{n}\hat{\theta}_{n,2}\right] \\
&= \frac{(n+1)^2}{n^2}V[\hat{\theta}_{n,2}] \\
&= \frac{(n+1)^2}{n^2} \cdot \frac{n}{(n+1)^2(n+2)}\theta^2 \\
&= \frac{1}{n(n+2)}\theta^2
\end{aligned}
$$

It is easy to check that for $n \geq 1$:

$$3n \leq \frac{(n+1)}{2}(n+2) \leq n(n+2)$$

Thus

$$MQE(\hat{\theta}_{n,3}) \leq MQE(\hat{\theta}_{n,2}) \leq MQE(\hat{\theta}_{n,1})$$

and we conclude that $MQE(\hat{\theta}_{n,3})$ it's the best estimator for θ.

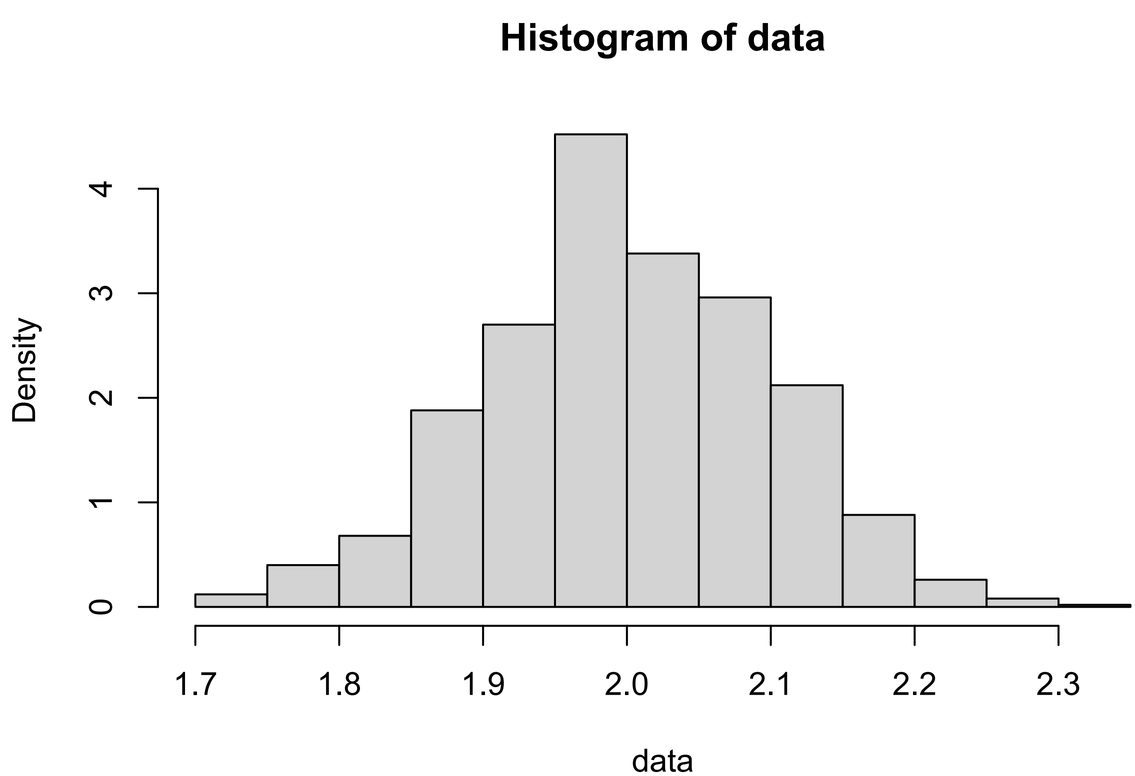
Practical class

1. Load the data in R software.
2. Propose a model for the variables associated to this file.

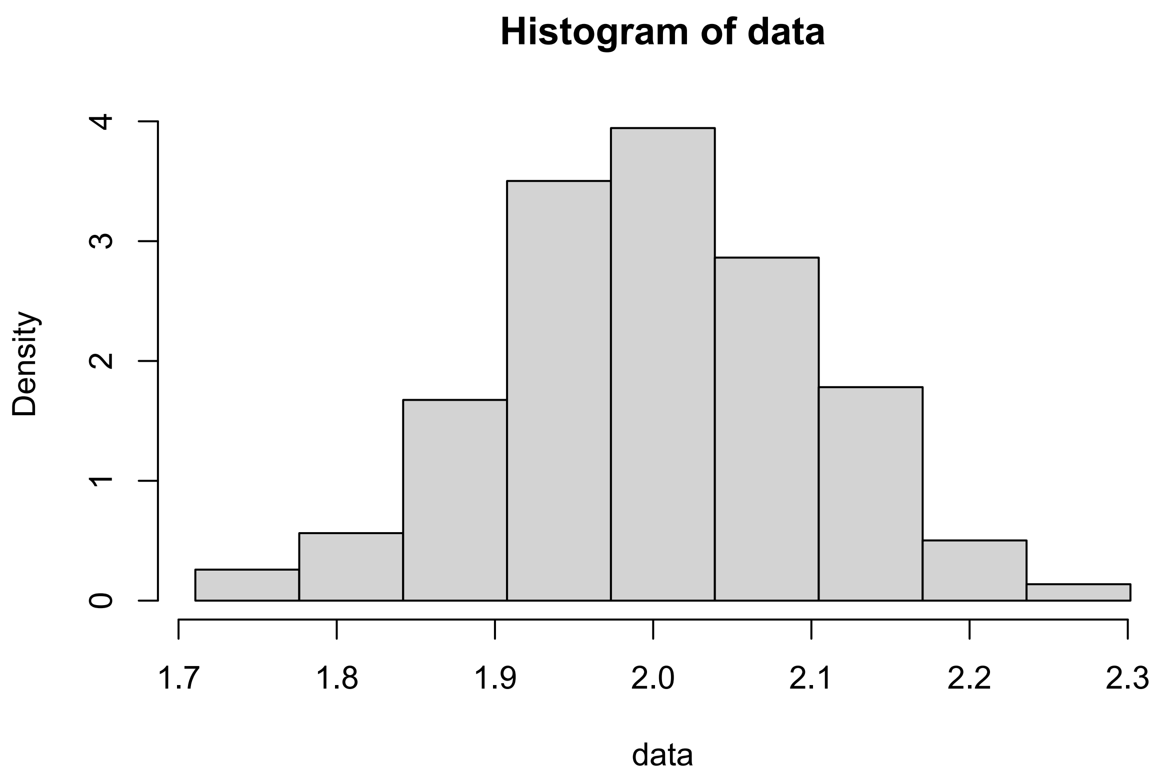
- Make a visualization of this.

```
data <- as.matrix(read.table("data/data1.txt"))

## Some comments about the data and data types.
```

```
hist(data, freq = FALSE)
```



```r
## Be careful about getting conclusions

data1_min = min(data)
data1_max = max(data)

hist(data, freq = FALSE, breaks = seq(data1_min, data1_max, length = 10))
```



We guess a Gaussian distribution.

How to estimate the parameters of the distribution?

1. Compute the empirical mean and statistical variance.
2. Plot the theoretical density that we guess.
3. Test the goodness of fitness.

Compute the empirical mean and statistical variance.

```r
mhu <- mean(data)

# When you compute the variance be careful and read the documentation
# in this case we have:
# "The denominator n - 1 is used which gives an unbiased estimator of the (co)variance
# for i.i.d. observations"

sigma2 <- var(data)
```

```r
## You also can do it manually:

sigma2 <- 1/(nrow(data)-1)*sum((data - mean(data))^2)
```

Plot the theoretical density that we guess.

```r
## This generates a list, with all the hist information.

H <- hist(data, freq = FALSE, plot = FALSE) ## ignore warning message
```

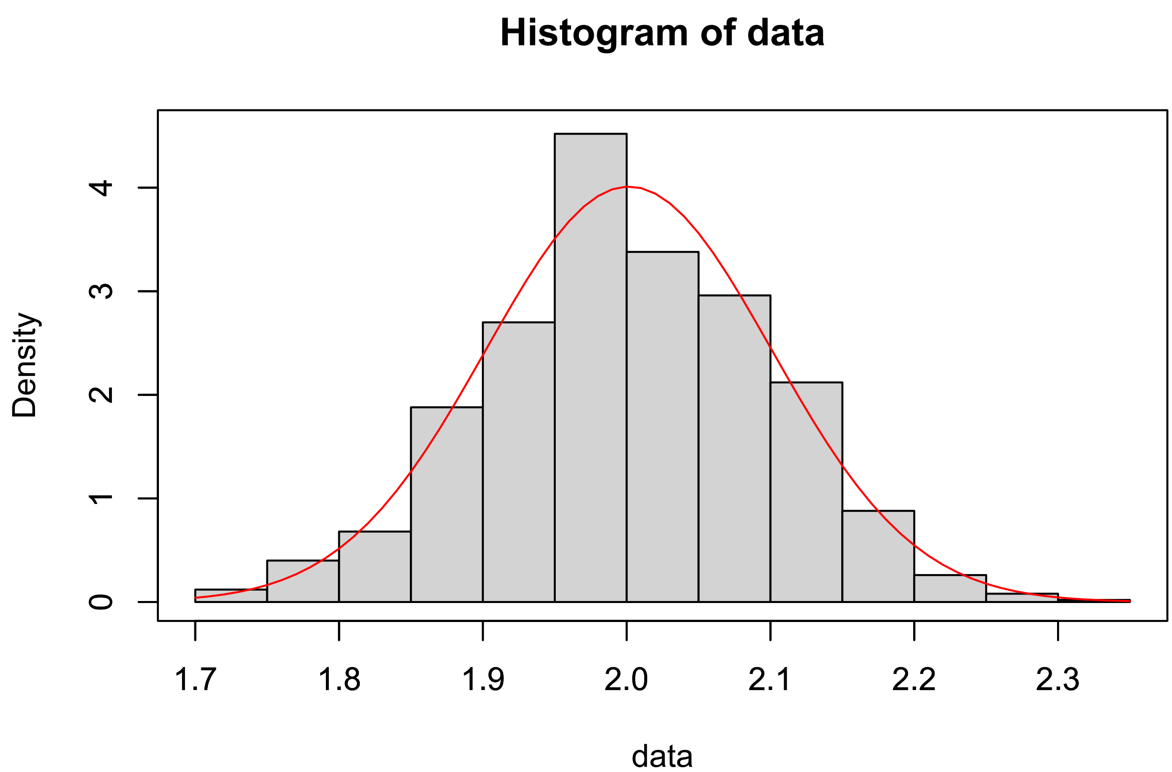
```
## Warning in hist.default(data, freq = FALSE, plot = FALSE): argument 'freq' is
## not made use of
```

```r
H
```

```
## $breaks
##  [1] 1.70 1.75 1.80 1.85 1.90 1.95 2.00 2.05 2.10 2.15 2.20 2.25 2.30 2.35
##
## $counts
##  [1]   6  20  34  94 135 226 169 148 106  44  13   4   1
##
## $density
##  [1] 0.12 0.40 0.68 1.88 2.70 4.52 3.38 2.96 2.12 0.88 0.26 0.08 0.02
##
## $mids
##  [1] 1.725 1.775 1.825 1.875 1.925 1.975 2.025 2.075 2.125 2.175 2.225 2.275
## [13] 2.325
##
## $xname
## [1] "data"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
```

```r
## Histogram info to use.
limits <- H$breaks
lmin <- limits[1]
lmax <- limits[length(limits)]

## Create our plot

x <- seq(lmin, lmax, by = 0.01)
y <- dnorm(x, mhu, sqrt(sigma2))


y_max = max(y, H$density)

## Make the comparison
```

```r
hist(data, freq = FALSE, xlim = c(lmin, lmax), ylim = c(0, y_max*1.01))
par(new = TRUE)
plot(x, y, type = 'l', col = 'red'
     , xlim = c(lmin, lmax)
     , ylim = c(0, y_max*1.01)
     , xlab = ""
     , ylab = "")
```



Test the goodness of fitness.

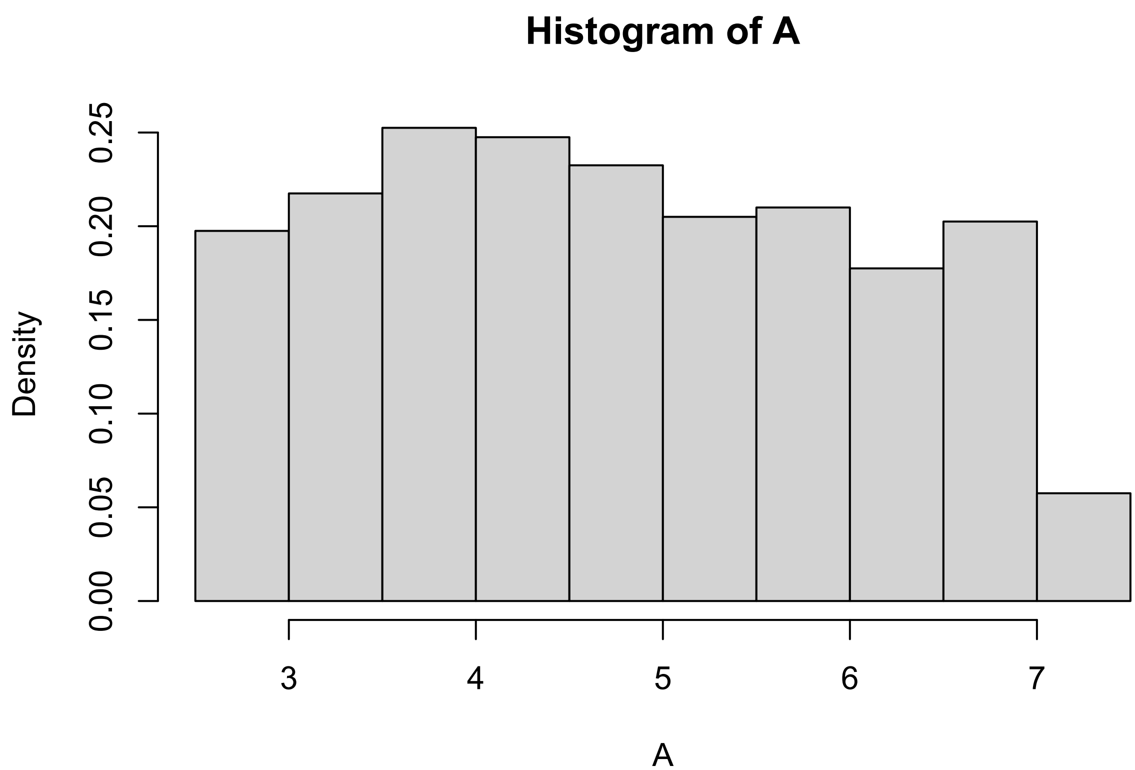
```r
ks.test(x, 'pnorm', mhu, sqrt(sigma2))
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  x
## D = 0.25147, p-value = 0.0003644
## alternative hypothesis: two-sided
```

Data set 2

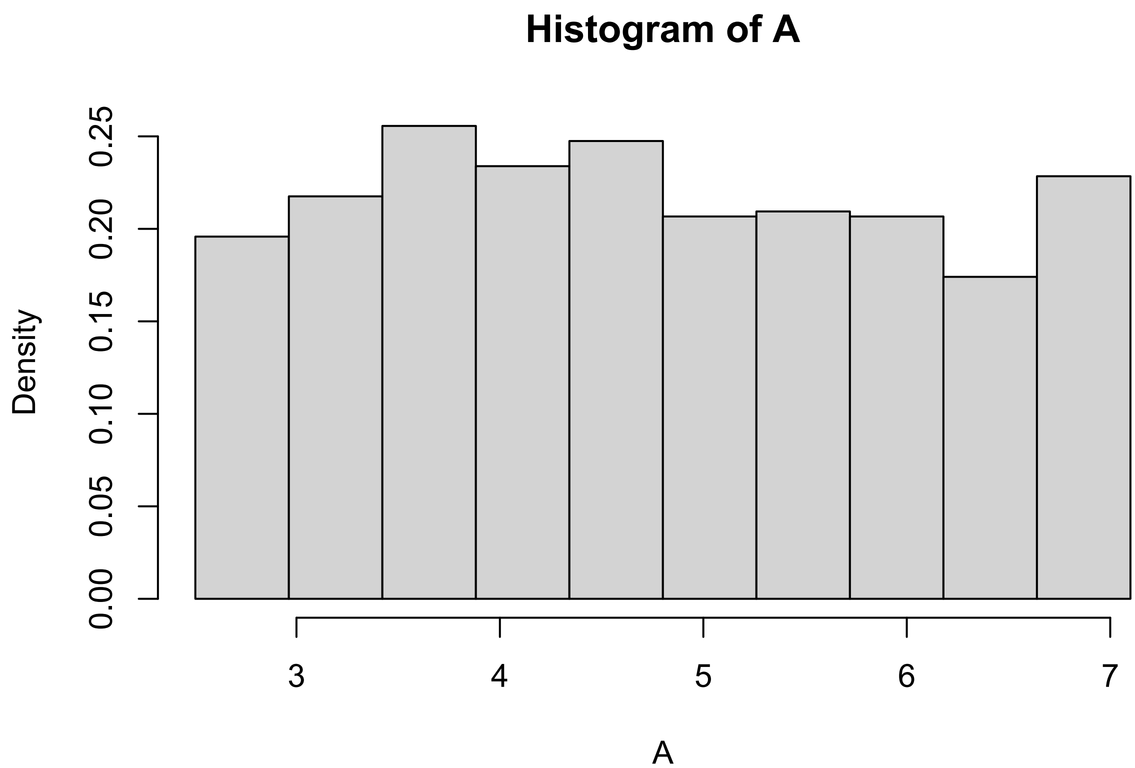
```r
## Loading the data
A <- as.matrix(read.table("data/data2.txt"))
```

```r
## Make visualizations
hist(A, freq = FALSE)
```

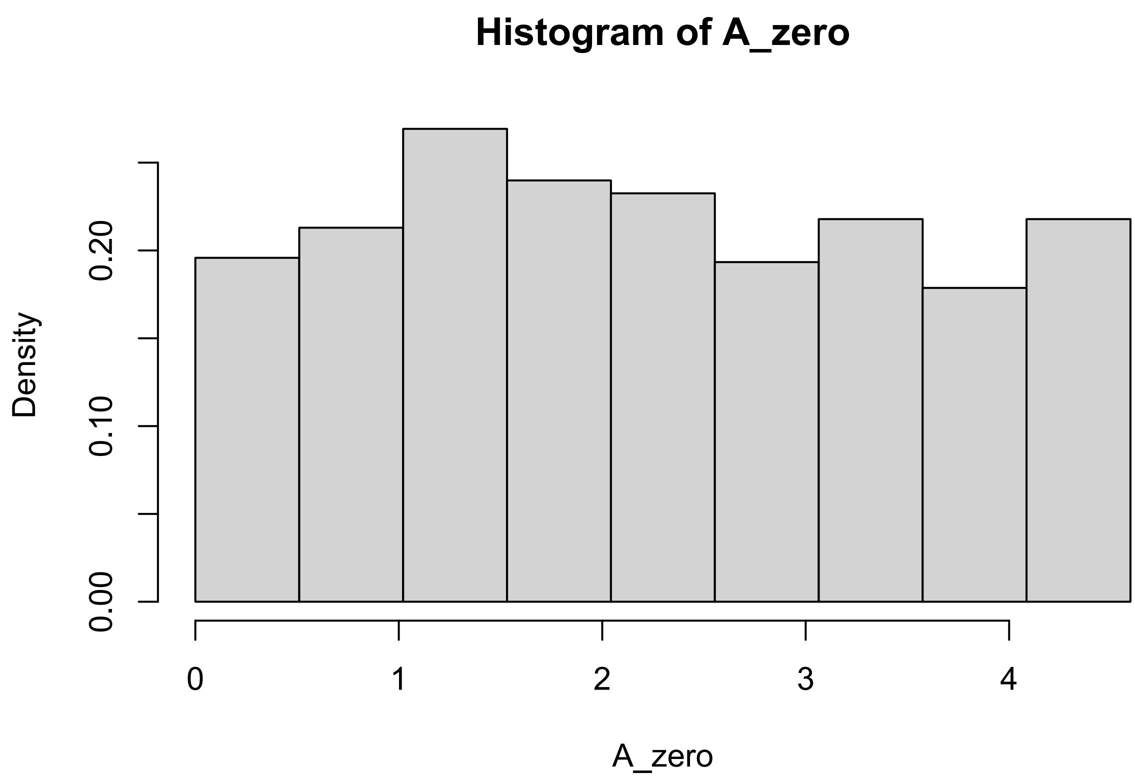



```r
## The right box exist because r creates one class with just one element.
## There is a formula that generates the right amount of classes, given by:
## K aproxx 1 + 3.22*log(n, 10) ## (in practice take the floor)


## Use right parameters:
n = length(A)
K = floor(1 + 3.22*log(n, 10))
A_min = min(A)
A_max = max(A)

A_mp = (A_max - A_min)/K

epsilon = (A_max - A_min)/10^9

limits = seq(A_min, A_max, by = A_mp)
```

```r
hist(A, freq = FALSE, breaks = limits)
```




```r
## We guess is an uniform distribution

## Make start in 0

A_zero = A - min(A)

A_min = min(A_zero)
A_max = max(A_zero)
hist(A_zero, freq = FALSE, breaks = seq(A_min, A_max, length = 10))
```

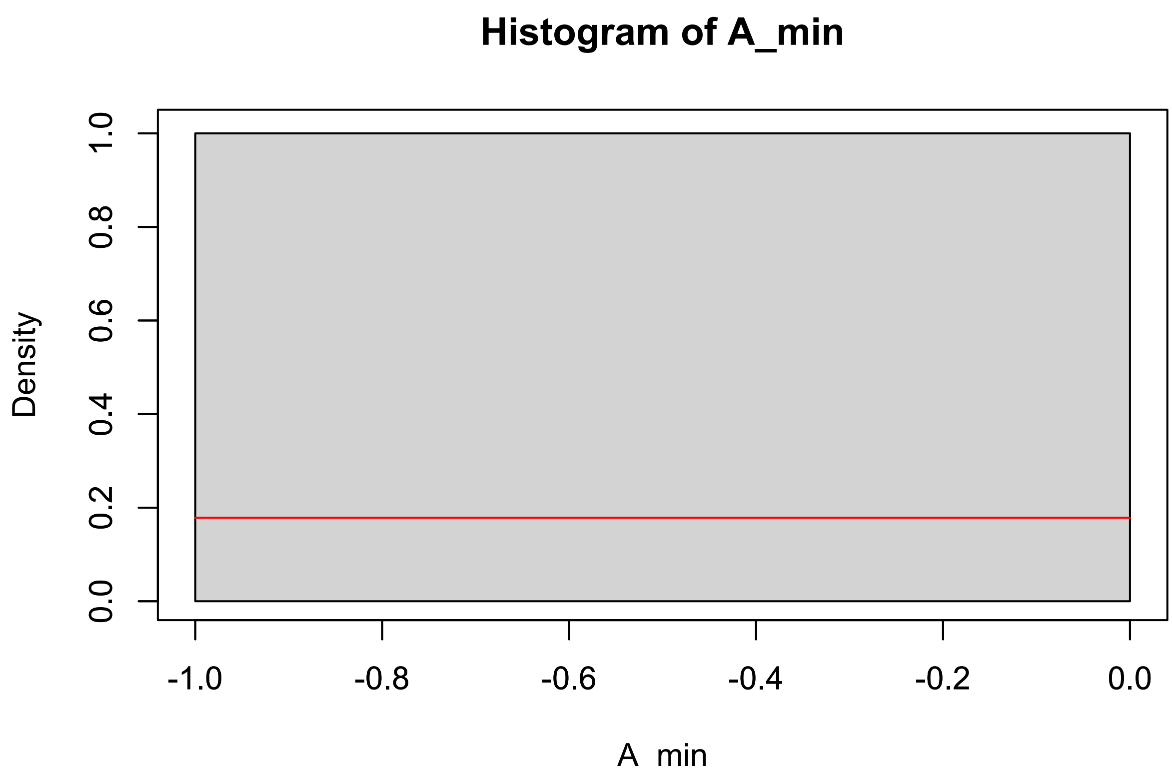



```r
## Estimate the parameter

n <- length(A_zero)

theta <- ((n+1)/n)*max(A_zero)

## Create our theoretical density plot


H <- hist(A_min, freq = FALSE, plot = FALSE) ## ignore warning message
```

```
## Warning in hist.default(A_min, freq = FALSE, plot = FALSE): argument 'freq' is
## not made use of
```

```r
## Histogram info to use.
limits <- H$breaks
lmin <- limits[1]
lmax <- limits[length(limits)]

## Create our plot

x <- seq(lmin, lmax, by = 0.01)
y <- dunif(x, min = lmin, max = theta )

y_max = max(y, H$density)
```

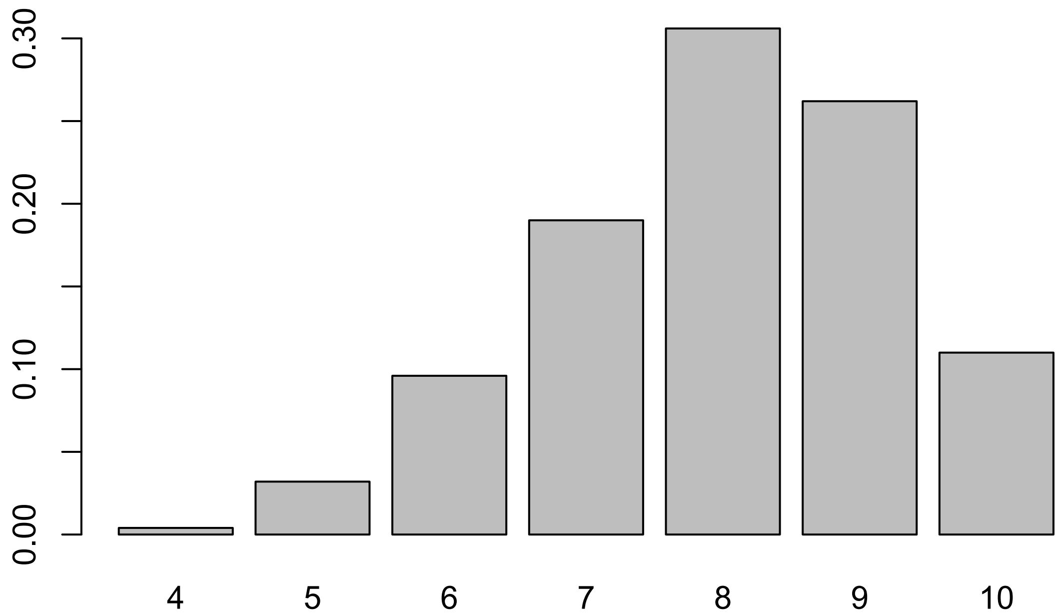
```r
## Make the comparison

hist(A_min, freq = FALSE, xlim = c(lmin, lmax), ylim = c(0, y_max*1.01))
par(new = TRUE)
plot(x, y, type = 'l', col = 'red'
     , xlim = c(lmin, lmax)
     , ylim = c(0, y_max*1.01)
     , xlab = ""
     , ylab = "")
```



Discrete case

data4.txt

```r
## Loading the data
A <- as.matrix(read.table("data/data4.txt"))

barplot(table(A)/length(A))
```




```
## A binomial distribution can be approach by a Gaussian distribution.

## Methods of moments
## Don't forget to correct the estimation of np to be an integer.
## and then correct the ph value




m = mean(A)
s2 = var(A)*(499/500)

ph = 1-s2/m
nh = m/ph

nh
```

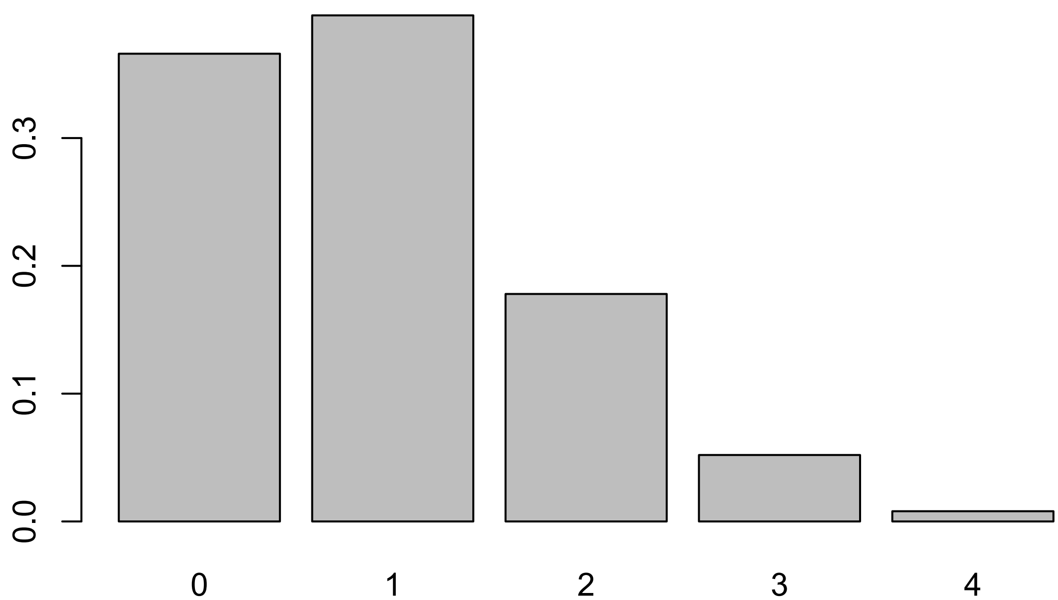
```
##          V1
## V1 10.0325
```

```
nh = 10

ph = m/nh

ph
```

```
## [1] 0.7988
```

data5.txt

```r
## Loading the data
A <- as.matrix(read.table("data/data5.txt"))

barplot(table(A)/length(A))
```




```r
## A binomial distribution can be approach by a Gaussian distribution.

## Methods of moments
## Don't forget to correct the estimation of np to be an integer.
## and then correct the ph value


m = mean(A)
s2 = var(A)*(499/500)

ph = 1-s2/m
nh = m/ph

nh
```

```
##          V1
## V1 7.38796
```

```r
nh = 7

ph = m/nh

ph
```

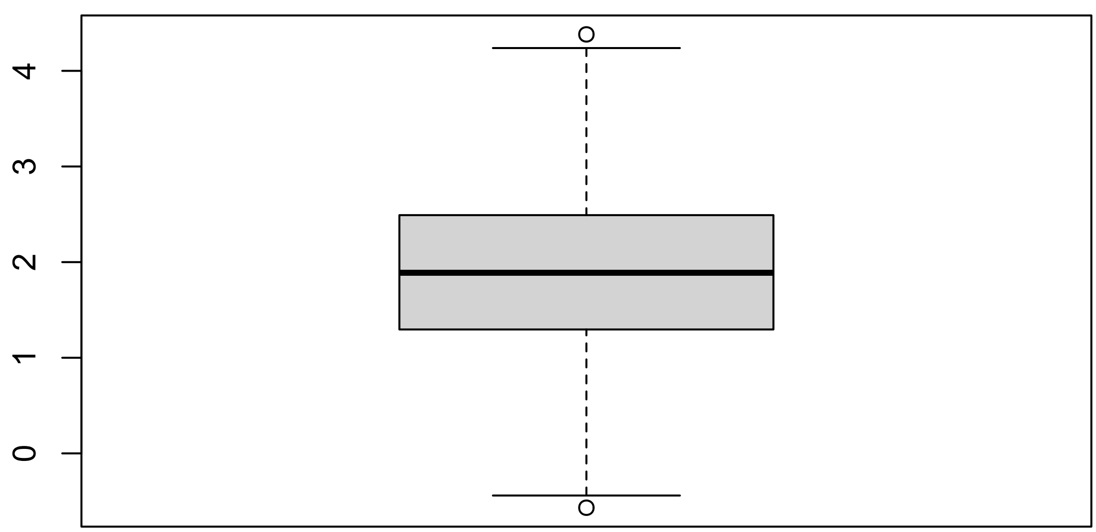
```
## [1] 0.1342857
```

```r
## When p is small is very hard to get correct estimations of n when you
## don't have enough number of observations
```

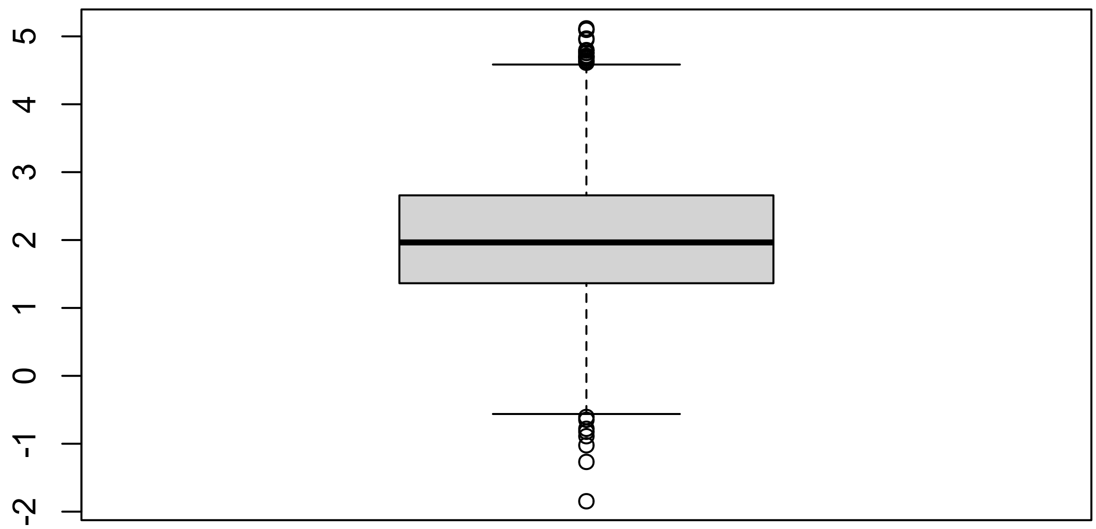
Day 3 (In progress)

Confidence interval

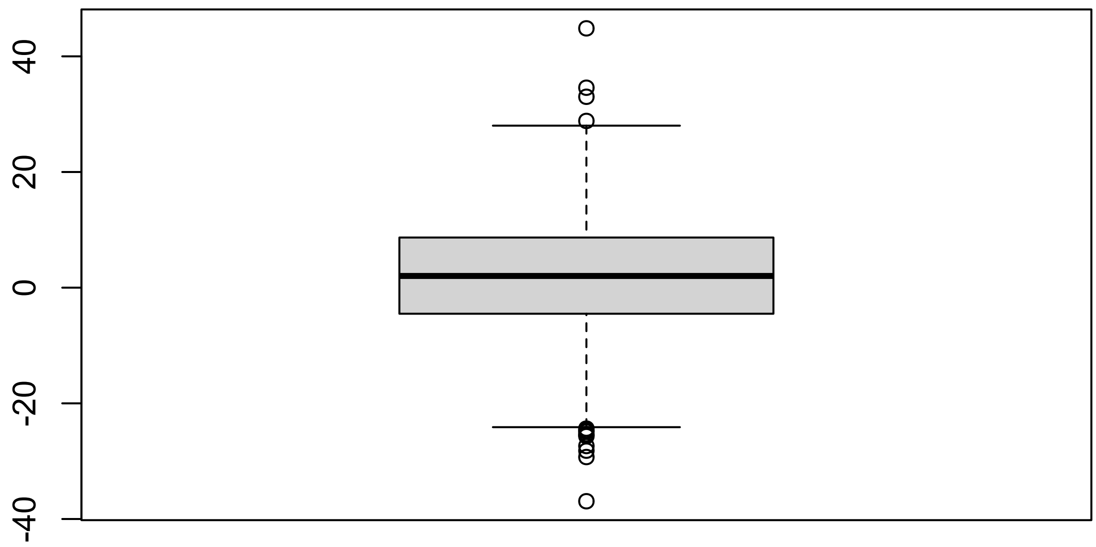
```r
m1 = c()

for (i in 1:50) {
  A = rnorm(500, 2, 1)

  m1 = c(m1, mean(A))
}

boxplot(A)
```



```r
## Increase the number of observations

m2 = c()

for (i in 1:50) {
  B = rnorm(2000, 2, 1)

  m2 = c(m2, mean(B))
}

boxplot(B)
```



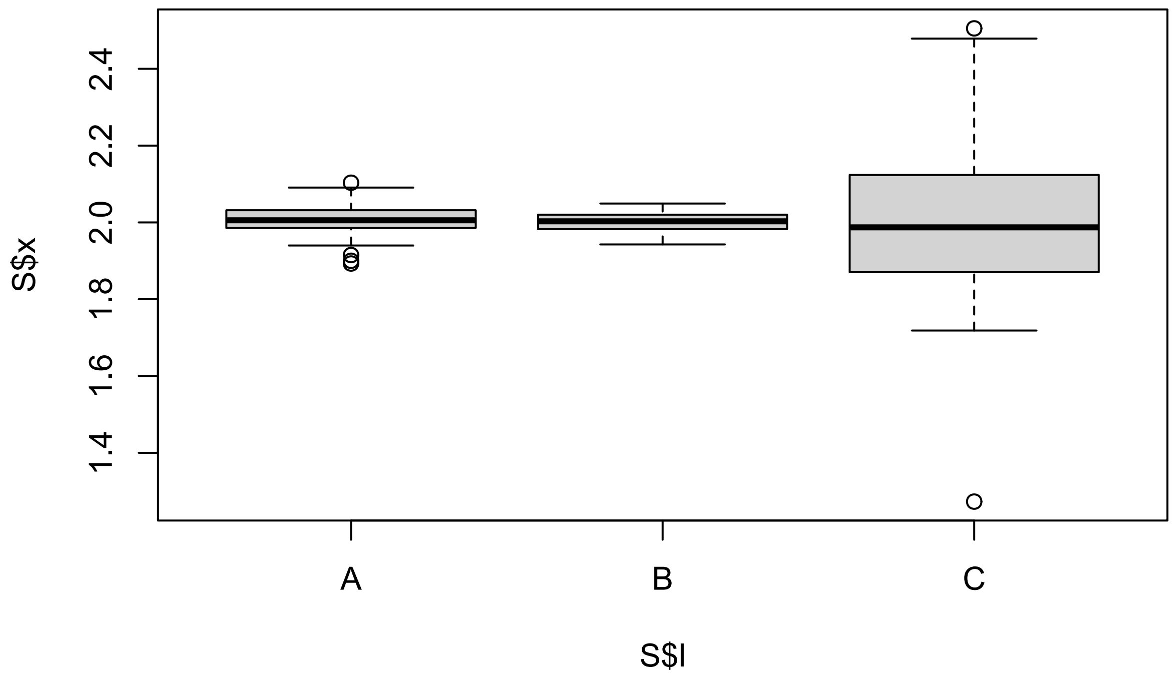

```r
## Change the variance

m3 = c()

for (i in 1:50) {
  C = rnorm(2000, 2, 10)

  m3 = c(m3, mean(C))
}

boxplot(C)
```




```r
## Create a data frame with all the observations

a = c(m1, m2, m3)
r = rep(c("A", "B", "C"), each = 50)

S = data.frame(x = a, l = r)
```

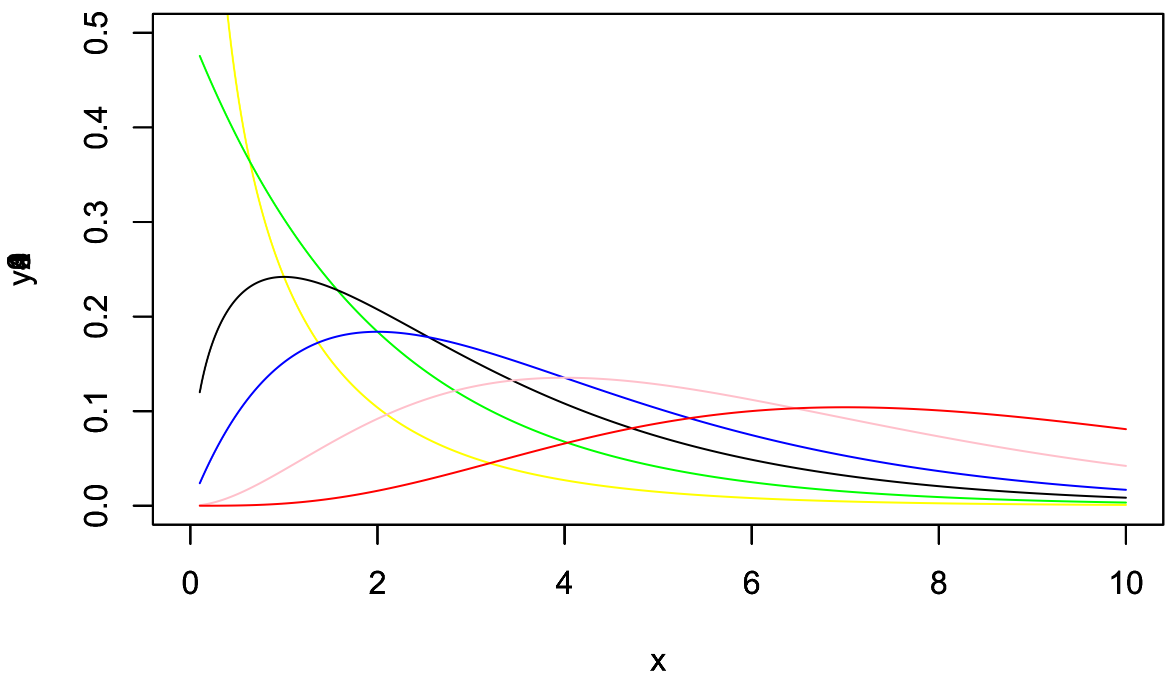
In this plot we can see that the approximation depends on the number of observations and the variance of the sample.

```r
boxplot(S$x ~ S$l)
```



Chi squared distribution

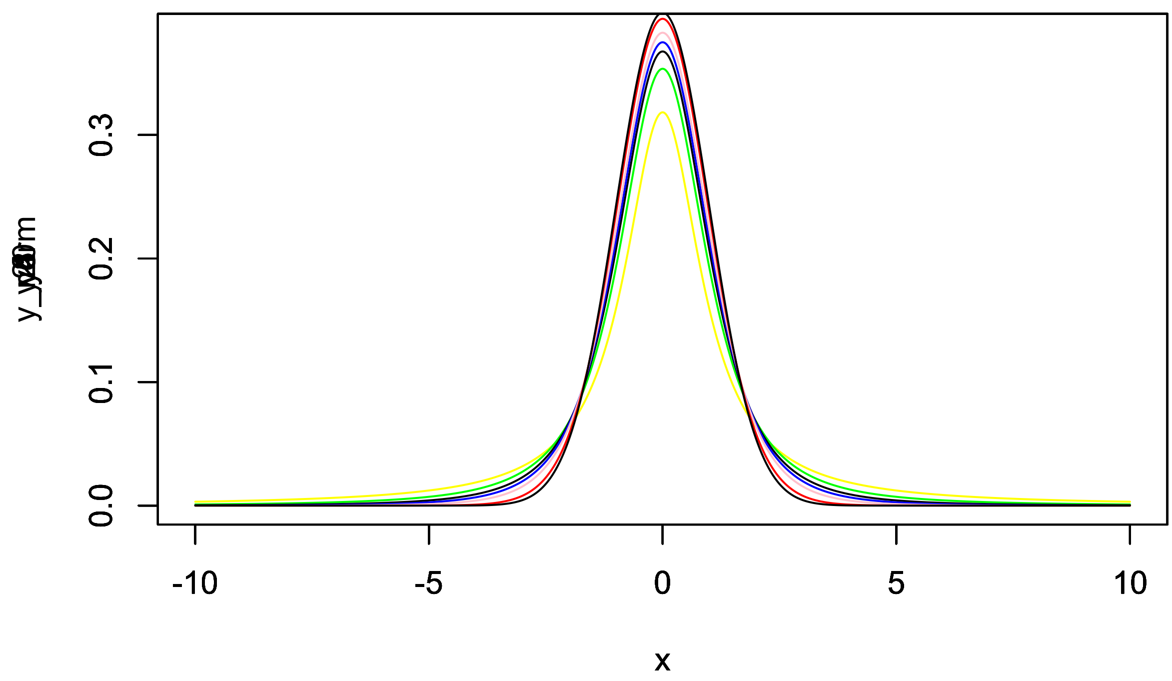
```
x = seq(0.1, 10, by =0.01)

y1 = dchisq(x, 1)
y2 = dchisq(x, 2)
y3 = dchisq(x, 3)
y4 = dchisq(x, 4)
y6 = dchisq(x, 6)
y9 = dchisq(x, 9)

y_max = max(y1, y2, y3, y4, y6, y9)

plot(x, y1, type = 'l', col = 'yellow', xlim = c(0, 10), ylim = c(0, 0.5))
par(new = TRUE)
plot(x, y2, type = 'l', col = 'green', xlim = c(0, 10), ylim = c(0, 0.5))
par(new = TRUE)
plot(x, y3, type = 'l', col = 'black', xlim = c(0, 10), ylim = c(0, 0.5))
par(new = TRUE)
plot(x, y4, type = 'l', col = 'blue', xlim = c(0, 10), ylim = c(0, 0.5))
par(new = TRUE)
plot(x, y6, type = 'l', col = 'pink', xlim = c(0, 10), ylim = c(0, 0.5))
par(new = TRUE)
plot(x, y9, type = 'l', col = 'red', xlim = c(0, 10), ylim = c(0, 0.5))
```



t-student distribution

```
x = seq(-10, 10, by =0.01)

y1 = dt(x, 1)
y2 = dt(x, 2)
y3 = dt(x, 3)
y4 = dt(x, 4)
y6 = dt(x, 6)
y20 = dt(x, 20)

y_norm = dnorm(x)

y_max = max(y1, y2, y3, y4, y6, y9)

plot(x, y1, type = 'l', col = 'yellow', xlim = c(-10, 10), ylim = c(0, y_max))
par(new = TRUE)
plot(x, y2, type = 'l', col = 'green', xlim = c(-10, 10), ylim = c(0, y_max))
par(new = TRUE)
plot(x, y3, type = 'l', col = 'black', xlim = c(-10, 10), ylim = c(0, y_max))
par(new = TRUE)
plot(x, y4, type = 'l', col = 'blue', xlim = c(-10, 10), ylim = c(0, y_max))
par(new = TRUE)
plot(x, y6, type = 'l', col = 'pink', xlim = c(-10, 10), ylim = c(0, y_max))
```

```r
par(new = TRUE)
plot(x, y20, type = 'l', col = 'red', xlim = c(-10, 10), ylim = c(0, y_max))
par(new = TRUE)
plot(x, y_norm, type = 'l', col = 'black', xlim = c(-10, 10), ylim = c(0, y_max))
```




```r
## A t-student distribution with a big enough k approach a Gaussian distribution
```

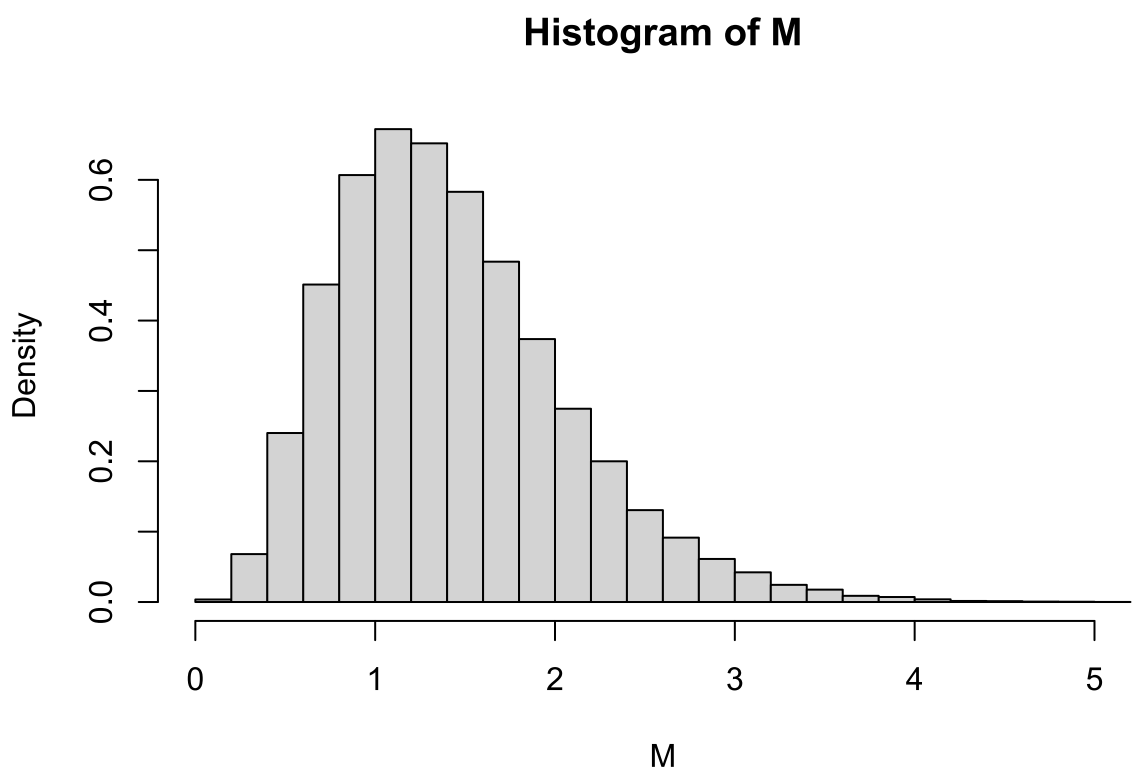
Day 4 (In progress)

Use of t-student table

Confidence interval

```r
## We compute N observations of \overline{X}_n.


## Number of r.v.
n <- 5

## Number of observations
N <- 50000

## Parameter exp distribution.
```

```r
lambda <- 0.7

## Create a matrix
A <- matrix(rexp(n*N, lambda), ncol = n)

## Compute the mean for each row
M <- apply(A, 1, mean)

## Visualization for M

hist(M, freq = F, breaks = 20)
```




```r
###

Mo <- sort(M)
y <- 1:N/N

max_Mo <- max(Mo)

x <- seq(0, max_Mo, by = 0.01)
mu <- 1/lambda
v <- 1/(lambda^2)/n

yt <- pnorm(x, mu, sqrt(v))
```

```
plot(Mo, y, type = 'l', xlim = c(0, max_Mo), ylim = c(0,1))
par(new = TRUE)
plot(x, yt, type = 'l', xlim = c(0, max_Mo), ylim = c(0,1), col = 'red')
```