Paris School of Business (PSB)

SERIE TEMPORELLE : MODELE ARIMA

De:
Arnaud Bruel YANKO KOUATCHOU

MSc Data Management

November 20, 2020

PLAN DU TRAVAIL

Plan du travail

1. Introduction
2. Séries temporelles
3. Modélisation des séries temporelles
4. Application sur *R*

PLAN DU TRAVAIL

Plan du travail

1. Introduction
2. Séries temporelles
3. Modélisation des séries temporelles
4. Application sur *R*

PLAN DU TRAVAIL

Plan du travail

1. Introduction
2. Séries temporelles
3. Modélisation des séries temporelles
4. Application sur *R*

PLAN DU TRAVAIL

Plan du travail

1. Introduction
2. Séries temporelles
3. Modélisation des séries temporelles
4. Application sur R

PLAN DU TRAVAIL

Introduction
Introduction

Séries temporelles

Definition 1

Une série chronologique (ou temporelle) est une succession d'observations au cours du temps: $U_t : t = 1, 2, \cdots, n, \cdots = (U_1, U_2, \cdots, U_n, \cdots)$

Particularité des séries temporelles

Tient à la présence d'une relation d'antériorité qui ordonne l'ensemble des informations. Les dates d'observations sont souvent équidistantes les unes des autres : on a des séries mensuelles, trimestrielles, etc, dans quel cas on peut les indexer par $t \in \mathbb{N}$

Séries temporelles

Exemple 1

- *Nombre de voyageurs par mois (SNCF) entre 2000 et 2019;*
- *Nombre de naissance féminine par mois au Cameroun entre 2008 et 2020;*
- *Nombre de voitures vendues par un garage, par trimestre entre 1995 et 1999;*
- *Nombres de contaminations journalière au Covid-19 en France entre Mars et Mai 2020*

Séries temporelles

But principal

Le but principal est le choix d'un modèle *(estimation)* raisonnable, qui permettra à partir des valeurs connues la prédiction des valeurs inobservables (comme les valeurs futures des séries temporelles, ou moins accessibles physiquement, couteuses, etc).

On veut à la fois:

- (a) enlever du bruit d'observation eventuelle;
- (b) *extrapoler* du connu au inconnu.

Séries temporelles

Domaine d'application

- Prospection et exploitation pétrolières et minières
- Traitement du signal
- Imagerie Médicale
- Océanographie, météorologie, hydrogeologie, environnement,$\cdots$
- Séries temporelles, appliquées en économie, finances, météo, médecine

séries temporelles

l'examen graphique de la série étudiée $(y_i, 1 \leq i \leq n)$ permet de dégager les composantes:

Composantes d'une serie temporelle

- La tendance $(f_i, 1 \leq i \leq n)$ représente l'évolution à long terme de la grandeur étudiée, et traduit l'aspect général de la série.
- Les variations saisonnières $(s_i, 1 \leq i \leq n)$. Mathématiquement, ce sont des fonctions périodiques, c'est-à-dire qu'il existe un entier p, appelé période, tel que:

$$s_i = s_{i+p}, \forall i \geq 1 \tag{1}$$

- *Cycles* $(c_i, 1 \leq i \leq n)$, qui regroupent des variations à période moins précise autour de la tendance, par exemple les phases économiques d'expansion et de récession.

séries temporelles

Composantes d'une serie temporelle

- Les fluctuations *irrégulières/résidues/bruit* $(e_i, 1 \leq i \leq n)$ sont des variations de faible intensité et de courte durée, et de nature aléatoire.
- Points de changement Ce sont des points où la série change complètement d'allure, par exemple de tendance.

Modélisation stochastique des séries temporelles

Catégories des series temporelles

On distingue 3 catégories de séries temporelles:

- Le modèle additif
- Le modèle multiplicatif
- modèle mixte

Rappel

un modèle additif sans saisonnalité, a une décomposition de la forme :

$$Y_t = m_t + \epsilon_t \tag{2}$$

où:

- m_t représente la "tendance" (intuitivement un "mouvement lisse à long terme"), qui sera la composante la plus importante dans la prévision.
- $\epsilon_t = Y_t - m_t$ sont les "résidus" qui restent après qu'on enlève la partie structurée m_t.

Stationnarité

Definition 2

Soit X un processus aléatoire indexé par $T = \mathbb{N}$ ou $\mathbb{Z}$. On dit que X est stationnaire (strict) si pour toute famille finie d'instants $t_1 ... t_r \in T$ et tout entier s, les lois jointes de $(X_{t_1}, \cdots, X_{t_r})$ et de $(X_{t_{1+s}}, \cdots, X_{t_{r+s}})$ sont les mêmes.

Fonction d'autocorrélation d'une série stationnaire

$$\rho_t = \frac{cov(y_t, y_{t-1})}{\sqrt{var(y_t)var(y_{t-1})}} = \frac{cov(y_t, y_{t-1})}{var(y_t)} = \frac{\gamma_t}{\gamma_0} \qquad (3)$$

Bruit blanc

Definition 3

Un bruit blanc $\{\epsilon_t\}$ est une suite de v.a. non corrélées (mais pas nécessairement indépendantes) de moyenne nulle et de variance constante σ_ϵ^2.

C'est donc une série faiblement stationnaire. On note $\{\epsilon_t\}$ tend vers $BB(0, \sigma_\epsilon^2)$.

Bruit blanc gaussien

Definition 4

Un bruit blanc gaussien $\{\epsilon_t\}$ est une suite de v.a. i.i.d $\mathcal{N}(0, \sigma_\epsilon^2)$..

Modèles de series

- **Série linéaire**
  Une série $\{y_t\}$ est dite linéaire si elle peut s'écrire :

$$y_t = \mu + \sum_{-\infty}^{+\infty} \psi_i \epsilon_{t-i} \tag{4}$$

- **Modèles autorégressifs**
  Considérons le modèle

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t, \epsilon_t \sim BB(0, \sigma_\epsilon^2). \tag{5}$$

où c et ϕ sont des constantes avec $\phi_p \neq 0$, appelé modèle autorégressif d'ordre p.
Pour p=1 le modèle

$$y_t = c + \phi y_{t-1} + \epsilon_t, \epsilon_t \sim BB(0, \sigma_\epsilon^2). \tag{6}$$

où c et ϕ sont des constantes, appelé modèle autorégressif d'ordre 1: $AR(1)$

Modèles de series

- **modèles moyennes mobiles**
  y_t est un processus moyenne mobile d'ordre q ($MA(q)$) si :

$$y_t = \mu + \epsilon + \theta_1 \epsilon_{t-1} + + \cdots , \theta_q \epsilon_{t-q}, \epsilon_t \sim BB(0, \sigma_\epsilon^2). \quad (7)$$

  avec $\theta_q \neq 0$

- **Processus** $ARMA(p, q)$
  yt obéit à un modèle $ARMA(p, q)$ s'il est stationnaire et vérifie :

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + + \cdots , \theta_q \epsilon_{t-q}. \quad (8)$$

  $\epsilon_t \sim BB(0, \sigma_\epsilon^2)$

Modèles de series

Les processus ARIMA sont des processus non stationnaires qui reviennent, après différenciation simple ou saisonnière, à des processus *ARMA*. C'est un modèle de séries stationnaires en différence qui comportent un tendance stochastique (et non déterministe) ou une saisonnalité stochastique (et non déterministe).

Modèles de series

Definition 5

Les modèles $ARIMA(p, d, q)$
On appelle processus $ARIMA(p, d, q)$ un processus non stationnaire X_t pour le quel le processus différencié d'ordre d, $Y_t = (1 - B)^d X_t, t \in \mathbb{Z}$ stationnaire, et vérifie une relation de récurrence $ARMA(p, q)$:

$$Y_t = \sum_{i=0}^{t-1} \psi^i \epsilon_{t-i} + \sum_{i=0}^{q} \theta_i \epsilon_{t-i}, \forall t \in \mathbb{Z} \tag{9}$$

où les ψ, θ sont des réels et ϵ_t est un bruit blanc de variance σ^2.

Modèles de series

Projet R

M. Henri LAUDE

Definition 6

Prévision des processus $ARIMA(p, d, q)$

Théorème 1

Dans le cas d'un modèle $ARIMA(p, d, q)$, la meilleure prévision linéaire au temps t est :

$$X_t\hat{(k)} = \mathbb{E}[X_{t+k}/F_t] = \sum_{i=1}^{p} \hat{\psi}_i X_t(\hat{k} - i) + \sum_{i=k}^{q} \theta_i \epsilon_{t+\hat{k}-i} \tag{10}$$

où les $\hat{\psi}_i$ sont les coefficients du polynôme $\psi(B)(1 - B)^d$

Application sur R

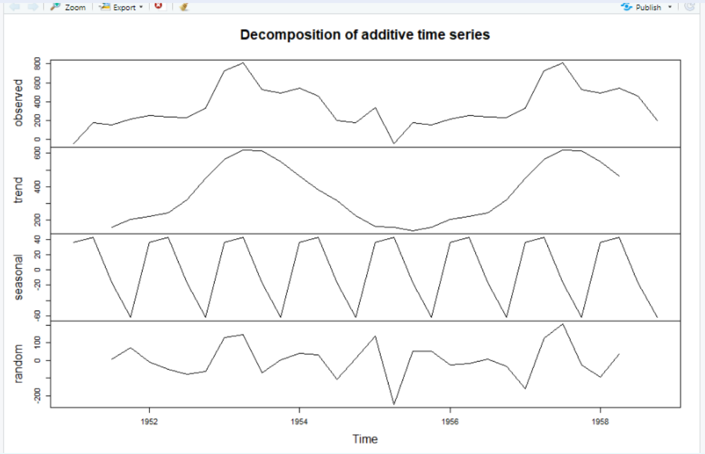
Definition 7

Modélisation

Ici on modélise une série temporelle dit "classique" qui est un exemple de série temporelle monter par moi même.

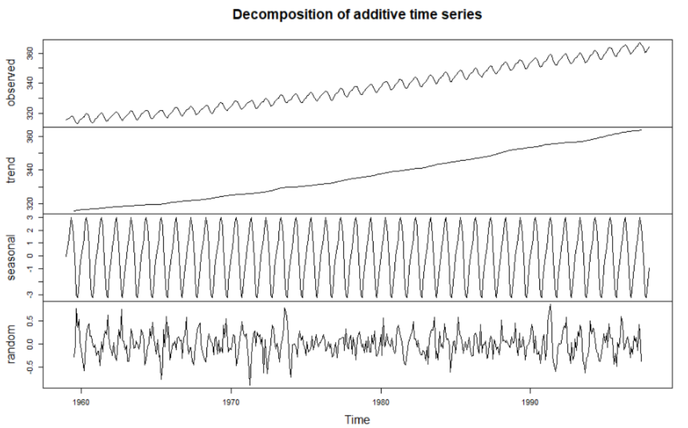
```
> x <- c(-50, 175, 149, 214, 247, 237, 225, 329, 729, 809, 530, 489, 540, 457, 195, 176, 337)
> x <- ts(x, start = c(1951, 1), end = c(1958, 4), frequency = 4)
> m <- decompose(x)
> plot(m)
> |
```

Application sur R

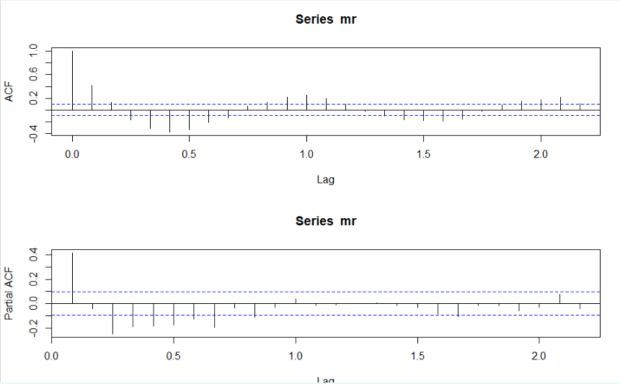


Application sur *R*


```
require(graphics)
m <- decompose(co2)
names(m)
plot(m)
mr<-window(m$random,start=c(1960,1),end=c(1996,4))
summary(mr)
layout(1:2)
acf(mr)
pacf(mr)
mrs<-as.ts(mr)
mr.ar <- ar(mrs, method = "mle")
mr.ar
```



Application sur *R*




```
Coefficients:
      1        2        3        4        5        6        7        8        9       10
 0.2377  -0.0509  -0.2358  -0.1941  -0.1961  -0.1688  -0.1043  -0.1962  -0.0110  -0.1151

Order selected 10  sigma^2 estimated as  0.04462
```

Application sur R

Modélisation: Modèle $ARIMA(1,1,2)$

Considérons le modèle $ARMA(1,2)$ suivant:

$$y_t = -0.9 - 0.8y_{t-1} + z_t - 0.3z_{t-1} + 0.6z_{t-2}, t = 1, \cdots, 200 \tag{11}$$

Avec z_t qui suit N$(0,4)$. On peut reécrire cette équation sous la forme:

$$y_t = -0.5 + \frac{1 - 0.3B + 0.6B^2}{1 + 0.8B} z_t, t = 1, \cdots, 200 \tag{12}$$

Le modèle donnée par l'équation (19) est un $ARIMA(1,1,2)$

Application sur *R*

Question : Comment simuler cette série et l'analyser ?

(1) Simulation de $ARMA(1,2)$ avec la fonction *arima.sim*

```
set.seed(121181)
yd.n = -0.5 + arima.sim(n = 200, list(ar = -0.8, ma = c(-0.3,
                                     + 0.6)), sd = 2, n.start = 50)
```

(2) Appliquer la fonction *diffinv* sur le $ARMA(1,2)$:

```
y2.int <- diffinv(yd.n)
```

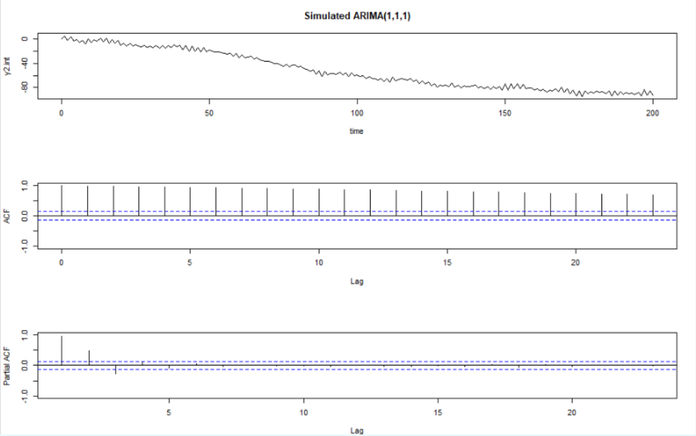
- L'analyse de l'*ACF* (la décroissante de manière non exponentielle) confirme la non stationnarité de la série.

```
op <- par(mfrow = c(3, 1))
plot(y2.int, type = "l", xlab = "time", main = "Simulated ARIMA(1,1,1)")
acf(y2.int, main = "", ylim = c(-1, 1))
pacf(y2.int, main = "", ylim = c(-1, 1))
par(op)
```

Application sur *R*

Projet R

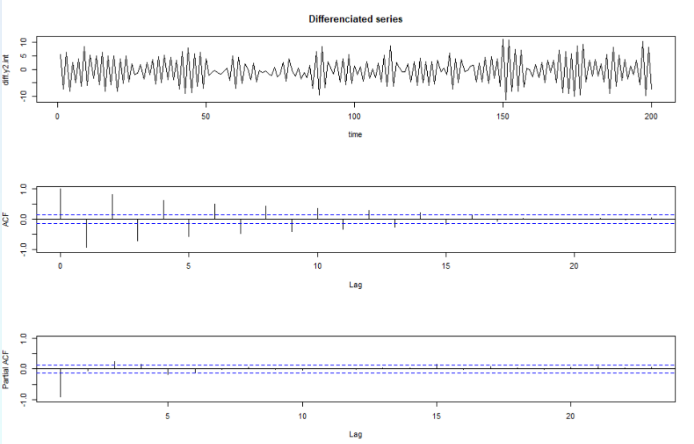
M. Henri LAUDE



Application sur *R*

- *ACF* et *PACF* de la série différenciée

```
diff.y2.int <- diff(y2.int)
op <- par(mfrow = c(3, 1))
plot(diff.y2.int, type = "l", xlab = "time", main = "Differenciated series")
acf(diff.y2.int, main = "", ylim = c(-1, 1))
pacf(diff.y2.int, main = "", ylim = c(-1, 1))
par(op)
```



(3) Modélisation $ARIMA(1, 1, 2)$

- utilisation la fonction *Arima* incluse dans le package *Forecast* pour modéliser le $ARIMA(1, 1, 2)$ de la serie initiale.

```
require(forecast)
```

```
install.packages("forecast")
require(forecast)
y2.fit = Arima(y2.int, order = c(1, 1, 2), include.drift = TRUE)
y2.fit
```

```
Series: y2.int
ARIMA(1,1,2) with drift

Coefficients:
          ar1      ma1     ma2    drift
      -0.8175  -0.2345  0.4635  -0.4608
s.e.   0.0475   0.0719  0.0743   0.0891

sigma^2 estimated as 3.544:  log likelihood=-409.5
AIC=828.99   AICc=829.3   BIC=845.48
>
```

NOUS JE VOUS REMERCIE POUR VOTRE AIMABLE ATTENTION