ECM 3A-DDEFi 2024-2025

Fonctions d'autocorrélation et d'autocorrélation partielle [1]

Objectifs :

1. *Appréhender la différence entre les probabilités et la statistique dans l'étude des séries chronolgiques.*
2. *Analyser les comportements des fonctions d'autocorrélation et d'autocorrélation partielle pour les processus AR, MA et ARMA.*

1 Introduction au logiciel R

Le logiciel R est un logiciel libre ressemblant au logiciel S-Plus et spécialisé en statistique. A partir du logiciel de base, un grand nombre de packages ont été créés par de nombreux utilisateurs, permettant ainsi d'aborder une grande diversité de problèmes statistiques à l'aide de techniques statistiques récentes. On compte actuellement plusieurs milliers de packages.
Le logiciel R a d'ailleurs attiré l'attention d'Oracle qui l'a incorporé Ã son offre de logiciels de base de données. Il existe aussi de nombreux connecteurs entre R et diffénts logiciels de bases de données tels que MongoDB, MariaDB, MySQL,...

1.1 What is R?

R is very similar to Scilab and Matlab in spirit. It is a free software avalailable for many OS. It is very convenient for conducting statistical analyzes.
R as Scilab and Matlab is a command line software. You can write scripts in an editor and save them in files. To execute a script, you can use instruction *source()*:

```
source("myscript.r")
```

Alternatively you can use R IDE called *RStudio*. It allows you to write scripts, save and excute them, visualize graphs and data.
If you need help, you can type

```
help(instruction_name)
```

or search the web. A good starting point is the official website of the CRAN:

```
http://cran.r-project.org/
```

There are also online manuals.
You can write comments in your program: use the sign # followed by your comments.

1.2 Data structures

Here is a quick list of common data structures in R

- data frame,
- list,
- matrix and vector,
- real and boolean.

[1]Version du 3/10/2022

In the sequel we will be mainly interested in data frame. A data frame is a data structure which can store different types of data together (e.g. numbers, strings,...).
Here is an example

```
# Defining a vector
vect1 <- 1:4
# Defining a list of letters
vect2 <- c("A","B","C","D")
# Defining a data frame called mydata
mydata <- data.frame(numbers=vect1,letters=vect2)
# Column names of the data frame mydata
colnames(mydata)
# Dimension of the data frame mydata: number of rows, number of columns
dim(mydata)
```

If you need to remove an object, you can use instruction rm():

```
rm(vect1)
```

To list all stored objects, you can use ls():

```
ls()
```

Let us see how to generate random numbers.

```
> set.seed(13) # Set seed of the random number generator
> rnorm(n=3) # Generate 3 standard Gaussian random variables
[1]  0.5543269 -0.2802719  1.7751634
```

If you want uniform random variables, you can use

```
> runif(n=3)
[1] 0.3574129 0.5914571 0.8654515
```

You can even choose the support of the uniform distribution, i.e. the lower and the upper bounds

```
> runif(n=3,min=-1,max=1)
[1]  0.36104732 -0.72586690  0.09389775
```

Before simulating other distributions, let us store the results in data structures such as vector, matrix or dataframe.
To store the samples in a vector, use

```
> vec1 <- runif(n=3)
> is.vector(vec1) # Check that vec1 is a vector
[1] TRUE
```

Next let us store the results in a matrix

```
> vec2 <- matrix(runif(n=6),3,2)
> is.vector(vec2) # Check that vec2 is not a vector
[1] FALSE
> is.matrix(vec2) # Check that vec2 is a matrix
[1] TRUE
```

The last data structure we are interested in is the data frame. This structure is versatile but there are some drawbacks.

```
> vec3 <- rnorm(n=3)
> vec4 <- runif(n=3)
> mydata <- data.frame(normalvar=vec3,unifvar=vec4)
> is.data.frame(mydata)
[1] TRUE
> colnames(mydata) # normalvar and unifvar are the column names
[1] "normalvar" "unifvar"
```

1.3 Reading data from an external file

It is easy to read data from an external file if this file is well organized: column names on the first line, classical field separators (blank space, semi-colon, tabulation or comma).

```
# Reading data from external file canadiandollar.csv
dollar.canadian <- read.table("canadiandollar.csv",sep=";",header=T)
# Reshaping the data frame
dollar.canadian <- dollar.canadian[,2:5]
# Generating a column with the number of months
month <- 1:dim(dollar.canadian)[1]
dollar.canadian <- cbind(month,dollar.canadian)
```

1.4 Loop

Although loops are not very efficient in R (as in Scilab and Matlab), you can use them for quickly and "dirtily" generating data. Here is an example to generate a series with a linear trend, seasonality equal to 12 and Gaussian noise.

```
n <- 120
vect1 <- rep(0,n)
for (t in 1:n) {
u <- rnorm(1,mean=0,sd=5)
vect1[t] <- 2+3*t+10*cos(pi/6*t) +u
}
vect2 <- 1:n
mydata <- data.frame(t=vect2,x=vect1)
plot(x~t,data=mydata,type="l")
```

1.5 Graphs

Graphs are easy in R. You can plot one graph. This graph is the rate Canadian Dollar versus Icelandic Krone.

```
plot(ISK~month,data=dollar.canadian,main="Canadian dollar for
one Icelandic Krone",xlab="Months",ylab="Rate ISK",type="l")
```

The options used above are the following

- xlab: name of the X-axis
- ylab: name of the Y-axis
- main: title of the graph
- type: type of interpolation (l means line)

You can draw several graphs at the same time but in different graphics windows.

```
# Rate of Swiss Francs
plot(CHF~month,data=dollar.canadian,xlab="Months",ylab="Rate CHF",type="l")
# Rate of Sterling Pound
X11()
plot(GBP~month,data=dollar.canadian,xlab="Months",ylab="Rate GBP",type="l")
```

Instruction X11() forces the second graph to be drawn in a new graphic window.

Exercise: modify the program above and draw graphs for the indian rupee, the swiss franc and the sterling pound. Find a way to save the graph in a file (e.g. file format jpeg, png, gif).

1.6 Writing functions

You can write functions in order to build modular programs. Let us write a function to generate data.

```
generatemysample <- function(n){
vect1 <- rep(0,n)
vect2 <- seq(1,n)
# Loop
for (t in 1:n) {
# generate a Gaussian random variable with mean=0 and standard deviation=5
u <- rnorm(1,mean=0,sd=5)
vect1[t] <- 2+3*t+10*cos(pi/6*t) +u
}
tempo <- data.frame(t=vect2,x=vect1)
return(tempo)}
```

The syntax is simple:

```
name_of_my_function <- function(parameters){
instructions
}
```

You can return a value (use the return instruction) or not.

2 Analyse de séries temporelles avec R

Nous allons simuler des données afin d'illustrer les résultats du cours sur les séries temporelles. Faire des simulations permet de contrôler les paramètres et donc mettre en évidence les comportements des séries temporelles connus théoriquement.

2.1 Processus non-stationnaire

Le but de l'exercice est de simuler le processus

$$X_t = X_{t-1} + \varepsilon_t, t \in \mathbb{N}^\star, \quad X_0 = 0.$$

Le processus $\{\varepsilon_t, t \in \mathbb{N}^\star\}$ est un bruit blanc gaussien de variance $\sigma^2 = 1$.

1) Simuler le processus X_t pour $t = 1, \ldots, n$ avec $n = 200$.

2) Tracer le processus X_t en fonction du temps. Pouvez-vous expliquer en quoi le comportement est non-stationnaire?

3) Tracer la fonction d'autocorrélation. Que constatez-vous sur son comportement?

2.2 Processus MA(q)

Nous souhaitons illustrer le comportement des fonctions d'autocorrélation et d'autocorrélation partielle pour les processus MA(q).
Nous allons d'abord simuler un processus MA(1) :

$$X_t = \varepsilon_t + \theta\varepsilon_{t-1}, t \in \mathbb{N}^\star.$$

1) On choisit $\theta = 0.9$.
Le processus est-il inversible?
Simuler le processus X_t pour $t = 1, \ldots, n$, avec $n = 1000$.
Tracer les fonctions d'autocorrélation et d'autocorrélation partielle. Quels comportements constatez-vous? Comparer avec les comportements théoriques présentés en cours.

2) On choisit $\theta = -0.9$.
Le processus est-il inversible?
Simuler le processus X_t pour $t = 1, \ldots, n$, avec $n = 1000$.
Tracer le processus X_t en fonction du temps.
Tracer les fonctions d'autocorrélation et d'autocorrélation partielle. Quels comportements constatez-vous? Comparer avec les comportements théoriques présentés en cours.

Nous allons maintenant simuler un processus MA(2) :

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}, t \in \mathbb{N}^\star.$$

3) On choisit $\theta_1 = 0.3$ et $\theta_2 = -0.4$.
Le processus est-il inversible?
Simuler le processus X_t pour $t = 1, \ldots, n$, avec $n = 1000$.
Tracer le processus X_t en fonction du temps.
Tracer les fonctions d'autocorrélation et d'autocorrélation partielle. Quels comportements constatez-vous? Comparer avec les comportements théoriques présentés en cours.

2.3 Processus AR(p)

Nous souhaitons illustrer le comportement des fonctions d'autocorrélation et d'autocorrélation partielle pour les processus AR(p).
Nous allons d'abord simuler un processus AR(1) :

$$X_t = \phi X_{t-1} + \varepsilon_t, t \in \mathbb{N}^\star.$$

4) On choisit $\phi = 0.7$.
Le processus est-il causal?
Simuler le processus X_t pour $t = 1, \ldots, n$, avec $n = 1000$.
Tracer le processus X_t en fonction du temps.
Tracer les fonctions d'autocorrélation et d'autocorrélation partielle. Quels comportements constatez-vous? Comparer avec les comportements théoriques présentés en cours.

5) On choisit $\phi = -0.7$.
Le processus est-il causal?
Simuler le processus X_t pour $t = 1, \ldots, n$, avec $n = 1000$.
Tracer le processus X_t en fonction du temps.
Tracer les fonctions d'autocorrélation et d'autocorrélation partielle. Quels comportements constatez-vous? Comparer avec les comportements théoriques présentés en cours.

6) On choisit $\phi = 0.99$.
Le processus est-il stationnaire? Est-il causal?
Simuler le processus X_t pour $t = 1, \ldots, n$, avec $n = 200$.
Tracer le processus X_t en fonction du temps.
Avez-vous l'impression que le processus X_t est stationnaire?
Tracer les fonctions d'autocorrélation et d'autocorrélation partielle. Quels comportements constatez-vous? Comparer avec les comportements théoriques présentés en cours.
Recommencer les étapes ci-dessus avec $n = 1000$ (vous pouvez augmenter cette valeur pour mieux appréhender le phńomène). Quelle est votre impression quant à la stationnarité du processus?

Nous allons maintenant simuler un processus AR(2) :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t, t \in \mathbb{N}^\star.$$

7) On choisit $\phi_1 = 1.2$ et $\phi_2 = -0.35$.
Le processus est-il causal?

Simuler le processus X_t pour $t = 1, \dots, n$, avec $n = 1000$.
Tracer les fonctions d'autocorrélation et d'autocorrélation partielle. Quels comportements constatez-vous?
Calculer la fonction d'autocovariance et la fonction d'autocorrélation. L'observation de la fonction d'autocorrélation pour votre simulation correspond-elle au comportement théorique?

8) On choisit $\phi_1 = 0.5$ et $\phi_2 = 0.24$.
Le processus est-il causal?
Simuler le processus X_t pour $t = 1, \dots, n$, avec $n = 1000$.
Tracer les fonctions d'autocorrélation et d'autocorrélation partielle. Quels comportements constatez-vous?
Calculer la fonction d'autocovariance et la fonction d'autocorrélation. L'observation de la fonction d'autocorrélation pour votre simulation correspond-elle au comportement théorique?

2.4 Processus ARMA(p,q)

Nous souhaitons illustrer le comportement des fonctions d'autocorrélation et d'autocorrélation partielle pour les processus ARMA(p).
Nous allons maintenant simuler un processus ARMA(1,1) :

$$X_t = \phi X_{t-1} + \varepsilon_t + \theta \varepsilon_{t-2}, t \in \mathbb{N}^\star.$$

9) On choisit $\phi = -0.5$ et $\theta = 0.25$.
Le processus est-il causal et inversible?
Simuler le processus X_t pour $t = 1, \dots, n$, avec $n = 1000$.
Tracer les fonctions d'autocorrélation et d'autocorrélation partielle. Quels comportements constatez-vous?
Calculer la fonction d'autocovariance et la fonction d'autocorrélation. L'observation de la fonction d'autocorrélation pour votre simulation correspond-elle au comportement théorique?

Nous allons maintenant simuler un processus ARMA(2,1) :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t + \theta \varepsilon_{t-1}, t \in \mathbb{N}^\star.$$

10) On choisit $\phi_1 = -0.5, \phi_2 = 0.4$ et $\theta = 0.25$.
Le processus est-il causal et inversible?
Simuler le processus X_t pour $t = 1, \dots, n$, avec $n = 1000$.
Tracer les fonctions d'autocorrélation et d'autocorrélation partielle. Quels comportements constatez-vous?
Calculer la fonction d'autocovariance et la fonction d'autocorrélation. L'observation de la fonction d'autocorrélation pour votre simulation correspond-elle au comportement théorique?