

# Séries chronologiques

Licence 3 MIASHS - Université de Bordeaux

Jérémie BIGOT

Polycopié rédigé à partir des notes de cours de Frédéric Proia

Année 2016-2017



# Avant-propos

**Remerciements** : ces notes de cours ont été essentiellement rédigées par Frédéric Proia qui a assuré cet enseignement en Licence MIASHS lors de son doctorat à l'Université de Bordeaux au sein de l'Institut de Mathématiques de Bordeaux. Suite à son départ à l'Université d'Angers en 2014, Frédéric Proia m'a communiqué l'ensemble de ses notes de cours que je me suis permis de regrouper sous la forme d'un unique polycopié. Je voudrais le remercier très chaleureusement pour son aide précieuse dans la préparation de cours.

Ce document est un support de cours pour une introduction à la modélisation de données temporelles au moyen des séries chronologiques, et à l'implémentation de ce type d'outils au travers du logiciel **R** pour l'analyse de données simulées ou réelles.

Pour de plus amples détails sur les méthodes de séries chronologiques et leur utilisation à l'aide du logiciel **R**, le lecteur intéressé pourra consulter les ouvrages suivants :

- Régis Bourbonnais, Michel Terraza. *Analyse des séries temporelles : applications à l'économie et à la gestion*. Dunod, 2004.
- Yves Aragon. *Séries temporelles avec R. Méthodes et cas*. Springer, Collection Pratique R, 1st edition, 2011.



# Table des matières

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Séries chronologiques et stationnarité</b>             | <b>7</b>  |
| 1.1      | Processus aléatoires . . . . .                            | 7         |
| 1.1.1    | Objectifs de la modélisation chronologique . . . . .      | 7         |
| 1.1.2    | Loi d'un processus aléatoire . . . . .                    | 8         |
| 1.2      | Processus du second ordre . . . . .                       | 8         |
| 1.2.1    | L'espace $L^2(\Omega, \mathcal{A}, \mathbb{P})$ . . . . . | 8         |
| 1.2.2    | Séries chronologiques . . . . .                           | 10        |
| 1.3      | Stationnarité . . . . .                                   | 10        |
| 1.3.1    | La stationnarité stricte . . . . .                        | 11        |
| 1.3.2    | La stationnarité du second ordre . . . . .                | 11        |
| 1.4      | Autocovariance et autocorrélation . . . . .               | 13        |
| 1.4.1    | La fonction d'autocovariance . . . . .                    | 13        |
| 1.4.2    | La fonction d'autocorrélation . . . . .                   | 13        |
| 1.4.3    | Un exemple détaillé . . . . .                             | 14        |
| <b>2</b> | <b>Le processus linéaire et la moyenne mobile</b>         | <b>17</b> |
| 2.1      | Deux opérateurs de chronologie . . . . .                  | 17        |
| 2.1.1    | L'opérateur retard . . . . .                              | 17        |
| 2.1.2    | L'opérateur différenciation . . . . .                     | 18        |
| 2.1.3    | Généralisation . . . . .                                  | 18        |
| 2.2      | Le processus linéaire . . . . .                           | 18        |
| 2.2.1    | Écriture . . . . .  | 19        |
| 2.2.2    | Stationnarité . . . . .                                   | 19        |
| 2.3      | Le processus moyenne mobile . . . . .                     | 20        |
| 2.3.1    | Écriture . . . . .  | 20        |
| 2.3.2    | Stationnarité . . . . .                                   | 21        |
| 2.3.3    | Estimation des paramètres . . . . .                       | 22        |
| 2.3.4    | Application à la prévision . . . . .                      | 24        |
| 2.4      | Quelques exemples simples . . . . .                       | 25        |
| 2.4.1    | Le processus MA(0) . . . . .                              | 25        |
| 2.4.2    | Le processus MA(1) . . . . .                              | 25        |
| 2.4.3    | Le processus MA(2) . . . . .                              | 26        |

|          |  |           |
|----------|--|-----------|
| <b>3</b> | <b>Le processus autorégressif</b>                      | <b>29</b> |
| 3.1      | Focus sur le processus AR(1) . . . . .                 | 30        |
| 3.1.1    | Écriture causale et stationnarité . . . . .            | 30        |
| 3.1.2    | Estimation des paramètres . . . . .                    | 31        |
| 3.1.3    | Modélisation et prévision . . . . .                    | 33        |
| 3.2      | Focus sur le processus AR( $p$ ) . . . . .             | 33        |
| 3.2.1    | Écriture causale et stationnarité . . . . .            | 34        |
| 3.2.2    | Estimation des paramètres . . . . .                    | 35        |
| 3.2.3    | Modélisation et prévision . . . . .                    | 38        |
| 3.3      | Quelques exemples simples . . . . .                    | 38        |
| 3.3.1    | Le processus AR(0) . . . . .                           | 38        |
| 3.3.2    | Le processus AR(1) . . . . .                           | 38        |
| 3.3.3    | Le processus AR(2) . . . . .                           | 39        |
| <b>4</b> | <b>Le processus ARMA et la non stationnarité</b>       | <b>41</b> |
| 4.1      | Le processus ARMA . . . . .                            | 41        |
| 4.1.1    | Stationnarité . . . . .                                | 42        |
| 4.1.2    | Un exemple détaillé : l'ARMA(1,1) . . . . .            | 43        |
| 4.1.3    | Causalité et inversibilité . . . . .                   | 45        |
| 4.1.4    | Estimation des paramètres . . . . .                    | 46        |
| 4.1.5    | Application à la prévision . . . . .                   | 47        |
| 4.2      | Introduction à la non stationnarité . . . . .          | 48        |
| 4.2.1    | Le processus ARIMA . . . . .                           | 48        |
| 4.2.2    | Détecter la stationnarité : le test ADF . . . . .      | 49        |
| 4.2.3    | Détecter la non stationnarité : le test KPSS . . . . . | 50        |
| 4.2.4    | En résumé... . . . .                                   | 50        |

# Chapitre 1

## Séries chronologiques et stationnarité

Nous introduisons dans ce premier chapitre quelques outils fondamentaux que nous utiliserons dans la suite du module pour modéliser certains phénomènes aléatoires évoluant avec le temps. Celui-ci contient donc essentiellement des définitions.

### 1.1 Processus aléatoires

**Définition 1.1.1** *On appelle « processus stochastique » toute famille  $(X_t)_{t \in T}$  de variables aléatoires définies sur un même espace de probabilité  $(\Omega, \mathcal{A}, \mathbb{P})$ , indexées par l'espace des temps  $T$  et à valeurs dans un même espace des états  $(E, \mathcal{E})$ .*

Nous rencontrerons le plus souvent des processus aléatoires à temps discret (avec  $T = \mathbb{N}$  ou  $T = \mathbb{Z}$ ) à valeurs réelles (avec  $E = \mathbb{R}$  et  $\mathcal{E} = \mathcal{B}(\mathbb{R})$ ). Un processus aléatoire  $(X_t)$  défini sur  $\mathbb{Z}$  possède un *passé asymptotique* ainsi qu'un *futur asymptotique*, et l'ensemble de ses valeurs est entièrement explicité par

$$X_{-\infty}, \dots, X_{-1}, X_0, X_1, \dots, X_{\infty}.$$

Au contraire, si un processus aléatoire  $(X_t)$  défini sur  $\mathbb{N}$  possède bien un futur asymptotique, il ne possède en revanche pas de passé asymptotique (qui est remplacé par des valeurs initiales). Il est caractérisé par

$$X_0, X_1, \dots, X_{\infty}.$$

De la même manière, un processus  $(X_t)$  défini sur  $\mathbb{N}^*$  possède un futur asymptotique et une valeur initiale  $X_1$  (et non plus  $X_0$ ).

#### 1.1.1 Objectifs de la modélisation chronologique

Une différence fondamentale avec les exemples considérés dans le module "modélisation statistique" est que, sauf quelques cas particuliers que nous pourrions étudier, le processus  $(X_t)_{t \in T}$  n'est pas formé de variables aléatoires indépendantes et identiquement distribuées. Au contraire, il existe généralement une corrélation forte entre

deux valeurs proches dans le temps, et le processus évolue en conséquence. Étant donné un processus  $(X_t)$ , nous nous fixons trois objectifs majeurs :

- Identification du type d'évolution du processus.
- Modélisation pertinente du processus et estimation des paramètres associés.
- Prédiction des valeurs futures du processus en fonction des observations passées.

### 1.1.2 Loi d'un processus aléatoire

La loi d'un échantillon indépendant et identiquement distribué est très facile à caractériser : il s'agit de sa loi parente. Dans le cas d'un processus aléatoire, le problème devient extrêmement compliqué voire parfois insoluble, car la présence de corrélations implique de connaître la loi jointe de toutes les valeurs du processus.

**Définition 1.1.2** Soit  $(X_t)_{t \in T}$  un processus aléatoire défini sur  $(\Omega, \mathcal{A}, \mathbb{P})$ . Sa loi est caractérisée par ses répartitions finies, c'est-à-dire par les lois de  $(X_t)_{t \in \mathcal{F}}$  pour toutes les parties finies  $\mathcal{F}$  de  $T$ .

En pratique,  $\mathcal{F}$  correspond à tous les vecteurs  $(t_1, \dots, t_n)$  avec  $t_1 < \dots < t_n$ , et ce pour tout  $n \in \mathbb{N}^*$ . Par exemple, si  $E = \mathbb{R}$  alors  $\mathcal{E} = \mathcal{B}(\mathbb{R})$  correspond à la tribu borélienne engendrée par les intervalles  $[a, b]$  de  $\mathbb{R}$ . Pour tout  $t \in T$ , la loi de la variable  $X_t$  est alors caractérisée par sa fonction de répartition,

$$F_{X_t}([a, b]) = \mathbb{P}(a \leq X_t \leq b).$$

Il s'ensuit que la loi du processus  $(X_t)$  ne pourra être connue que par l'intermédiaire de la fonction de répartition jointe,

$$\mathbb{P}(a_1 \leq X_{t_1} \leq b_1, \dots, a_n \leq X_{t_n} \leq b_n)$$

où  $a_1 < b_1, \dots, a_n < b_n$ , et ce pour tout  $n \in \mathbb{N}^*$ . À l'exception de cas particuliers (tels que les processus gaussiens ou les bruits blancs), cette loi reste difficile à calculer. C'est pourquoi l'on introduira par la suite le concept de stationnarité.

## 1.2 Processus du second ordre

D'une manière générale, nous parlerons de « processus du second ordre » lorsque le processus stochastique étudié prend ses valeurs dans  $L^2(\Omega, \mathcal{A}, \mathbb{P})$ .

### 1.2.1 L'espace $L^2(\Omega, \mathcal{A}, \mathbb{P})$

**Définition 1.2.1** On note  $L^2(\Omega, \mathcal{A}, \mathbb{P})$  l'espace des variables aléatoires réelles  $X$  définies sur  $(\Omega, \mathcal{A}, \mathbb{P})$  telles que

$$\mathbb{E}[X^2] < +\infty.$$

L'espace  $L^2(\Omega, \mathcal{A}, \mathbb{P})$  est un espace de Hilbert pour le produit scalaire défini, pour toutes variables  $X, Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ , par

$$\langle X, Y \rangle = \mathbb{E}[XY].$$

On a de plus  $\|X\|_2 = \sqrt{\langle X, X \rangle} = \sqrt{\mathbb{E}[X^2]}$ . À partir de l'espace  $L^2(\Omega, \mathcal{A}, \mathbb{P})$ , il est possible d'introduire formellement les notions de variance, de covariance et de corrélation qui sont des caractéristiques du second ordre.

**Définition 1.2.2** Soient  $X, Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ . La covariance de  $X$  et  $Y$  est définie par

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

En particulier, on a

$$\mathbb{V}(X) = \text{Cov}(X, X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Enfin, la corrélation entre  $X$  et  $Y$  est donnée par

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\mathbb{V}(X)}\sqrt{\mathbb{V}(Y)}}.$$

On dira que  $X$  et  $Y$  ne sont pas corrélées lorsque  $\text{Cov}(X, Y) = \text{Corr}(X, Y) = 0$ . De plus, si  $\mathbb{E}[X] = \mathbb{E}[Y] = 0$ , alors la non corrélation équivaut à l'orthogonalité. En effet, on a alors  $\langle X, Y \rangle = 0$  ce qui signifie que, relativement au produit scalaire de  $L^2(\Omega, \mathcal{A}, \mathbb{P})$ ,  $X$  et  $Y$  sont orthogonales.

**Proposition 1.2.1** Soient  $X, Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ . Si  $X$  et  $Y$  sont indépendantes, alors

$$\text{Cov}(X, Y) = \text{Corr}(X, Y) = 0.$$

Cependant, la réciproque est généralement fausse.

**Démonstration.** Si  $X$  et  $Y$  sont indépendantes, alors  $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$  et il est donc immédiat de voir que  $\text{Cov}(X, Y) = \text{Corr}(X, Y) = 0$ . Pour montrer que la réciproque est fausse, considérons un contre-exemple. Prenons  $X \sim \mathcal{N}(0, 1)$  et  $Y = \varepsilon X$  où  $\varepsilon \sim \mathcal{R}(1/2)$ , c'est-à-dire la loi de Rademacher caractérisée par

$$\mathbb{P}(\varepsilon = -1) = \frac{1}{2} \quad \text{et} \quad \mathbb{P}(\varepsilon = 1) = \frac{1}{2}.$$

On a bien sûr  $\mathbb{E}[X] = 0$ ,  $\mathbb{E}[X^2] = 1$  et  $\mathbb{E}[\varepsilon] = -1/2 + 1/2 = 0$ . De plus, supposons que  $\varepsilon$  est indépendante de  $X$ . Alors,

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[\varepsilon X^2] = \mathbb{E}[\varepsilon] \mathbb{E}[X^2] = 0$$

par indépendance de  $\varepsilon$  et de  $X$ . Et pourtant, il est clair que  $X$  et  $Y$  ne sont pas indépendantes puisque  $Y = \varepsilon X$ .  $\square$

Il est possible de montrer que la réciproque est vraie dans le cas où  $X$  et  $Y$  forment un vecteur gaussien. On a également, par l'inégalité de Cauchy-Schwarz,

$$\begin{aligned} |\text{Cov}(X, Y)| &= |\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]| \\ &\leq \sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2]} \sqrt{\mathbb{E}[(Y - \mathbb{E}[Y])^2]} = \sqrt{\mathbb{V}(X)} \sqrt{\mathbb{V}(Y)}. \end{aligned}$$

Cela nous permet d'établir que

$$-1 \leq \text{Corr}(X, Y) \leq 1.$$

## 1.2.2 Séries chronologiques

**Définition 1.2.3** On appelle « processus du second ordre » tout processus aléatoire  $(X_t)_{t \in T}$  à valeurs dans  $\mathbb{R}$  tel que, pour tout  $t \in T$ ,

$$X_t \in L^2(\Omega, \mathcal{A}, \mathbb{P}).$$

Sa moyenne  $m(\cdot)$  et sa covariance  $\Gamma(\cdot, \cdot)$  sont définies, pour tous  $t, s \in T$ , par

$$m(t) = \mathbb{E}[X_t] \quad \text{et} \quad \Gamma(t, s) = \text{Cov}(X_t, X_s).$$

La moyenne et la covariance d'un processus stochastique sont donc des fonctions du temps, puisque le processus évolue. En particulier,  $m(\cdot)$  est une fonction de  $T$  dans  $\mathbb{R}$  alors que  $\Gamma(\cdot, \cdot)$  est une fonction de  $T^2$  dans  $\mathbb{R}$ .

**Proposition 1.2.2** Soit  $(X_t)_{t \in T}$  un processus du second ordre. Alors, sa covariance est une fonction symétrique.

**Démonstration.** Cette démonstration est immédiate. En effet, pour tout  $t, s \in T$ ,

$$\Gamma(t, s) = \text{Cov}(X_t, X_s) = \mathbb{E}[X_t X_s] - \mathbb{E}[X_t] \mathbb{E}[X_s] = \text{Cov}(X_s, X_t) = \Gamma(s, t).$$

□

**Définition 1.2.4** On appelle « série chronologique » tout processus du second ordre  $(X_t)_{t \in T}$  indexé par l'espace des temps  $T = \mathbb{N}$  ou  $T = \mathbb{Z}$ .

Il peut aussi s'agir d'un espace des temps translaté, tel que  $\mathbb{N}^*$  si notre première observation se fait à  $t = 1$  et non à  $t = 0$ . De plus, puisqu'une série chronologique est par définition du second ordre, on a nécessairement, pour tout  $t \in T$ ,

$$\mathbb{E}[X_t^2] < +\infty.$$

## 1.3 Stationnarité

La notion de stationnarité caractérise la capacité d'un processus à se décorrélérer totalement de l'indice temporel. Ainsi la loi du processus, bien que restant souvent inconnue, sera bien plus aisée à manipuler par l'intermédiaire de ses propriétés d'espérance et de covariance.

### 1.3.1 La stationnarité stricte

**Définition 1.3.1** Une série chronologique  $(X_t)_{t \in T}$  est dite « strictement stationnaire » (ou « fortement stationnaire ») si, pour tout  $n \geq 1$  et tout vecteur  $(t_1, \dots, t_n) \in T^n$ , les vecteurs

$$(X_{t_1}, \dots, X_{t_n}) \quad \text{et} \quad (X_{t_1+k}, \dots, X_{t_n+k})$$

ont la même loi, et ce pour tout décalage temporel  $k \in \mathbb{Z}$ .

Cette définition illustre bien le fait que l'indice temporel ne joue plus aucun rôle dans le comportement de la série chronologique puisque, en considérant tous les vecteurs de même taille translatés dans le temps, la loi du processus reste inchangée. **La loi d'un processus strictement stationnaire ne dépend donc pas du temps.** En pratique, nous sommes confrontés au même problème que précédemment : il est nécessaire d'être en mesure d'évaluer la loi jointe du processus, ce qui peut s'avérer très difficile. C'est pourquoi l'on introduit généralement une stationnarité plus faible et moins contraignante.

### 1.3.2 La stationnarité du second ordre

**Définition 1.3.2** Une série chronologique  $(X_t)_{t \in T}$  est dite « stationnaire au second ordre » (ou « faiblement stationnaire », ou même simplement « stationnaire ») si sa moyenne  $m(\cdot)$  et sa covariance  $\Gamma(\cdot, \cdot)$  sont invariantes par translation dans le temps. Autrement dit, pour tout  $t, s \in T$  et tout décalage temporel  $k \in \mathbb{Z}$ , on a

1.  $\mathbb{E}[X_t] = m(t) = m$ ,
2.  $\text{Cov}(X_t, X_s) = \Gamma(t, s) = \Gamma(t + k, s + k)$ .

Cette définition implique que  $T = \mathbb{Z}$ , mais il n'est pas difficile de la généraliser à  $T = \mathbb{N}$ . Nous pouvons en outre en déduire que la variance du processus est constante. En effet,

$$\mathbb{V}(X_t) = \Gamma(t, t) = \Gamma(0, 0)$$

en choisissant  $k = -t$  comme décalage temporel. **Ainsi, l'espérance et la variance d'un processus stationnaire au second ordre sont constantes tandis que sa covariance ne dépend que du décalage temporel entre les deux valeurs considérées.** Comme corollaire immédiat, nous en déduisons que la covariance, initialement fonction de deux variables  $t$  et  $s$ , est, pour un processus stationnaire au second ordre, fonction que d'une seule variable. En effet, avec  $k = -s$ ,

$$\Gamma(t, s) = \Gamma(t - s, 0) = \Gamma(h, 0) = \gamma(h),$$

en posant  $h = t - s$ .

**Proposition 1.3.1** Une série chronologique  $(X_t)_{t \in T}$  strictement stationnaire est stationnaire au second ordre. Cependant, la réciproque est généralement fausse.

**Démonstration.** Si  $(X_t)_{t \in T}$  est strictement stationnaire, alors, selon la Définition 1.3.1 avec  $n = 1$ , toutes ses valeurs  $(X_t)$  ont la même loi qui ne dépend pas de  $t$ . Ainsi,  $\mathbb{E}[X_t]$  et  $\mathbb{V}(X_t)$  sont constantes, pour tout  $t \in T$ . De plus, selon la Définition 1.3.1 avec  $n = 2$ , tous les couples  $(X_t, X_s)$  ont la même loi que les couples  $(X_{t+k}, X_{s+k})$ , pour tous  $t, s \in T$  et tout décalage temporel  $k \in \mathbb{Z}$ . Il s'ensuit que  $\text{Cov}(X_t, X_s) = \text{Cov}(X_{t-s}, X_0)$  et la covariance ne dépend donc que de l'amplitude  $t - s$ . La série chronologique  $(X_t)_{t \in T}$  est donc stationnaire au second ordre. Pour montrer que la réciproque est généralement fautive, considérons un contre-exemple. Soit  $(X_t)_{t \in \mathbb{N}}$  le processus défini par

$$X_t = (-1)^t Z$$

où  $Z$  est une variable aléatoire non symétrique (la loi de  $Z$  n'est pas la même que la loi de  $-Z$ ), telle que  $\mathbb{E}[Z] = 0$  et  $\mathbb{E}[Z^2] = 1$ . Pour tout  $t \in \mathbb{N}$ , on a

$$\mathbb{E}[X_t] = (-1)^t \mathbb{E}[Z] = 0 \quad \text{et} \quad \mathbb{E}[X_t^2] = (-1)^{2t} \mathbb{E}[Z^2] = 1.$$

De plus, pour tous  $t, s \in \mathbb{N}$ ,

$$\text{Cov}(X_t, X_s) = (-1)^t (-1)^s \mathbb{E}[Z^2] = (-1)^{t+s} = (-1)^{t-s}$$

car, puisque  $(-1)^{-2s} = 1$ , il est clair que  $(-1)^{t+s} = (-1)^{t+s} (-1)^{-2s} = (-1)^{t-s}$ . Ainsi, l'espérance et la variance de  $X_t$  sont constantes et la covariance entre  $X_t$  et  $X_s$  ne dépend que de  $t - s$ , ce qui fait que la série chronologique  $(X_t)$  est stationnaire au second ordre, selon la Définition 1.3.2. Cependant, comme  $X_t = -X_{t-1} = \pm Z$  (selon la parité de  $t$ ), la loi de  $X_t$  et la loi de  $X_{t-1}$  sont différentes (puisque la loi de  $Z$  n'est pas symétrique). En conclusion,  $(X_t)$  n'est pas strictement stationnaire.  $\square$

Là encore, il est possible de montrer que la réciproque est vraie dans le cas d'un processus gaussien. Un vecteur gaussien étant entièrement déterminé par son espérance et sa covariance, on a alors l'équivalence entre stricte stationnarité et stationnarité du second ordre. L'exemple le plus trivial de série chronologique stationnaire est le bruit blanc.

**Définition 1.3.3** On appelle « bruit blanc » tout processus  $(X_t)_{t \in T}$  de variables aléatoires centrées, de même variance  $\sigma^2$  et non corrélées. Pour tous  $t, s \in T$ , on a

$$\mathbb{E}[X_t] = m(t) = 0 \quad \text{et} \quad \text{Cov}(X_t, X_s) = \Gamma(t, s) = \begin{cases} \sigma^2 & \text{si } t = s, \\ 0 & \text{si } t \neq s. \end{cases}$$

Le bruit blanc est ainsi stationnaire au second ordre (son espérance est constante et sa covariance ne dépend que du décalage temporel  $h = t - s$ ).

**Définition 1.3.4** On appelle « bruit blanc fort » tout processus  $(X_t)_{t \in T}$  de variables aléatoires indépendantes et identiquement distribuées, centrées et de variance  $\sigma^2$ .

Le bruit blanc fort est donc strictement stationnaire puisque ses variables sont indépendantes (donc de covariance nulle) et de même loi. En vertu de ce que nous avons vu précédemment, la non corrélation est équivalente à l'indépendance dans le cas gaussien. Cela signifie qu'un bruit blanc gaussien est nécessairement un bruit blanc fort.

## 1.4 Autocovariance et autocorrélation

La covariance d'une série chronologique stationnaire est fonction d'une seule variable :  $h = t - s$ , le décalage temporel entre les deux observations. Nous pouvons ainsi bâtir une *fonction d'autocovariance* et une *fonction d'autocorrélation* sur cette variable.

### 1.4.1 La fonction d'autocovariance

**Définition 1.4.1** Soit  $(X_t)_{t \in T}$  une série chronologique stationnaire. Sa « fonction d'autocovariance »  $\gamma(\cdot)$  est définie, pour tout  $h \in \mathbb{Z}$ , par

$$\gamma(h) = \Gamma(h, 0) = \text{Cov}(X_h, X_0).$$

Son abréviation usuelle est « ACV ». Puisque  $(X_t)$  est stationnaire, il est clair que pour tout  $t \in T$ , on a

$$\text{Cov}(X_{t+h}, X_t) = \dots = \text{Cov}(X_{h+1}, X_1) = \text{Cov}(X_h, X_0) = \gamma(h),$$

ce qui montre bien que l'autocovariance d'un processus stationnaire ne dépend que du décalage temporel  $h$  entre deux observations, et non de l'instant  $t$  de la mesure. Comme nous l'avons montré, la fonction d'autocovariance du bruit blanc est caractérisée par

$$\gamma(h) = \begin{cases} \sigma^2 & \text{si } h = 0, \\ 0 & \text{si } |h| \neq 0. \end{cases}$$

### 1.4.2 La fonction d'autocorrélation

**Définition 1.4.2** Soit  $(X_t)_{t \in T}$  une série chronologique stationnaire. Sa « fonction d'autocorrélation »  $\rho(\cdot)$  est définie, pour tout  $h \in \mathbb{Z}$ , par

$$\rho(h) = \text{Corr}(X_h, X_0) = \frac{\gamma(h)}{\gamma(0)}.$$

Son abréviation usuelle est « ACF ». Puisque  $(X_t)$  est stationnaire, il est clair que pour tout  $t \in T$ , on a

$$\text{Corr}(X_{t+h}, X_t) = \dots = \text{Corr}(X_{h+1}, X_1) = \text{Corr}(X_h, X_0) = \rho(h),$$

ce qui montre également que l'autocorrélation d'un processus stationnaire ne dépend que du décalage temporel  $h$  entre deux observations, et non de l'instant  $t$  de la mesure. De même, la fonction d'autocorrélation du bruit blanc est caractérisée par

$$\rho(h) = \begin{cases} 1 & \text{si } h = 0, \\ 0 & \text{si } |h| \neq 0. \end{cases}$$

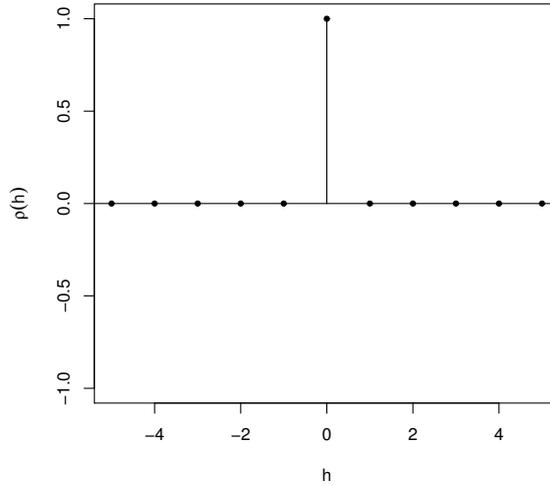
**Proposition 1.4.1** Soit  $(X_t)_{t \in T}$  une série chronologique stationnaire. Alors, sa fonction d'autocovariance et sa fonction d'autocorrélation sont paires.

**Démonstration.** Immédiate. En effet, pour tout  $h \in \mathbb{Z}$ ,

$$\gamma(-h) = \text{Cov}(X_{-h}, X_0) = \text{Cov}(X_0, X_h) = \gamma(h),$$

par stationnarité. Il en va de même pour  $\rho(-h)$ .  $\square$

On représente souvent l'ACF par un « autocorrélogramme », c'est-à-dire sous la forme d'un peigne représentant l'ampleur de chaque  $\rho(h)$ . Ci-dessous, nous avons représenté l'autocorrélogramme du bruit blanc, le plus simple que l'on puisse imaginer.



### 1.4.3 Un exemple détaillé

Nous allons conclure ce premier chapitre par un exemple détaillé. On considère le processus aléatoire  $(X_t)_{t \in \mathbb{Z}}$  défini, pour tout  $t \in \mathbb{Z}$ , par

$$X_t = \begin{cases} A + \varepsilon_t & \text{si } t \text{ est pair,} \\ B + \varepsilon_t & \text{si } t \text{ est impair,} \end{cases}$$

où  $A \sim \mathcal{N}(\mu_A, \tau_A^2)$ ,  $B \sim \mathcal{N}(\mu_B, \tau_B^2)$ ,  $\text{Cov}(A, B) = \delta$  et  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2$ , indépendant de  $A$  et de  $B$ . La première question que l'on se pose : le processus  $(X_t)$  forme-t-il une série chronologique stationnaire ? Tout d'abord,

$$\mathbb{V}(X_t) = \tau_A^2 + \sigma^2 < +\infty \quad \text{ou} \quad \mathbb{V}(X_t) = \tau_B^2 + \sigma^2 < +\infty$$

selon que  $t \in \mathbb{Z}$  est pair ou impair. Donc  $(X_t)$  est bien un processus du second ordre défini sur  $\mathbb{Z}$ , c'est donc une série chronologique. On voit également que

$$\mathbb{E}[X_t] = \mu_A \quad \text{ou} \quad \mathbb{E}[X_t] = \mu_B$$

selon la parité de  $t$ . Ainsi, si  $\mu_A \neq \mu_B$ ,  $(X_t)$  n'est pas stationnaire. Supposons donc que  $\mu_A = \mu_B = \mu$ . Alors,  $(X_t)$  n'est toujours pas stationnaire tant que  $\tau_A^2 \neq \tau_B^2$  puisque sa variance reste différente pour deux valeurs consécutives de  $t$ . Supposons donc que  $\tau_A^2 = \tau_B^2 = \tau^2$ . On a alors, pour tout  $t \in \mathbb{Z}$ ,

$$\mathbb{E}[X_t] = \mu \quad \text{et} \quad \mathbb{V}(X_t) = \tau^2 + \sigma^2.$$

Par ailleurs,

$$\text{Cov}(X_t, X_{t-1}) = \text{Cov}(A + \varepsilon_t, B + \varepsilon_{t-1}) = \text{Cov}(A, B) = \delta,$$

par indépendance entre  $(\varepsilon_t)$  et  $A, B$ . De plus, si  $t$  est pair,

$$\text{Cov}(X_t, X_{t-2}) = \text{Cov}(A + \varepsilon_t, A + \varepsilon_{t-2}) = \mathbb{V}(A) = \tau^2$$

alors que, si  $t$  est impair,

$$\text{Cov}(X_t, X_{t-2}) = \text{Cov}(B + \varepsilon_t, B + \varepsilon_{t-2}) = \mathbb{V}(B) = \tau^2.$$

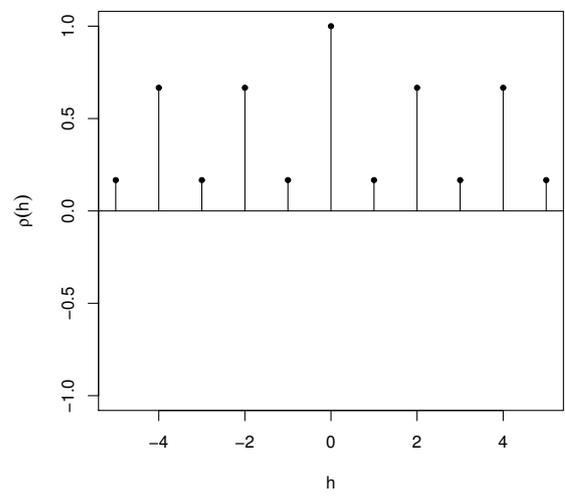
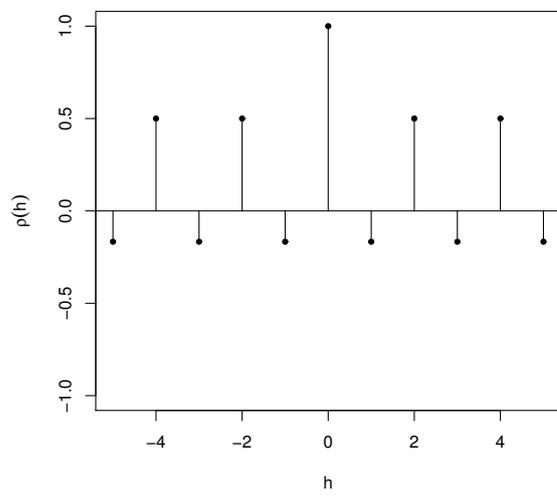
Il est facile de voir de la même manière que  $\text{Cov}(X_t, X_{t-h}) = \delta$  si  $h$  est impair, alors que  $\text{Cov}(X_t, X_{t-h}) = \tau^2$  si  $h$  est pair mais non nul. Ainsi, la covariance entre deux valeurs du processus ne dépend bien que du décalage temporel entre ces valeurs, et  $(X_t)$  est stationnaire. Sa fonction d'autocovariance est donnée, pour tout  $h \in \mathbb{Z}$ , par

$$\gamma(h) = \begin{cases} \tau^2 + \sigma^2 & \text{si } h = 0, \\ \tau^2 & \text{si } h \text{ est pair mais non nul,} \\ \delta & \text{si } h \text{ est impair.} \end{cases}$$

Sa fonction d'autocorrélation est donc donnée, pour tout  $h \in \mathbb{Z}$ , par

$$\rho(h) = \begin{cases} 1 & \text{si } h = 0, \\ \tau^2/(\tau^2 + \sigma^2) & \text{si } h \text{ est pair mais non nul,} \\ \delta/(\tau^2 + \sigma^2) & \text{si } h \text{ est impair.} \end{cases}$$

Si l'on suppose que  $\sigma^2 = 1$ ,  $\tau^2 = 1$  et  $\delta = -1/3$ , on obtient l'autocorrélogramme ci-dessous à gauche, alors que celui de droite correspond à  $\sigma^2 = 1$ ,  $\tau^2 = 2$  et  $\delta = 1/2$ .



# Chapitre 2

## Le processus linéaire et la moyenne mobile

Nous présentons tout d'abord dans ce chapitre deux opérateurs de chronologie. Par la suite, nous proposons quelques notions liées au processus linéaire du second ordre ainsi que ses propriétés de stationnarité. Nous en déduisons enfin celles du processus moyenne mobile qui est un cas particulier de processus linéaire.

### 2.1 Deux opérateurs de chronologie

Le traitement des processus chronologiques est souvent simplifié par l'intermédiaire des deux opérateurs que nous allons présenter.

#### 2.1.1 L'opérateur retard

**Définition 2.1.1** *On appelle « opérateur retard » (et l'on note  $L$ ) l'opérateur qui à une donnée chronologique associe sa valeur précédente dans le temps.*

Il peut s'agir de quantités déterministes. Par exemple, on a

$$Lt = t - 1 \quad \text{et} \quad Lf(t) = f(t - 1)$$

si  $t$  représente le temps sur l'espace discret  $\mathbb{Z}$ , et  $f$  une fonction déterministe du temps. L'opérateur s'applique aussi sur des quantités aléatoires. En particulier, si  $(X_t)$  est une série chronologique définie sur  $\mathbb{Z}$ , on a

$$LX_t = X_{t-1}.$$

La notation  $L$  provient de l'anglais *lag*, mais il est fréquent de rencontrer aussi la notation  $B$  (comme *backward*) ou, de manière plus francisée, la notation  $R$  (comme *retard*).

### 2.1.2 L'opérateur différenciation

**Définition 2.1.2** On appelle « opérateur différenciation » (et l'on note  $\Delta$ ) l'opérateur qui à une donnée chronologique associe la valeur de son dernier incrément dans le temps.

Reprenant les exemples précédents, on a

$$\Delta t = (1 - L)t = t - (t - 1) = 1 \quad \text{et} \quad \Delta f(t) = (1 - L)f(t) = f(t) - f(t - 1).$$

De même, pour une série chronologique  $(X_t)$  définie sur  $\mathbb{Z}$ , on a

$$\Delta X_t = (1 - L)X_t = X_t - X_{t-1}.$$

On vérifie facilement que l'on a

$$X_t = LX_t + \Delta X_t$$

et donc que l'on a l'équivalence  $\Delta \equiv 1 - L$  en termes d'opérateurs.

### 2.1.3 Généralisation

Nous venons de présenter les deux opérateurs à l'ordre 1 : lorsque le retard porte sur la valeur précédente. Cependant, on peut facilement les généraliser à l'ordre  $h$  quelconque. Ainsi, appliqué à la série chronologique  $(X_t)$ , pour tout  $h \geq 1$ , on note

$$L^h X_t = \underbrace{(L \circ L \circ \dots \circ L)}_{h \text{ fois}} X_t = L(L(\dots LX_t \dots)) = X_{t-h}.$$

De même,

$$\Delta^h X_t = (1 - L)^h X_t \quad \text{et} \quad \Delta_h X_t = (1 - L^h)X_t = X_t - X_{t-h}.$$

Par exemple lorsque  $h = 2$ , on obtient

$$L^2 X_t = X_{t-2}, \quad \Delta_2 X_t = (1 - L^2)X_t = X_t - X_{t-2}$$

et

$$\begin{aligned} \Delta^2 X_t &= (1 - L)^2 X_t = (1 - 2L + L^2)X_t \\ &= X_t - 2LX_t + L^2 X_t \\ &= X_t - 2X_{t-1} + X_{t-2} \\ &= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) = \Delta(\Delta X_t). \end{aligned}$$

## 2.2 Le processus linéaire

Le *processus linéaire* désigne de manière générique l'ensemble des processus formés à partir d'une combinaison linéaire de valeurs provenant d'un bruit blanc. C'est une classe très générale de processus et la plupart de ceux que nous serons amenés à considérer en sont des sous-classes. D'où l'intérêt d'en étudier ses propriétés, ce qui nous permettra par la suite de se ramener à des cas particuliers sans systématiquement avoir à refaire tous les calculs de covariances.

### 2.2.1 Écriture

**Définition 2.2.1** On dit que la série chronologique  $(X_t)$  définie sur  $\mathbb{Z}$  est un « processus linéaire » si elle est définie, pour tout  $t \in \mathbb{Z}$ , par

$$X_t = \sum_{k=-\infty}^{\infty} c_k \varepsilon_{t-k} \quad \text{avec} \quad \sum_{k=-\infty}^{\infty} c_k^2 < \infty$$

où  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2$  et  $(c_k)$  est une suite de nombres réels.

On dit encore que la suite  $(c_k)$  est dans  $\ell^2(\mathbb{Z})$ , l'espace des suites réelles de carré sommable sur  $\mathbb{Z}$ . Nous verrons par la suite que la stationnarité du processus repose sur cette condition. Cette écriture est donc **anticipative** : elle s'exprime en fonction du passé mais aussi à travers le futur du processus. On peut en effet décomposer, à l'instant présent  $t$ ,

$$\underbrace{X_t}_{\text{présent}} = \underbrace{\sum_{k=1}^{\infty} c_k \varepsilon_{t-k}}_{\text{passé}} + \underbrace{c_0 \varepsilon_t}_{\text{présent}} + \underbrace{\sum_{k=-\infty}^{-1} c_k \varepsilon_{t-k}}_{\text{futur}}.$$

### 2.2.2 Stationnarité

Tout d'abord, il est clair que, pour tout  $t \in \mathbb{Z}$ ,

$$\mathbb{E}[X_t] = \sum_{k=-\infty}^{\infty} c_k \mathbb{E}[\varepsilon_{t-k}] = 0$$

puisque  $(\varepsilon_t)$  est un bruit blanc. Par ailleurs, l'absence de corrélation dans  $(\varepsilon_t)$  entraîne que toutes ses autocovariances sont nulles (par définition du bruit blanc !) et donc que

$$\mathbb{V}(X_t) = \sum_{k=-\infty}^{\infty} c_k^2 \mathbb{V}(\varepsilon_{t-k}) = \sigma^2 \sum_{k=-\infty}^{\infty} c_k^2 < \infty,$$

d'où l'utilité de l'hypothèse  $(c_k) \in \ell^2(\mathbb{Z})$ . La variance étant finie,  $(X_t)$  est bien un processus du second ordre. De plus, ni son espérance ni sa variance ne dépendent du temps. Il nous reste à étudier ses autocovariances. À l'ordre 1,

$$\text{Cov}(X_t, X_{t-1}) = \text{Cov}\left(\sum_{k=-\infty}^{\infty} c_k \varepsilon_{t-k}, \sum_{i=-\infty}^{\infty} c_i \varepsilon_{t-1-i}\right) = \sum_{k=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} c_k c_{\ell-1} \text{Cov}(\varepsilon_{t-k}, \varepsilon_{t-1-\ell})$$

en posant  $\ell = i + 1$  dans la deuxième somme. Comme

$$\text{Cov}(\varepsilon_{t-k}, \varepsilon_{t-\ell}) = \begin{cases} \sigma^2 & \text{si } k = \ell \\ 0 & \text{si } k \neq \ell, \end{cases}$$

on voit que seuls les couples pour lesquels  $k = \ell$  n'engendreront pas de covariance nulle. En conséquence, pour chaque indice  $k$  de la première somme, il existe **un seul** indice  $\ell$  de la seconde somme (tel que  $\ell = k$ ) dont on doit tenir compte. Il vient

$$\mathbb{C}\text{ov}(X_t, X_{t-1}) = \gamma(1) = \sigma^2 \sum_{k=-\infty}^{\infty} c_k c_{k-1}.$$

De la même manière, en posant  $\ell = h + 1$ , on trouve pour tout  $h \geq 1$ ,

$$\mathbb{C}\text{ov}(X_t, X_{t-h}) = \sum_{k=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} c_k c_{\ell-h} \mathbb{C}\text{ov}(\varepsilon_{t-k}, \varepsilon_{t-\ell})$$

ce qui nous conduit directement à

$$\gamma(h) = \sigma^2 \sum_{k=-\infty}^{\infty} c_k c_{k-h}.$$

La fonction d'autocovariance ne dépendant que de  $h$ , cela nous confirme qu'un processus linéaire est stationnaire. Cette dernière étant de plus paire, nous avons pour tout  $h \in \mathbb{Z}$ ,

$$\gamma(h) = \sigma^2 \sum_{k=-\infty}^{\infty} c_k c_{k-|h|} \quad \text{et} \quad \rho(h) = \frac{\sum_{k=-\infty}^{\infty} c_k c_{k-|h|}}{\sum_{k=-\infty}^{\infty} c_k^2}$$

à partir de quoi l'on retrouve la variance  $\gamma(0)$  et le fait que  $\rho(0) = 1$ .

## 2.3 Le processus moyenne mobile

Un cas particulier de processus linéaire que l'on rencontre très souvent en pratique est la *moyenne mobile*. Cette dernière est communément utilisée pour lisser des séries chronologiques (faire disparaître les perturbations locales), ou pour modéliser les séries à très courte mémoire. Nous allons la définir puis travailler à l'estimation de ses paramètres et à son application à la prévision.

### 2.3.1 Écriture

**Définition 2.3.1** *On dit que la série chronologique  $(X_t)$  définie sur  $\mathbb{Z}$  est un « processus moyenne mobile d'ordre  $q$  » si elle est définie, pour tout  $t \in \mathbb{Z}$ , par*

$$X_t = \mu + \varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_q \varepsilon_{t-q}$$

où  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2$  et  $b_q \neq 0$ .

La notation usuelle des moyennes mobiles est MA, de l'anglais *moving average*. Pour signifier que le processus  $(X_t)$  est engendré par une moyenne mobile d'ordre  $q$ , nous noterons

$$(X_t) \sim \text{MA}(q).$$

**Remarque 2.3.1** *Attention aux notations ! Les parenthèses autour de  $(X_t)$  sont ici indispensables : elles signifient que l'on considère le processus, alors que  $X_t$  sans parenthèses désigne la variable aléatoire extraite du processus à l'instant  $t$ .*

En général, l'écriture  $\text{MA}(q)$  se condense à l'aide des opérateurs chronologiques. Soit le polynôme défini, pour tout  $z \in \mathbb{C}$ , par  $\mathcal{B}(z) = 1 + b_1 z + \dots + b_q z^q$ . Alors, on écrira

$$X_t = \mu + \mathcal{B}(L) \varepsilon_t.$$

En effet, l'application de l'opérateur retard conduit à

$$\mathcal{B}(L) \varepsilon_t = (1 + b_1 L + \dots + b_q L^q) \varepsilon_t = \varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_q \varepsilon_{t-q}.$$

Nous verrons dans un chapitre ultérieur que les racines du polynôme  $\mathcal{B}$  jouent un rôle dans le comportement asymptotique du processus.

### 2.3.2 Stationnarité

Pour l'étude de la stationnarité de  $(X_t)$ , nous allons tout d'abord remarquer que la moyenne mobile n'est qu'un cas particulier de processus linéaire. En effet, si nous choisissons  $c_{-\infty}, \dots, c_{-1} = 0$ ,  $c_0 = 1$ ,  $c_1, \dots, c_q = b_1, \dots, b_q$  et  $c_{q+1}, \dots, c_\infty = 0$ , alors il est clair que l'écriture  $\text{MA}(q)$  est identique à celle proposée dans la Définition 2.2.1. De plus, suite à cette identification des paramètres, on vérifie que

$$\sum_{k=-\infty}^{\infty} c_k^2 = 1 + \sum_{k=1}^q b_k^2 < \infty,$$

ce qui confirme que la moyenne mobile est bien un processus du second ordre. Il s'ensuit, puisque nous l'avons montré dans le cas du processus linéaire, que **la moyenne mobile est un processus stationnaire**. Ses caractéristiques sont alors déduites par identification. En particulier, pour tout  $t \in \mathbb{Z}$ ,

$$\mathbb{E}[X_t] = \mu, \quad \mathbb{V}(X_t) = \sigma^2 \left( 1 + \sum_{k=1}^q b_k^2 \right)$$

et, pour tout  $|h| \leq q$ ,

$$\gamma(h) = \sigma^2 \left( b_{|h|} + \sum_{k=1}^{q-|h|} b_k b_{k+|h|} \right) \quad \text{et} \quad \rho(h) = \frac{b_{|h|} + \sum_{k=1}^{q-|h|} b_k b_{k+|h|}}{1 + \sum_{k=1}^q b_k^2}.$$

Ces expressions correspondent exactement à celles établies pour le processus linéaire, en identifiant  $c_{-\infty}, \dots, c_{-1} = 0$ ,  $c_0 = 1$ ,  $c_1, \dots, c_q = b_1, \dots, b_q$  et  $c_{q+1}, \dots, c_\infty = 0$ . Par ailleurs, si  $|h| > q$ , alors on voit que

$$\gamma(h) = \rho(h) = 0.$$

Ce résultat est très intuitif dans la mesure où le processus  $MA(q)$  engendre  $q$  retards. Deux valeurs distantes d'un décalage supérieur à  $q$  ne sont donc plus corrélées. En particulier pour  $h = q + 1$ ,

$$\begin{aligned} \text{Cov}(X_t, X_{t-q-1}) &= \text{Cov}(\varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_q \varepsilon_{t-q}, \varepsilon_{t-q-1} + b_1 \varepsilon_{t-q-2} + \dots + b_q \varepsilon_{t-2q-1}) \\ &= 0 \end{aligned}$$

et il en va bien sûr de même pour tous les décalages  $|h|$  supérieurs à  $q$ .

**Remarque 2.3.2** *Pour être tout à fait rigoureux, l'intercept n'apparaissant pas dans la définition du processus linéaire, il faudrait construire le processus  $(Y_t)$  donné, pour tout  $t \in \mathbb{Z}$ , par*

$$Y_t = X_t - \mu$$

*et identifier le processus  $(Y_t)$  avec un processus linéaire. Cela ne change bien évidemment absolument rien dans notre contexte, puisque  $\mu$  n'est qu'une constante qui ne dépend pas du temps et n'intervient pas dans les calculs de covariances.*

L'ensemble de ces résultats nous conduit au résumé suivant.

**Proposition 2.3.1** *Soit  $(X_t)$  un processus  $MA(q)$  défini sur  $\mathbb{Z}$ . Alors,  $(X_t)$  est stationnaire et ses fonctions d'autocovariance et d'autocorrélation s'annulent dès que  $|h| > q$ .*

**Démonstration.** La démonstration a été faite dans la Section 2.2. □

La principale utilité de ce résultat est qu'un simple coup d'oeil à un autocorrélogramme peut nous conduire à considérer une modélisation chronologique  $MA(q)$ , lorsque les autocorrélations s'annulent à partir d'un certain rang  $q$ . Ainsi, les deux exemples ci-dessous (qui seront étudiés en TD) peuvent faire l'objet d'une modélisation  $MA(1)$  et  $MA(3)$ , respectivement.

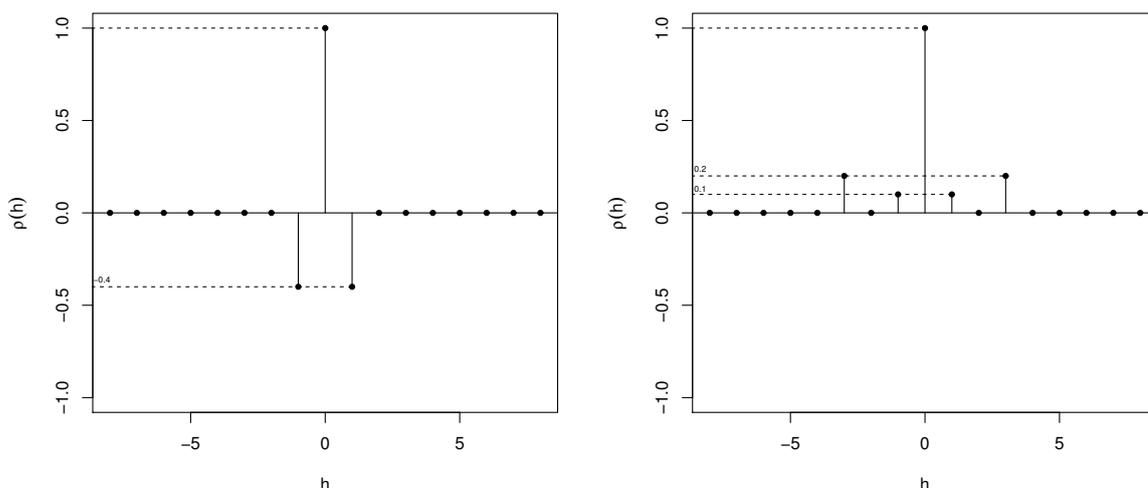
### 2.3.3 Estimation des paramètres

L'estimation des paramètres d'un processus passe tout d'abord par l'observation d'une *trajectoire* de ce processus, c'est-à-dire d'une réalisation de  $(X_t)$  sur un intervalle des temps fini,  $t \in \{1, \dots, n\}$  par exemple. L'expression du modèle  $MA$  fait intervenir un obstacle majeur en vue de l'estimation de ses paramètres  $\mu, b_1, \dots, b_q, \sigma^2$  : il n'est pas possible d'exprimer le bruit du modèle en fonction de variables observables. En effet, si dans un modèle de régression linéaire standard il est possible d'écrire

$$\varepsilon_t = Y_t - \beta_0 - \beta_1 X_t$$

puis d'élever cette expression au carré avant de la minimiser par rapport aux paramètres, en revanche il est clair que cette opération est irréalisable dans le cadre d'un modèle  $MA$ . Ainsi, il est certes possible d'écrire

$$\varepsilon_t = X_t - \mu - b_1 \varepsilon_{t-1} - \dots - b_q \varepsilon_{t-q}$$



mais les variables intervenant à droite de l'égalité, à l'exception de  $X_t$ , ne sont pas observables. Il s'ensuit qu'un modèle MA **ne peut pas être estimé par moindres carrés**. Pour mieux comprendre les choses, considérons l'exemple simple du modèle MA(1) centré ( $\mu = 0$ ) observé sur  $t \in \{1, \dots, n\}$ , donné par

$$X_t = \varepsilon_t + b_1 \varepsilon_{t-1}$$

avec par convention  $\varepsilon_0 = 0$ . Le raisonnement par moindres carrés standard nous conduirait au choix de l'estimateur

$$\hat{b}_1 = \frac{\sum_{t=1}^n \varepsilon_{t-1} X_t}{\sum_{t=1}^n \varepsilon_{t-1}^2}$$

mais **ce raisonnement est faux!** En effet, sur un jeu de données réelles, cette quantité ne pourra pas être évaluée puisque le processus  $(\varepsilon_t)$  n'est pas observable, contrairement aux processus  $(X_t)$  et  $(Y_t)$  dans le cadre de la régression linéaire classique. C'est pourquoi l'on utilise généralement le maximum de vraisemblance comme méthode d'estimation.

**Proposition 2.3.2** *Supposons que le polynôme  $\mathcal{B}$  possède toutes ses racines à l'extérieur du cercle unité, c'est-à-dire que pour tout  $z \in \mathbb{C}$  tel que  $|z| \leq 1$ ,  $\mathcal{B}(z) \neq 0$ . Soient  $\hat{\mu}, \hat{b}_1, \dots, \hat{b}_q, \hat{\sigma}^2$  les estimateurs du maximum de vraisemblance dans le modèle MA( $q$ ) donné, pour tout  $t \in \{q+1, \dots, n\}$ , par*

$$X_t = \mu + \mathcal{B}(L) \varepsilon_t = \mu + \varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_q \varepsilon_{t-q}$$

où  $b_q \neq 0$  et  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2$ . Alors, le vecteur  $(\hat{\mu}, \hat{b}_1, \dots, \hat{b}_q, \hat{\sigma}^2)$  est un estimateur consistant de  $(\mu, b_1, \dots, b_q, \sigma^2)$ .

*Démonstration.* Admise. □

Dans la pratique, cela passe par l'approximation de la vraisemblance du modèle par une loi normale. On obtient une fonction de vraisemblance très complexe et non linéaire (en raison des autocorrélations et donc du fait que les variables  $(X_t)$  ne sont pas indépendantes), que les logiciels de calcul (R, S+, Matlab, etc.) maximisent numériquement. **Les estimateurs du maximum de vraisemblance des paramètres d'un modèle MA n'ont donc pas d'expression explicite, mais ils sont approximables par les logiciels de calcul.** Notons pour conclure que l'intercept  $\mu$  peut être estimé indépendamment des autres paramètres par la moyenne empirique d'un échantillon observé,

$$\hat{\mu} = \bar{X}_n = \frac{1}{n} \sum_{t=1}^n X_t.$$

Nous en verrons un exemple en TD.

### 2.3.4 Application à la prévision

Une fois le modèle correctement posé (par exemple en étudiant un autocorrélogramme) et les paramètres estimés (à l'aide d'un logiciel), la chronologie du modèle nous invite à nous intéresser à la problématique de la prévision : nous avons observé le processus jusqu'à l'instant  $n$ , que pouvons-nous dire sur la valeur la plus probable à l'instant  $n + 1$  ? Là encore nous sommes confrontés au caractère non observable du processus  $(\varepsilon_t)$ , et nous allons devoir passer par une étape d'évaluation des résidus de modélisation  $(\tilde{\varepsilon}_t)$ , qui seront eux observables. Il existe plusieurs méthodes de prévision à partir des processus MA, mais nous en proposerons seulement une ici, la plus intuitive. L'algorithme est le suivant.

1. On considère une série chronologique  $(X_t)$  observée sur  $t \in \{1, \dots, n\}$ , modélisée par un modèle MA( $q$ ) et dont les paramètres  $\mu, b_1, \dots, b_q, \sigma^2$  ont été estimés par maximum de vraisemblance à l'aide d'un logiciel de calcul.
2. On initialise les  $q$  premières valeurs du processus résiduel  $(\tilde{\varepsilon}_t)$  de manière arbitraire, de sorte que  $\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_q \in L^2(\Omega, \mathcal{A}, \mathbb{P})$  pour que le processus reste du second ordre.
3. Pour toutes les valeurs observées du processus situées dans l'intervalle de temps  $t \in \{q + 1, \dots, n\}$ ,

(a) on reconstruit

$$\tilde{X}_t = \hat{\mu} + \hat{b}_1 \tilde{\varepsilon}_{t-1} + \dots + \hat{b}_q \tilde{\varepsilon}_{t-q},$$

(b) puis on évalue le résidu de modélisation correspondant,

$$\tilde{\varepsilon}_t = X_t - \tilde{X}_t = X_t - \hat{\mu} - \hat{b}_1 \tilde{\varepsilon}_{t-1} - \dots - \hat{b}_q \tilde{\varepsilon}_{t-q}.$$

4. Une fois le processus reconstruit  $(\tilde{X}_t)$  et les résidus de modélisation  $(\tilde{\varepsilon}_t)$  évalués jusqu'au dernier instant connu  $n$ , on prédit le premier instant inconnu par

$$\tilde{X}_{n+1} = \hat{\mu} + \hat{b}_1 \tilde{\varepsilon}_n + \dots + \hat{b}_q \tilde{\varepsilon}_{n-q+1}.$$

## 2.4 Quelques exemples simples

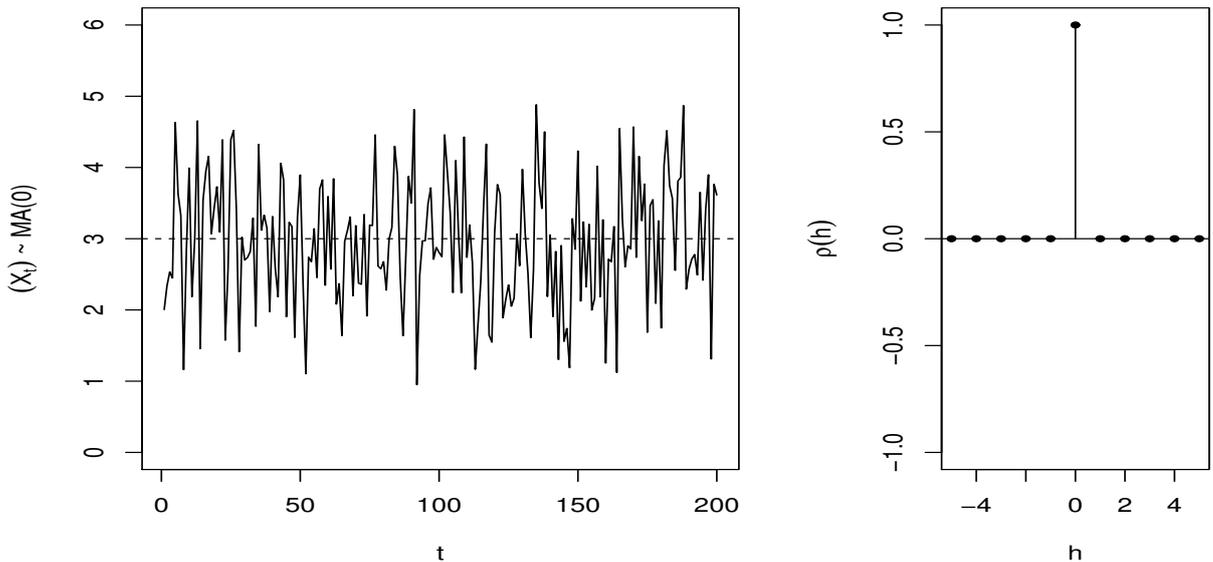
Pour conclure ce chapitre, considérons les 3 processus MA les plus simples. Nous en illustrons une trajectoire ainsi que l'autocorrélogramme associé.

### 2.4.1 Le processus MA(0)

Pour tout  $t \in \mathbb{Z}$ , le processus MA(0) est engendré par

$$X_t = \mu + \varepsilon_t$$

où  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2$ . Ainsi,  $(X_t)$  forme ici un simple bruit blanc décentré. L'exemple ci-dessous est simulé avec  $\mu = 3$ ,  $(\varepsilon_t) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$  et  $n = 200$ . Le



processus MA(0) est donc stationnaire et caractérisé par  $\mathbb{E}[X_t] = \mu$ ,

$$\gamma(h) = \begin{cases} \sigma^2 & \text{si } h = 0 \\ 0 & \text{si } h \neq 0 \end{cases} \quad \text{et} \quad \rho(h) = \begin{cases} 1 & \text{si } h = 0 \\ 0 & \text{si } h \neq 0. \end{cases}$$

### 2.4.2 Le processus MA(1)

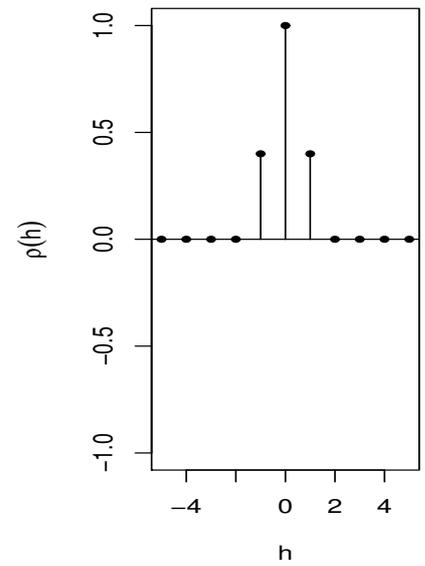
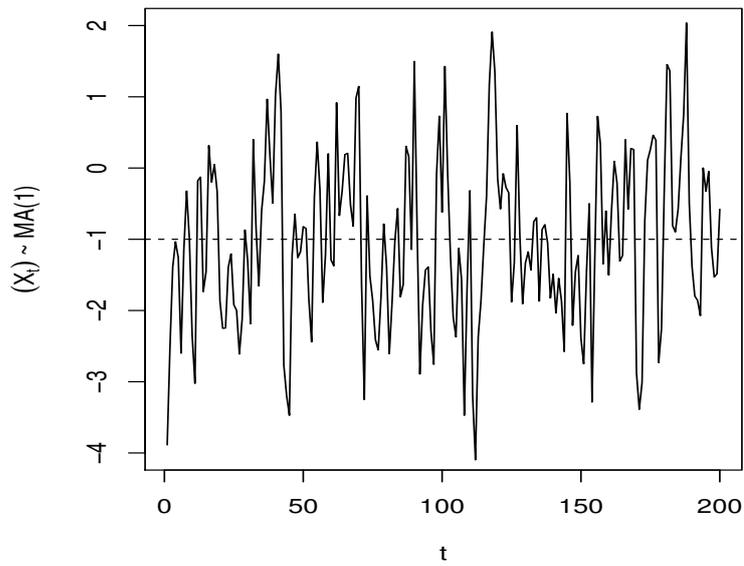
Pour tout  $t \in \mathbb{Z}$ , le processus MA(1) est engendré par

$$X_t = \mu + \varepsilon_t + b_1 \varepsilon_{t-1}$$

où  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2$ .

L'exemple ci-dessous est simulé avec  $\mu = -1$ ,  $b_1 = 0.5$ ,  $(\varepsilon_t) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$  et  $n = 200$ . Le processus MA(1) est donc stationnaire et caractérisé par  $\mathbb{E}[X_t] = \mu$ ,

$$\gamma(h) = \begin{cases} \sigma^2(1 + b_1^2) & \text{si } h = 0 \\ \sigma^2 b_1 & \text{si } |h| = 1 \\ 0 & \text{si } |h| > 1 \end{cases} \quad \text{et} \quad \rho(h) = \begin{cases} 1 & \text{si } h = 0 \\ \frac{b_1}{1+b_1^2} & \text{si } |h| = 1 \\ 0 & \text{si } |h| > 1. \end{cases}$$

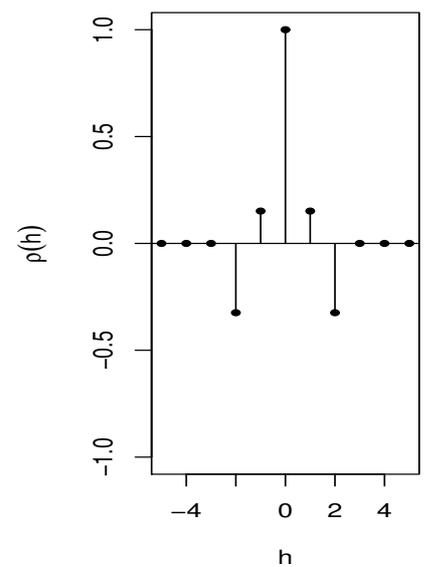
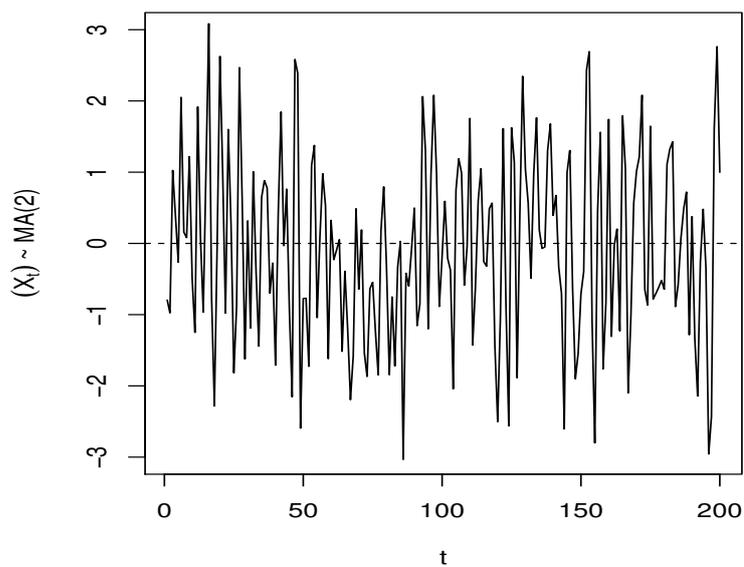


### 2.4.3 Le processus MA(2)

Pour tout  $t \in \mathbb{Z}$ , le processus MA(2) est engendré par

$$X_t = \mu + \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2}$$

où  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2$ . L'exemple ci-dessous est simulé avec  $\mu = 0$ ,  $b_1 = 0.7$ ,  $b_2 = -0.6$ ,  $(\varepsilon_t) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$  et  $n = 200$ .



Le processus MA(2) est donc stationnaire et caractérisé par  $\mathbb{E}[X_t] = \mu$ ,

$$\gamma(h) = \begin{cases} \sigma^2(1 + b_1^2 + b_2^2) & \text{si } h = 0 \\ \sigma^2(b_1 + b_1 b_2) & \text{si } |h| = 1 \\ \sigma^2 b_2 & \text{si } |h| = 2 \\ 0 & \text{si } |h| > 2 \end{cases} \quad \text{et} \quad \rho(h) = \begin{cases} 1 & \text{si } h = 0 \\ \frac{b_1 + b_1 b_2}{1 + b_1^2 + b_2^2} & \text{si } |h| = 1 \\ \frac{b_2}{1 + b_1^2 + b_2^2} & \text{si } |h| = 2 \\ 0 & \text{si } |h| > 2. \end{cases}$$



# Chapitre 3

## Le processus autorégressif

Nous introduisons dans ce chapitre le processus autorégressif, peut-être le plus utilisé en pratique pour modéliser les données chronologiques (température, indice boursier, cycle des marées, ventes, etc.) L'idée est de régresser le phénomène aléatoire directement sur son passé, de manière linéaire. Nous en étudions les propriétés de stationnarité et de causalité, nous en estimons les paramètres de deux manières différentes, puis nous abordons la problématique de la prédiction.

**Définition 3.0.1** *On dit que la série chronologique  $(X_t)$  définie sur  $\mathbb{Z}$  est un « processus autorégressif d'ordre  $p$  » si elle est définie, pour tout  $t \in \mathbb{Z}$ , par*

$$X_t = \mu + a_1 X_{t-1} + \dots + a_p X_{t-p} + \varepsilon_t$$

où  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2$  et  $a_p \neq 0$ .

La notation usuelle des processus autorégressifs est AR, de l'anglais *autoregressive*. Pour signifier que  $(X_t)$  est engendré par un processus autorégressif d'ordre  $p$ , nous noterons

$$(X_t) \sim \text{AR}(p).$$

**Remarque 3.0.1** *Attention aux notations ! Les parenthèses autour de  $(X_t)$  sont ici indispensables : elles signifient que l'on considère le processus, alors que  $X_t$  sans parenthèses désigne la variable aléatoire extraite du processus à l'instant  $t$ .*

En général, l'écriture  $\text{AR}(p)$  se condense à l'aide des opérateurs chronologiques. Soit le polynôme défini, pour tout  $z \in \mathbb{C}$ , par  $\mathcal{A}(z) = 1 - a_1 z - \dots - a_p z^p$ . Alors, on écrira

$$\mathcal{A}(L)X_t = \mu + \varepsilon_t.$$

En effet, l'application de l'opérateur retard conduit à

$$\mathcal{A}(L)X_t = (1 - a_1 L - \dots - a_p L^p) X_t = X_t - a_1 X_{t-1} - \dots - a_p X_{t-p}.$$

Nous verrons que là encore, comme pour les processus MA, les racines du polynôme  $\mathcal{A}$  jouent un rôle dans le comportement asymptotique du processus.

### 3.1 Focus sur le processus AR(1)

Un cas particulier d'intérêt pratique considérable est le processus autorégressif du premier ordre, pour  $p = 1$ . Nous allons le définir, étudier ses propriétés de stationnarité puis travailler à l'estimation de ses paramètres.

#### 3.1.1 Écriture causale et stationnarité

Pour tout  $t \in \mathbb{Z}$ , considérons le processus engendré par

$$X_t = \mu + \theta X_{t-1} + \varepsilon_t$$

où  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2$ . Pour un AR(1), on note souvent le paramètre  $\theta$  plutôt que  $a_1$ . Chercher l'écriture causale de  $(X_t)$  revient à exprimer  $X_t$  à l'instant  $t$  en fonction de son présent et de son passé seuls. Ici, on a clairement

$$\begin{aligned} X_t &= \mu + \theta X_{t-1} + \varepsilon_t \\ &= (\mu + \theta\mu) + \theta^2 X_{t-2} + \theta \varepsilon_{t-1} + \varepsilon_t \\ &= (\mu + \theta\mu + \theta^2\mu) + \theta^3 X_{t-3} + \theta^2 \varepsilon_{t-2} + \theta \varepsilon_{t-1} + \varepsilon_t \\ &= \dots \\ &= \mu \sum_{k=0}^{\infty} \theta^k + \sum_{k=0}^{\infty} \theta^k \varepsilon_{t-k} = \frac{\mu}{1-\theta} + \sum_{k=1}^{\infty} \theta^k \varepsilon_{t-k} + \varepsilon_t. \end{aligned}$$

Ce raisonnement est bien sûr valable si et seulement si  $|\theta| < 1$ , ce qui permet en particulier de rendre le passé asymptotique " $\theta^\infty X_{-\infty}$ " (notez les guillemets!) négligeable. Ainsi, le processus AR(1) s'exprime comme un processus linéaire dès que  $|\theta| < 1$ . Par identification avec l'expression donnée dans le Chapitre II, on a  $c_{-\infty} = \dots = c_{-1} = 0$ ,  $c_0 = 1$  et  $c_k = \theta^k$  pour  $k \in \{1, 2, \dots\}$ . **Il s'ensuit que le processus AR(1) défini sur  $\mathbb{Z}$  est stationnaire dès que son paramètre est en module strictement inférieur à 1.** De l'expression obtenue ci-dessus, il vient immédiatement, pour tout  $t \in \mathbb{Z}$ ,

$$\mathbb{E}[X_t] = \frac{\mu}{1-\theta}.$$

De plus, si nous reprenons la valeurs de l'ACV et de l'ACF des processus linéaires, on obtient par identification

$$\mathbb{V}(X_t) = \sigma^2 \sum_{k=0}^{\infty} c_k^2 = \sigma^2 \sum_{k=0}^{\infty} \theta^{2k} = \frac{\sigma^2}{1-\theta^2}.$$

Enfin, pour  $h \geq 1$ ,

$$\text{Cov}(X_t, X_{t-h}) = \sigma^2 \sum_{k=h}^{\infty} c_k c_{k-h} = \sigma^2 \sum_{k=h}^{\infty} \theta^k \theta^{k-h} = \sigma^2 \sum_{i=0}^{\infty} \theta^{i+h} \theta^i = \frac{\sigma^2 \theta^h}{1-\theta^2}$$

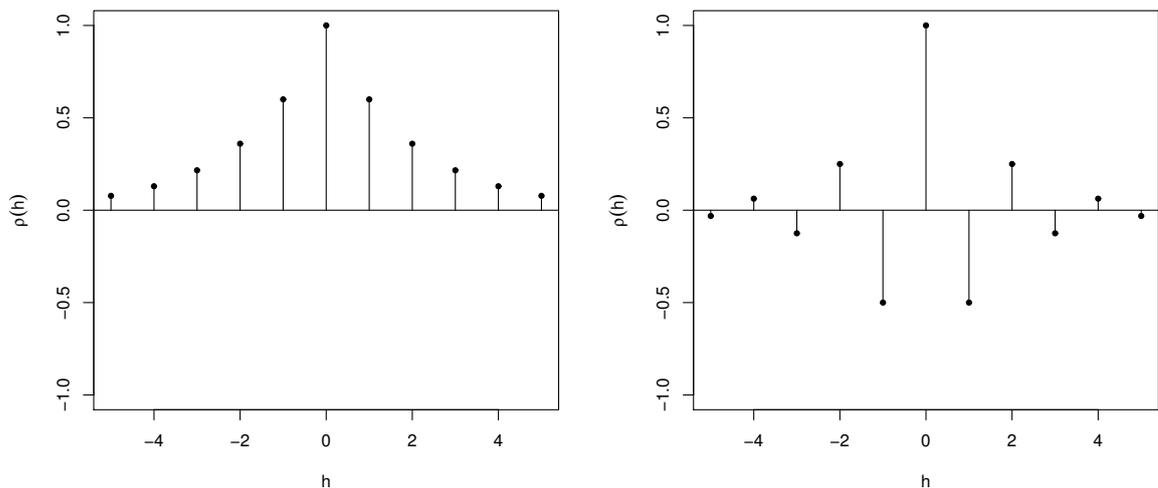
en posant  $i = k - h$ . On obtient ainsi l'ACV et l'ACF du processus AR(1). Pour  $h \in \mathbb{Z}$ ,

$$\gamma(h) = \frac{\sigma^2 \theta^{|h|}}{1 - \theta^2} \quad \text{et} \quad \rho(h) = \frac{\gamma(h)}{\gamma(0)} = \theta^{|h|}.$$

On a bien sûr  $(c_k) \in \ell^2(\mathbb{Z})$  comme le montre le calcul de la variance ci-dessus, ce qui confirme bien que  $(X_t)$  est un processus du second ordre.

**Remarque 3.1.1** *Attention : si  $|\theta| < 1$ , le processus est stationnaire sur  $\mathbb{Z}$ . Comme nous le verrons en TD, le processus n'est pas nécessairement stationnaire sur  $\mathbb{N}$ .*

Nous représentons ci-dessous deux exemples d'ACF pour le processus AR(1) : le premier a pour paramètre  $\theta = 0.6$ , le second  $\theta = -0.5$ . Nous remarquons que, en accord avec l'expression de  $\rho(h)$  et contrairement aux processus MA, **l'ACF d'un processus autorégressif stationnaire décroît géométriquement vite mais ne s'annule jamais.**



### 3.1.2 Estimation des paramètres

On remarque tout d'abord qu'il suffit de faire le changement de variable

$$Y_t = X_t - \frac{\mu}{1 - \theta}$$

pour obtenir un processus AR(1) centré, l'espérance de  $(X_t)$  étant estimée naturellement par  $\bar{X}_n$ , la moyenne empirique du processus observé sur  $t \in \{1, \dots, n\}$  par exemple. Nous allons donc considérer que  $\mu = 0$  dans ce qui suit, par souci de

simplification. Tout comme dans le cadre du modèle de régression linéaire simple, on a, pour tout  $1 \leq t \leq n$ ,

$$\varepsilon_t = X_t - \theta X_{t-1} \quad \text{et donc} \quad \sum_{t=1}^n \varepsilon_t^2 = \sum_{t=1}^n (X_t - \theta X_{t-1})^2.$$

L'observation d'une trajectoire du processus sur  $t \in \{1, \dots, n\}$  nous conduit naturellement à l'estimateur des moindres carrés donné par

$$\hat{\theta}_n = \frac{\sum_{t=1}^n X_{t-1} X_t}{\sum_{t=1}^n X_{t-1}^2}$$

où l'on considère arbitrairement que  $X_0 = 0$  (presque sûrement).

**Remarque 3.1.2** *Si l'on décide malgré tout de travailler sur le processus décentré (avec  $\mu \neq 0$ ), l'estimation de  $\theta$  se fait de la même manière, en remplaçant  $X_t$  et  $X_{t-1}$  par  $X_t - \bar{X}_n$  et  $X_{t-1} - \bar{X}_n$ , respectivement, dans l'expression ci-dessus.*

**Proposition 3.1.1** *Supposons que  $|\theta| < 1$  dans le modèle AR(1). Alors, l'estimateur des moindres carrés  $\hat{\theta}_n$  de  $\theta$  est consistant, c'est-à-dire que*

$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta.$$

De plus, on a la normalité asymptotique

$$\sqrt{n} (\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1 - \theta^2).$$

**Démonstration.** Admise. □

Un estimateur consistant de la variance résiduelle  $\sigma^2$  est alors donné par

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{t=1}^n (X_t - \hat{\theta}_n X_{t-1})^2.$$

De plus, on déduit de la normalité asymptotique que

$$\frac{\sqrt{n} (\hat{\theta}_n - \theta)}{\sqrt{1 - \hat{\theta}_n^2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

par le lemme de Slutsky, puisque  $\hat{\theta}_n^2$  est un estimateur consistant de  $\theta^2$ . Cela entraîne qu'un intervalle de confiance **asymptotique** de niveau  $1 - \alpha$  (avec  $0 < \alpha < 1$ ) pour  $\theta$  est donné par

$$\left[ \hat{\theta}_n - \frac{u_{1-\alpha/2}}{\sqrt{n}} \sqrt{1 - \hat{\theta}_n^2}, \hat{\theta}_n + \frac{u_{1-\alpha/2}}{\sqrt{n}} \sqrt{1 - \hat{\theta}_n^2} \right]$$

où  $u_{1-\alpha/2}$  est le quantile d'ordre  $1 - \alpha/2$  de la loi  $\mathcal{N}(0, 1)$ .

### 3.1.3 Modélisation et prévision

Si la modélisation et la prévision nécessitaient d'avoir recours à un algorithme évaluant pas à pas les résidus dans le cas du processus MA, tout est ici beaucoup plus simple. En effet, le régresseur  $X_{t-1}$  est **observé** à l'instant  $t$ . Il s'ensuit que, pour tout  $t \in \{1, \dots, n\}$ , nous reconstruisons simplement

$$\tilde{X}_t = \hat{\theta}_n X_{t-1} \quad \text{avec pour résidu} \quad \tilde{\varepsilon}_t = X_t - \tilde{X}_t = X_t - \hat{\theta}_n X_{t-1}.$$

Une fois le processus observé jusqu'à l'instant  $n$ , nous souhaitons prédire  $X_{n+1}$ . Nous avons alors tout simplement

$$\tilde{X}_{n+1} = \hat{\theta}_n X_n.$$

Si le processus est décentré, nous ajoutons  $\bar{X}_n$  à nos reconstructions et à notre prévision.

## 3.2 Focus sur le processus AR( $p$ )

Nous abordons désormais le processus autorégressif d'ordre  $p$ , qui est donc la généralisation du processus AR(1). Nous allons sommairement étudier ses propriétés de stationnarité puis travailler à l'estimation de ses paramètres. Avant cela, nous proposons un lemme d'analyse qui nous sera utile par la suite.

**Définition 3.2.1** Soit  $\mathcal{A}$  le polynôme d'ordre  $p$  défini, pour tout  $z \in \mathbb{C}$ , par

$$\mathcal{A}(z) = 1 - \sum_{k=1}^p a_k z^k.$$

On dit que  $\mathcal{A}$  est un « polynôme causal » si toutes ses racines sont en dehors du disque unité. En d'autres termes,  $\mathcal{A}$  est causal si, pour tout  $z \in \mathbb{C}$  tel que  $|z| \leq 1$ ,  $\mathcal{A}(z) \neq 0$ .

**Lemme 3.2.1** Soit  $\mathcal{A}$  un polynôme causal et soit  $r \in \mathbb{R}$  tel que  $1 < r < |z_{\min}|$  où  $z_{\min}$  est la racine de  $\mathcal{A}$  de plus petit module. Alors,  $\mathcal{A}$  est inversible dans la boule de centre 0 et de rayon  $r$ . C'est-à-dire que, pour tout  $z \in \mathbb{C}$  tel que  $|z| \leq r$ , on a

$$\mathcal{A}^{-1}(z) = \mathcal{P}(z) = \sum_{k=0}^{\infty} p_k z^k$$

où les coefficients  $p_0, p_1, \dots$  peuvent être calculés explicitement en fonction de  $a_1, a_2, \dots, a_p$ .

**Démonstration.** Admise. □

On a en particulier

$$\begin{aligned} \mathcal{A}(z) \mathcal{P}(z) &= \mathcal{A}(z) \mathcal{A}^{-1}(z) = (1 - a_1 z - a_2 z^2 - \dots - a_p z^p)(p_0 + p_1 z + p_2 z^2 + \dots) \\ &= p_0 + (p_1 - a_1 p_0)z + (p_2 - a_1 p_1 - a_2 p_0)z^2 + \dots \\ &= 1 \end{aligned}$$

ce qui permet d'identifier

$$\begin{cases} p_0 &= 1 \\ p_1 &= a_1 p_0 = a_1 \\ p_2 &= a_1 p_1 + a_2 p_0 = a_1^2 + a_2 \\ p_3 &= a_1 p_2 + a_2 p_1 + a_3 p_0 = a_1^3 + 2a_1 a_2 + a_3 \\ &\vdots \end{cases}$$

Il est important de noter que le polynôme  $\mathcal{A}$  est de degré  $p$  fini tandis que son polynôme inverse  $\mathcal{P}$  est de degré infini.

### 3.2.1 Écriture causale et stationnarité

À l'aide des outils que l'on vient de proposer, il est clair que la causalité de  $\mathcal{A}$  va jouer un rôle dans la stationnarité de  $(X_t)$  lorsque ce dernier est engendré par un  $\text{AR}(p)$ , à travers son écriture linéaire. En effet, le modèle (avec  $\mu = 0$  pour simplifier les calculs) s'écrit, pour tout  $t \in \mathbb{Z}$ ,

$$\mathcal{A}(L) X_t = \varepsilon_t$$

où  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2$  et, pour tout  $z \in \mathbb{C}$ ,  $\mathcal{A}(z) = 1 - a_1 z - \dots - a_p z^p$ . Ainsi, si le polynôme autorégressif  $\mathcal{A}$  est causal, on peut écrire

$$X_t = \mathcal{A}^{-1}(L) \varepsilon_t = \sum_{k=0}^{\infty} p_k L^k \varepsilon_t = \sum_{k=0}^{\infty} p_k \varepsilon_{t-k}$$

qui est l'écriture causale de  $(X_t)$  sous forme de processus linéaire (donc stationnaire), et l'on peut montrer que  $(p_k) \in \ell^2(\mathbb{Z})$ . On voit que  $(X_t)$  s'écrit comme un processus  $\text{MA}(\infty)$ . **Il s'ensuit que le processus  $\text{AR}(p)$  défini sur  $\mathbb{Z}$  est stationnaire dès que son polynôme autorégressif  $\mathcal{A}$  est causal.** On parlera alors par extension de « processus autorégressif causal ». On tire des expressions associées aux processus linéaires que, pour tout  $t \in \mathbb{Z}$ ,

$$\mathbb{E}[X_t] = 0, \quad \mathbb{V}(X_t) = \sigma^2 \sum_{k=0}^{\infty} p_k^2 \quad \text{et} \quad \text{Cov}(X_t, X_{t-h}) = \sigma^2 \sum_{k=0}^{\infty} p_k p_{k+|h|}$$

pour tout  $h \in \mathbb{Z}$ . Ces expressions sont explicitement calculables une fois que les coefficients  $(p_k)$  ont été identifiés (dépendant de l'ordre  $p$  du processus). En particulier, si  $p = 1$  alors  $p_k = \theta^k$  et l'on retrouve exactement les résultats proposés dans la section précédente. Par ailleurs, pour conserver l'analogie avec le cas  $p = 1$ , on remarque que le polynôme autorégressif est alors caractérisé par  $\mathcal{A}(z) = 1 - \theta z$  et que son unique racine est bien évidemment donnée par

$$z_0 = \frac{1}{\theta}.$$

Dès lors, la causalité de  $\mathcal{A}$  s'exprime par  $|z_0| > 1$ , et donc par  $|\theta| < 1$  : on retrouve bien la condition de stationnarité de l' $\text{AR}(1)$ . Nous étudierons en TD la stationnarité de plusieurs exemples de processus  $\text{AR}(p)$  avec  $p > 1$ .

**Remarque 3.2.1** *Attention : là encore et tout comme pour l'AR(1), un processus autorégressif causal est stationnaire sur  $\mathbb{Z}$ , mais le même processus défini sur  $\mathbb{N}$  n'est pas nécessairement stationnaire.*

Concluons cette partie en supposant que le processus  $(X_t)$  est causal (donc stationnaire) mais pas centré, et donc que  $\mu \neq 0$ . On a alors, pour tout  $t \in \mathbb{Z}$ ,  $\mathbb{E}[X_t] = m$  et l'identification de  $m$  se fait en résolvant l'équation

$$m = \mu + a_1 m + a_2 m + \dots + a_p m$$

obtenue en passant à l'espérance dans l'expression du processus AR( $p$ ), constante par stationnarité. Il vient alors

$$m = \frac{\mu}{1 - a_1 - a_2 - \dots - a_p}$$

qui représente l'espérance du processus stationnaire, et qui s'estime naturellement par  $\bar{X}_n$  sur une trajectoire  $t \in \{p+1, \dots, n\}$ .

### 3.2.2 Estimation des paramètres

Nous allons proposer deux méthodes d'estimation des paramètres dans un modèle AR( $p$ ). Nous ne nous attarderons pas sur les moindres carrés, car la stratégie est exactement la même que celle vue au Chapitre V du module de « modélisation statistique ». Par contre, nous détaillerons une méthode simple et efficace dite « de Yule-Walker ». Soit une trajectoire observée sur  $t \in \{p+1, \dots, n\}$  d'un processus  $(X_t)$  centré ( $\mu = 0$ ) engendré par un AR( $p$ ) stationnaire dont le bruit blanc  $(\varepsilon_t)$  est de variance  $\sigma^2$ . On considère arbitrairement que  $X_0 = 0$  (presque sûrement).

#### Les moindres carrés

Nous cherchons à minimiser

$$\sum_{t=p+1}^n \varepsilon_t^2 = \sum_{t=p+1}^n (X_t - a_1 X_{t-1} - \dots - a_p X_{t-p})^2.$$

Nous savons que l'estimateur des moindres carrés dans un modèle de régression linéaire multiple de paramètre  $\beta$  est donné par

$$\hat{\beta} = (X^t X)^{-1} X^t Y$$

où la matrice  $X$  regroupe les variables explicatives et le vecteur  $Y$  les variables à expliquer. En adaptant cette expression à notre cadre de travail, on a  $\beta = a$  (le vecteur des paramètres) et

$$a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}, \quad X = \begin{pmatrix} X_p & X_{p-1} & \dots & X_1 \\ X_{p+1} & X_p & \dots & X_2 \\ \vdots & \vdots & & \vdots \\ X_{n-1} & X_{n-2} & \dots & X_{n-p} \end{pmatrix}, \quad Y = \begin{pmatrix} X_{p+1} \\ X_{p+2} \\ \vdots \\ X_n \end{pmatrix} \quad \text{et} \quad \varepsilon = \begin{pmatrix} \varepsilon_{p+1} \\ \varepsilon_{p+2} \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

**Proposition 3.2.1** *Supposons que le polynôme  $\mathcal{A}$  est causal. Soit  $\hat{a}_n = (X^t X)^{-1} X^t Y$  l'estimateur des moindres carrés du paramètre  $a$  dans le modèle  $AR(p)$  centré donné, pour tout  $t \in \{p+1, \dots, n\}$ , par*

$$\mathcal{A}(L) X_t = \varepsilon_t$$

où  $a_p \neq 0$  et  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2$ . Alors, le vecteur  $\hat{a}_n$  est un estimateur consistant de  $a$ .

**Démonstration.** Admise. □

**Remarque 3.2.2** *En pratique et dans un cadre chronologique, il arrive que des corrélations fortes soient présentes entre les colonnes de la matrice  $X$ , ce qui est susceptible de perturber numériquement l'inversion de la matrice  $X^t X$ . On ajoute généralement la matrice identité  $I_p$  d'ordre  $p$  pour assurer l'inversibilité. En effet, on voit par construction que  $X^t X$  est semi-définie positive. Ainsi, pour tout  $\tau \in \mathbb{R}^p$  non nul, on a*

$$\tau^t (X^t X + I_p) \tau = \tau^t X^t X \tau + \|\tau\|^2 > 0$$

ce qui assure que  $X^t X + I_p$  est définie positive, donc inversible. En outre, cela ne perturbe pas les résultats de convergence.

### La méthode de Yule-Walker

Une méthode plus simple à mettre en œuvre repose sur l'étude de l'ACV du processus stationnaire sur  $\mathbb{Z}$ . En effet, en multipliant de chaque côté de l'égalité par  $X_{t-1}$ , il est clair que, pour tout  $t \in \mathbb{Z}$ , on a

$$X_{t-1} X_t = a_1 X_{t-1}^2 + a_2 X_{t-1} X_{t-2} + \dots + a_p X_{t-1} X_{t-p} + X_{t-1} \varepsilon_t.$$

Par passage à l'espérance en utilisant les propriétés de stationnarité, on obtient l'équation

$$\gamma(1) = a_1 \gamma(0) + a_2 \gamma(1) + \dots + a_p \gamma(p-1).$$

De la même manière, en multipliant cette fois par  $X_{t-2}$ , on obtient

$$\gamma(2) = a_1 \gamma(1) + a_2 \gamma(0) + \dots + a_p \gamma(p-2).$$

On répète cette opération  $p$  fois, jusqu'à multiplier par  $X_{t-p}$ . Il en résulte alors un système de  $p$  équations à  $p$  inconnues, donné par

$$\begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \dots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \dots & \gamma(p-2) \\ \vdots & \vdots & \vdots & & \vdots \\ \gamma(p-1) & \gamma(p-2) & \gamma(p-3) & \dots & \gamma(0) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(p) \end{pmatrix}$$

où la forme de la matrice explicative (symétrique et constante par diagonales) en fait une « matrice de Toeplitz », dont on peut montrer qu'elle est inversible dans

le cas d'un processus autorégressif stationnaire. On parle du « système de Yule-Walker » ou encore des « équations de Yule-Walker ». Bien entendu, ce système est inexploitable pour estimer le vecteur  $a$  puisque les valeurs de l'ACV sont inconnues. Cependant, on sait les estimer de façon consistante. Pour tout  $h \in \mathbb{Z}$  et sur une trajectoire  $t \in \{p+1, \dots, n\}$  observée du processus, alors

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=p+1}^{n-|h|} X_t X_{t+|h|} \xrightarrow{\mathbb{P}} \gamma(h).$$

Comme précédemment, on remplace  $X_t$  et  $X_{t+|h|}$  par  $X_t - \bar{X}_n$  et  $X_{t+|h|} - \bar{X}_n$ , respectivement, si le processus possède un intercept  $\mu \neq 0$ . Finalement, l'estimateur  $\check{a}_n$  de Yule-Walker du paramètre  $a$  est donné par

$$\check{a}_n = \hat{\Gamma}^{-1} \hat{\gamma}$$

où

$$\hat{\Gamma} = \begin{pmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \hat{\gamma}(2) & \dots & \hat{\gamma}(p-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \hat{\gamma}(1) & \dots & \hat{\gamma}(p-2) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \hat{\gamma}(p-1) & \hat{\gamma}(p-2) & \hat{\gamma}(p-3) & \dots & \hat{\gamma}(0) \end{pmatrix} \quad \text{et} \quad \hat{\gamma} = \begin{pmatrix} \hat{\gamma}(1) \\ \hat{\gamma}(2) \\ \vdots \\ \hat{\gamma}(p) \end{pmatrix}.$$

**Proposition 3.2.2** *Supposons que le polynôme  $\mathcal{A}$  est causal. Soit  $\check{a}_n = \hat{\Gamma}^{-1} \hat{\gamma}$  l'estimateur de Yule-Walker du paramètre  $a$  dans le modèle  $AR(p)$  centré donné, pour tout  $t \in \{p+1, \dots, n\}$ , par*

$$\mathcal{A}(L) X_t = \varepsilon_t$$

où  $a_p \neq 0$  et  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2$ . Alors, le vecteur  $\check{a}_n$  est un estimateur consistant de  $a$ .

**Démonstration.** Admise. □

Enfin, si nous multiplions de chaque côté de l'égalité par  $X_t$ , nous obtenons comme dernière équation

$$\gamma(0) = a_1 \gamma(1) + a_2 \gamma(2) + \dots + a_p \gamma(p) + \sigma^2$$

ce qui nous conduit à l'estimateur de Yule-Walker consistant de la variance  $\sigma^2$ , donné par

$$\check{\sigma}_n^2 = \hat{\gamma}(0) - \check{a}_{1,n} \hat{\gamma}(1) - \check{a}_{2,n} \hat{\gamma}(2) - \dots - \check{a}_{p,n} \hat{\gamma}(p)$$

où  $\check{a}_{1,n}, \check{a}_{2,n}, \dots, \check{a}_{p,n}$  sont les composantes de  $\check{a}_n$ .

**Remarque 3.2.3** *Les estimateurs des moindres carrés  $\hat{a}_n$  et de Yule-Walker  $\check{a}_n$  sont équivalents dans un contexte de stationnarité aux termes de bord près, comme nous le verrons en TD pour le processus  $AR(1)$ . Ils possèdent les mêmes propriétés de convergence et de normalité asymptotique. En supposant que la perturbation  $(\varepsilon_t)$  est gaussienne, on peut également établir l'estimateur  $\tilde{a}_n$  du maximum de vraisemblance du paramètre  $a$ , et montrer qu'il possède lui aussi les mêmes propriétés que  $\hat{a}_n$  et  $\check{a}_n$ . Là encore, nous le montrerons en TD pour le processus  $AR(1)$ .*

### 3.2.3 Modélisation et prévision

La problématique de la reconstruction des données et de la prévision est très simple pour le processus autorégressif, car les variables explicatives sont observables. Pour tout  $t \in \{p + 1, \dots, n\}$ , nous reconstruisons simplement

$$\tilde{X}_t = \hat{a}_{1,n} X_{t-1} + \hat{a}_{2,n} X_{t-2} + \dots + \hat{a}_{p,n} X_{t-p}$$

avec pour résidu

$$\tilde{\varepsilon}_t = X_t - \tilde{X}_t = X_t - \hat{a}_{1,n} X_{t-1} - \hat{a}_{2,n} X_{t-2} - \dots - \hat{a}_{p,n} X_{t-p}.$$

Une fois le processus observé jusqu'à l'instant  $n$ , nous souhaitons prédire  $X_{n+1}$ . Nous avons alors

$$\tilde{X}_{n+1} = \hat{a}_{1,n} X_n + \hat{a}_{2,n} X_{n-1} + \dots + \hat{a}_{p,n} X_{n-p+1}.$$

Lorsque la variable  $X_{n+1}$  est observée (donc après la prédiction), nous obtenons l'erreur de prédiction donnée par

$$\tilde{\varepsilon}_{n+1} = X_{n+1} - \tilde{X}_{n+1}.$$

La reconstruction et la prédiction sont ici faites par moindres carrés, mais il est bien sûr possible de reconstruire les données ou de les prédire à l'aide de la méthode de Yule-Walker ou du maximum de vraisemblance. Il suffit alors de remplacer  $\hat{a}_n$  par  $\tilde{a}_n$  ou  $\bar{a}_n$ , le cas échéant. Si le processus est décentré, nous ajoutons  $\bar{X}_n$  à nos reconstructions et à notre prévision.

## 3.3 Quelques exemples simples

Pour conclure ce chapitre, considérons les 3 processus AR les plus simples. Nous en illustrons une trajectoire ainsi que l'autocorrélogramme associé.

### 3.3.1 Le processus AR(0)

Pour tout  $t \in \mathbb{Z}$ , le processus AR(0) est engendré par

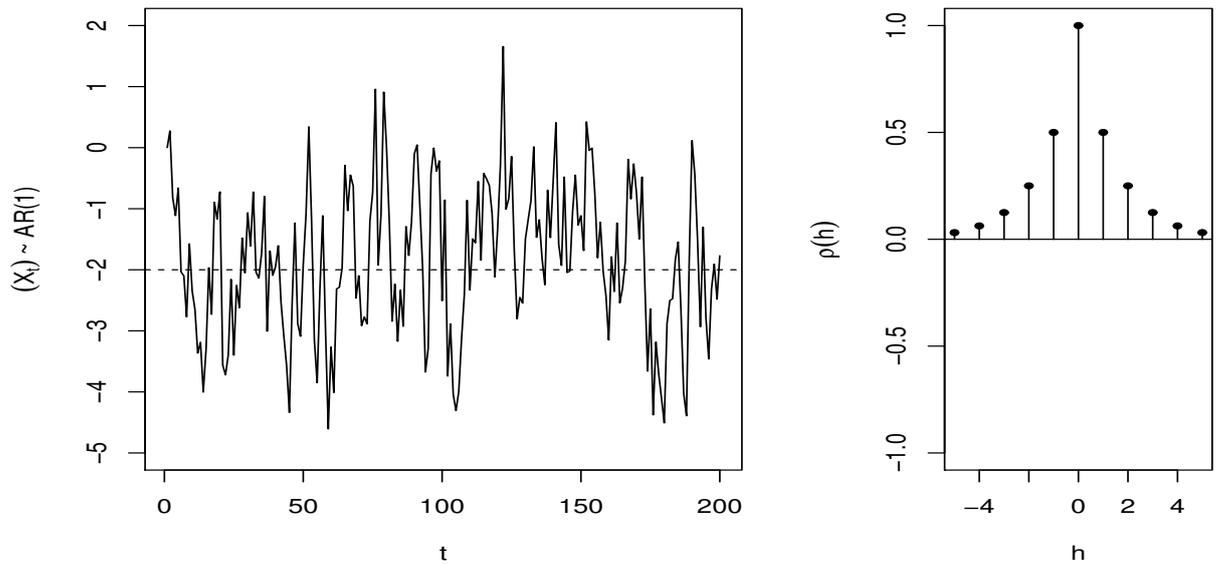
$$X_t = \mu + \varepsilon_t$$

où  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2$ . Ainsi, nous retrouvons l'exemple donné dans le chapitre précédent, car un AR(0) et un MA(0) ont la même expression.

### 3.3.2 Le processus AR(1)

Pour tout  $t \in \mathbb{Z}$ , le processus AR(1) est engendré par

$$X_t = \mu + a_1 X_{t-1} + \varepsilon_t$$



où  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2$ . L'exemple ci-dessous est simulé avec  $\mu = -1$ ,  $a_1 = 0.5$ ,  $(\varepsilon_t) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$  et  $n = 200$ .

Le processus AR(1) de paramètre  $a_1 = 0.5$  est donc stationnaire sur  $\mathbb{Z}$ , et son espérance est ici caractérisée par

$$\mathbb{E}[X_t] = \frac{\mu}{1 - a_1} = -2.$$

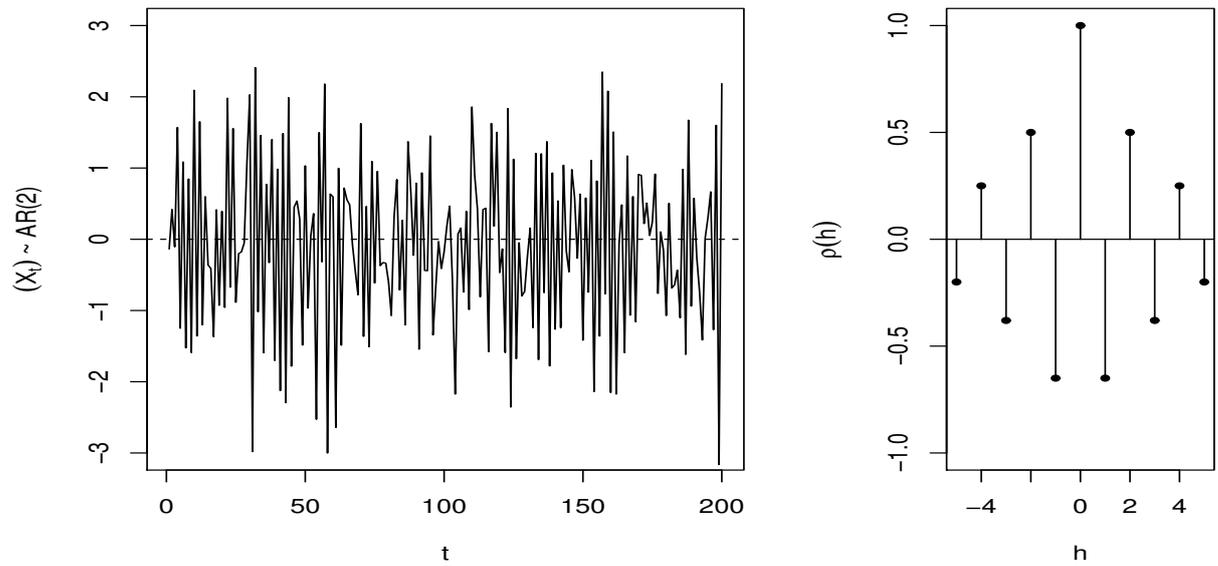
Ses propriétés de variance et de covariance sont largement décrites dans la première section du chapitre.

### 3.3.3 Le processus AR(2)

Pour tout  $t \in \mathbb{Z}$ , le processus AR(2) est engendré par

$$X_t = \mu + a_1 X_{t-1} + a_2 X_{t-2} + \varepsilon_t$$

où  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2$ . L'exemple ci-dessous est simulé avec  $\mu = 0$ ,  $a_1 = -0.5$ ,  $a_2 = 0.2$ ,  $(\varepsilon_t) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$  et  $n = 200$ .



Le processus AR(2) de paramètres  $a_1 = -0.5$  et  $a_2 = 0.2$  est donc stationnaire sur  $\mathbb{Z}$ , et son espérance est ici caractérisée par

$$\mathbb{E}[X_t] = 0.$$

Ses propriétés de variance et de covariance sont compliquées à calculer comme nous l'avons vu dans la seconde section, mais elles restent identifiables.

# Chapitre 4

## Le processus ARMA et la non stationnarité

Nous abordons désormais le processus ARMA qui se présente comme la combinaison d'une dynamique AR et d'une perturbation MA. Nous étudierons ses propriétés habituelles de stationnarité, d'estimation ainsi qu'un algorithme de prédiction. Nous terminerons le chapitre par la problématique de la non stationnarité, ne permettant plus d'estimer correctement les paramètres du modèle : comment la détecter et comment la gérer ? Le Chapitre IV doit ainsi être appréhendé comme la généralisation des Chapitres II et III.

### 4.1 Le processus ARMA

Le processus ARMA permet de modéliser un processus autorégressif engendré par une moyenne mobile. Ainsi, sur  $\mathbb{Z}$ , si  $(Y_t)$  est un MA( $q$ ) et si  $(X_t)$  est un AR( $p$ ) engendré par  $(Y_t)$ , alors  $(X_t)$  est un ARMA( $p, q$ ). On comprend alors que le processus ARMA héritera des propriétés de stationnarité de l'AR( $p$ ) et des procédures d'estimation du MA( $q$ ).

**Définition 4.1.1** *On dit que la série chronologique  $(X_t)$  définie sur  $\mathbb{Z}$  est un « processus autorégressif moyenne mobile d'ordre  $(p, q)$  » si elle est définie, pour tout  $t \in \mathbb{Z}$ , par*

$$X_t = \mu + a_1 X_{t-1} + \dots + a_p X_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_q \varepsilon_{t-q}$$

où  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2$ ,  $a_p \neq 0$  et  $b_q \neq 0$ .

La notation usuelle des autorégressifs moyennes mobiles est ARMA, de l'anglais *autoregressive moving average*. Pour signifier que le processus  $(X_t)$  est un autorégressif moyenne mobile d'ordre  $(p, q)$ , nous noterons

$$(X_t) \sim \text{ARMA}(p, q).$$

**Remarque 4.1.1** *Attention aux notations ! Les parenthèses autour de  $(X_t)$  sont ici indispensables : elles signifient que l'on considère le processus, alors que  $X_t$  sans parenthèses désigne la variable aléatoire extraite du processus à l'instant  $t$ .*

En général, l'écriture  $\text{ARMA}(p, q)$  se condense à l'aide des opérateurs chronologiques. Soient les polynômes définis, pour tout  $z \in \mathbb{C}$ , par  $\mathcal{A}(z) = 1 - a_1 z - \dots - a_p z^p$  et  $\mathcal{B}(z) = 1 + b_1 z + \dots + b_q z^q$ . Alors, on écrira

$$\mathcal{A}(L) X_t = \mu + \mathcal{B}(L) \varepsilon_t.$$

En effet, l'application de l'opérateur retard conduit à

$$\mathcal{A}(L)X_t = (1 - a_1 L - \dots - a_p L^p) X_t = X_t - a_1 X_{t-1} - \dots - a_p X_{t-p}$$

pour la partie AR, et à

$$\mathcal{B}(L) \varepsilon_t = (1 + b_1 L + \dots + b_q L^q) \varepsilon_t = \varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_q \varepsilon_{t-q}$$

pour la partie MA.

**Remarque 4.1.2** *Il est clair qu'un  $\text{ARMA}(0, q)$  est un  $\text{MA}(q)$  avec  $\mathcal{A}(z) = 1$ , alors qu'un  $\text{ARMA}(p, 0)$  est un  $\text{AR}(p)$  avec  $\mathcal{B}(z) = 1$ .*

### 4.1.1 Stationnarité

Nous commençons par étudier les propriétés de stationnarité du processus ARMA, et nous avons besoin pour cela de son écriture causale. À l'aide des outils d'inversibilité des polynômes développés dans le Chapitre III, nous savons que la causalité de  $\mathcal{A}$  permettra d'écrire le processus  $\text{ARMA}(p, q)$  sous une forme  $\text{MA}(\infty)$ . En effet, si nous supposons que  $\mathcal{A}$  est causal, alors il est inversible pour tout  $|z| \leq 1$  et, pour tout  $t \in \mathbb{Z}$ , on voit que

$$X_t = \mathcal{A}^{-1}(L) \mu + \mathcal{A}^{-1}(L) \mathcal{B}(L) \varepsilon_t = m + \mathcal{Q}(L) \varepsilon_t$$

où la constante  $m$  et le polynôme  $\mathcal{Q}$  d'ordre infini sont identifiables. Soit  $\mathcal{P} = \mathcal{A}^{-1}$  le polynôme d'ordre infini permettant d'inverser  $\mathcal{A}$ . Pour tout  $z \in \mathbb{C}$ , on a

$$\mathcal{A}(z) \mathcal{P}(z) = (1 - a_1 z - \dots - a_p z^p)(p_0 + p_1 z + p_2 z^2 + \dots) = 1$$

ce qui permet d'identifier la suite  $(p_k) \in \ell^2(\mathbb{N})$ , ainsi que nous l'avons fait au cours du Chapitre III. Maintenant, la relation

$$\mathcal{P}(z) \mathcal{B}(z) = (p_0 + p_1 z + p_2 z^2 + \dots)(1 + b_1 z + \dots + b_q z^q) = q_0 + q_1 z + q_2 z^2 + \dots$$

permet à son tour d'identifier la suite  $(q_k) \in \ell^2(\mathbb{N})$ , à l'origine du polynôme  $\mathcal{Q}$ . Il s'ensuit que, pour tout  $t \in \mathbb{Z}$ ,

$$X_t = m + \sum_{k=0}^{\infty} q_k \varepsilon_{t-k}$$

qui est bien un processus linéaire, et donc stationnaire. Ainsi, **le processus ARMA( $p, q$ ) défini sur  $\mathbb{Z}$  est stationnaire dès que son polynôme autorégressif  $\mathcal{A}$  est causal.** On tire des expressions associées aux processus linéaires que, pour tout  $t \in \mathbb{Z}$ ,

$$\mathbb{E}[X_t] = m, \quad \mathbb{V}(X_t) = \sigma^2 \sum_{k=0}^{\infty} q_k^2 \quad \text{et} \quad \text{Cov}(X_t, X_{t-h}) = \sigma^2 \sum_{k=0}^{\infty} q_k q_{k+|h|}$$

pour tout  $h \in \mathbb{Z}$ . Ces expressions sont explicitement calculables une fois que les coefficients ( $q_k$ ) ont été identifiés (dépendant de l'ordre ( $p, q$ ) du processus). L'espérance quant à elle satisfait la relation

$$\mathbb{E}[X_t] = \mu + a_1 \mathbb{E}[X_{t-1}] + \dots + a_p \mathbb{E}[X_{t-p}] + \mathbb{E}[\varepsilon_t] + b_1 \mathbb{E}[\varepsilon_{t-1}] + \dots + \mathbb{E}[b_q \varepsilon_{t-q}]$$

et donc, par stationnarité,

$$m = \mu + a_1 m + \dots + a_p m.$$

L'espérance d'un processus ARMA( $p, q$ ) stationnaire est donc identique à celle d'un processus AR( $p$ ) stationnaire, à savoir

$$m = \frac{\mu}{1 - a_1 - a_2 - \dots - a_p}$$

et ce résultat est tout à fait logique dans la mesure où la perturbation engendrant la dynamique autorégressive est centrée, que ce soit un bruit blanc (pour l'AR) ou un MA (pour l'ARMA). La perturbation MA influe seulement sur la variance du processus.

**Remarque 4.1.3** *Attention : là encore et tout comme pour l'AR, un processus ARMA causal est stationnaire sur  $\mathbb{Z}$ , mais le même processus défini sur  $\mathbb{N}$  n'est pas nécessairement stationnaire. Tout dépend de la densité des valeurs initiales.*

## 4.1.2 Un exemple détaillé : l'ARMA(1,1)

Pour éclaircir les idées, nous allons détailler le cas de l'ARMA(1,1). Ce dernier s'écrit, pour tout  $t \in \mathbb{Z}$ ,

$$X_t = \mu + a_1 X_{t-1} + \varepsilon_t + b_1 \varepsilon_{t-1}$$

où ( $\varepsilon_t$ ) est un bruit blanc de variance  $\sigma^2$ . Les polynômes associés s'écrivent  $\mathcal{A}(z) = 1 - a_1 z$  et  $\mathcal{B}(z) = 1 + b_1 z$ , et nous savons que la causalité de  $\mathcal{A}$  est vérifiée si et seulement si  $|a_1| < 1$ . Cette condition garantit alors la stationnarité de ( $X_t$ ). Tout d'abord, on a

$$\mathbb{E}[X_t] = \frac{\mu}{1 - a_1} = m.$$

Par ailleurs, identifions la suite ( $q_k$ ). On a déjà vu dans la Section 1 du Chapitre III que, dans le cas d'un AR(1), on a, pour tout  $k \in \mathbb{N}$ ,

$$p_k = a_1^k.$$

Ainsi, l'égalité polynomiale

$$\begin{aligned} (p_0 + p_1 z + p_2 z^2 + \dots)(1 + b_1 z) &= (1 + a_1 z + a_1^2 z^2 + \dots)(1 + b_1 z) \\ &= q_0 + q_1 z + q_2 z^2 + \dots \end{aligned}$$

nous conduit, par identification, à

$$\begin{cases} q_0 = 1 \\ q_1 = a_1 + b_1 \\ q_2 = a_1^2 + a_1 b_1 = a_1(a_1 + b_1) \\ q_3 = a_1^3 + a_1^2 b_1 = a_1^2(a_1 + b_1) \\ \vdots \end{cases}$$

D'une manière générale, on a  $q_0 = 1$  et, pour tout  $k \in \mathbb{N}^*$ ,  $q_k = a_1^{k-1}(a_1 + b_1)$ . Nous en déduisons que

$$\begin{aligned} \mathbb{V}(X_t) &= \sigma^2 \sum_{k=0}^{\infty} q_k^2 = \sigma^2 \left[ 1 + (a_1 + b_1)^2 \sum_{k=1}^{\infty} a_1^{2(k-1)} \right] \\ &= \sigma^2 \left[ 1 + \frac{(a_1 + b_1)^2}{1 - a_1^2} \right] \\ &= \frac{\sigma^2}{1 - a_1^2} (b_1^2 + 2a_1 b_1 + 1) = \gamma(0). \end{aligned}$$

Enfin, pour tout  $h \neq 0$ ,

$$\begin{aligned} \text{Cov}(X_t, X_{t-h}) &= \sigma^2 \sum_{k=0}^{\infty} q_k q_{k+|h|} = \sigma^2 (a_1 + b_1) \left[ a_1^{|h|-1} + (a_1 + b_1) a_1^{|h|} \sum_{k=1}^{\infty} a_1^{2(k-1)} \right] \\ &= \sigma^2 (a_1 + b_1) a_1^{|h|-1} \left[ 1 + \frac{a_1(a_1 + b_1)}{1 - a_1^2} \right] \\ &= \frac{\sigma^2 (a_1 + b_1)(1 + a_1 b_1) a_1^{|h|-1}}{1 - a_1^2} = \gamma(h). \end{aligned}$$

On remarque que, lorsque  $b_1 = 0$ , on a

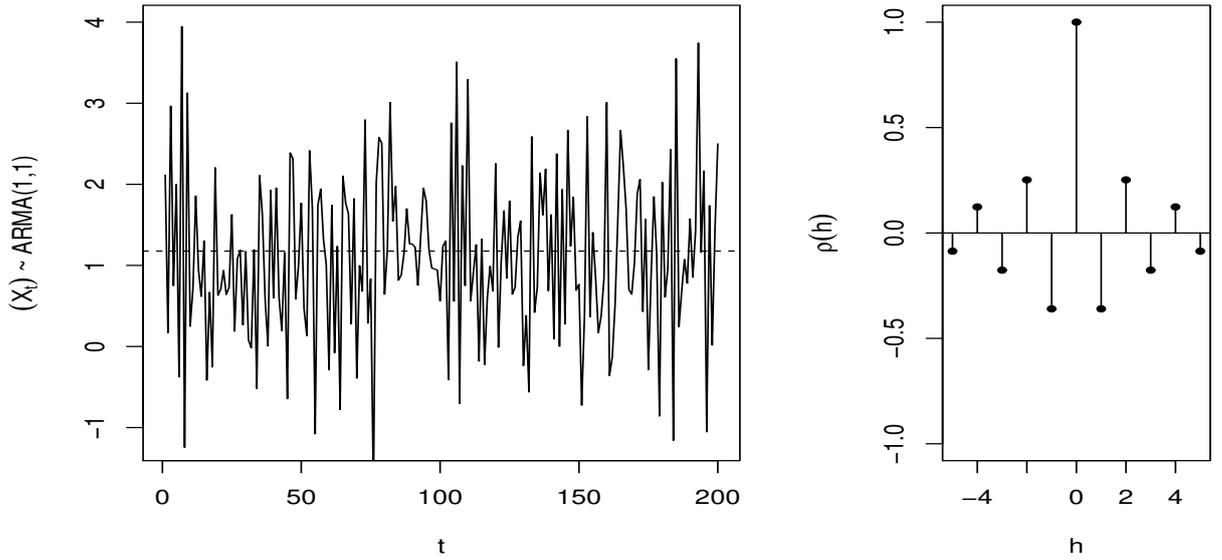
$$m = \frac{\mu}{1 - a_1}, \quad \gamma(0) = \frac{\sigma^2}{1 - a_1^2} \quad \text{et} \quad \gamma(h) = \frac{\sigma^2 a_1^{|h|}}{1 - a_1^2},$$

ce qui est caractéristique de l'AR(1). De même, lorsque  $a_1 = 0$ ,

$$m = \mu, \quad \gamma(0) = \sigma^2 (1 + b_1^2) \quad \text{et} \quad \gamma(h) = 0,$$

dès que  $h \neq 0$ , ce qui est caractéristique du MA(1). L'exemple ci-dessous est simulé avec  $\mu = 2$ ,  $a_1 = -0.7$ ,  $b_1 = 0.4$ ,  $(\varepsilon_t) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$  et  $n = 200$ .

On observe que la valeur moyenne du processus vaut  $2/1.7 \approx 1.18$ , et que l'ACF décroît géométriquement vite mais, contrairement aux processus MA, **ne s'annule jamais définitivement**.



### 4.1.3 Causalité et inversibilité

**Définition 4.1.2** On dit que le processus  $ARMA(p, q)$  défini sur  $\mathbb{Z}$  est un « processus causal » si son polynôme autorégressif  $\mathcal{A}$  est causal.

Un ARMA causal admet donc une représentation  $MA(\infty)$ , par l'intermédiaire de l'inverse de son polynôme autorégressif. Ainsi, pour tout  $t \in \mathbb{Z}$ ,

$$X_t = \mathcal{A}^{-1}(L)\mu + \mathcal{A}^{-1}(L)\mathcal{B}(L)\varepsilon_t = m + \mathcal{Q}(L)\varepsilon_t$$

comme nous l'avons vu précédemment, où le polynôme  $\mathcal{Q}$  d'ordre infini est identifiable et

$$m = \mathcal{A}^{-1}(L)\mu = \frac{\mu}{1 - a_1 - \dots - a_p}.$$

Un ARMA causal est donc stationnaire sur  $\mathbb{Z}$ .

**Définition 4.1.3** On dit que le processus  $ARMA(p, q)$  défini sur  $\mathbb{Z}$  est un « processus inversible » si son polynôme moyenne mobile  $\mathcal{B}$  est causal.

Un ARMA inversible admet donc une représentation  $AR(\infty)$ , par l'intermédiaire de l'inverse de son polynôme moyenne mobile. Ainsi, pour tout  $t \in \mathbb{Z}$ ,

$$\mathcal{B}^{-1}(L)\mathcal{A}(L)X_t = \mathcal{R}(L)X_t = \mathcal{B}^{-1}(L)\mu + \varepsilon_t = m' + \varepsilon_t$$

où le polynôme  $\mathcal{R}$  d'ordre infini est identifiable et

$$m' = \mathcal{B}^{-1}(L)\mu = \frac{\mu}{1 - b_1 - \dots - b_q}.$$

Un ARMA inversible n'est donc pas nécessairement stationnaire sur  $\mathbb{Z}$ , tout dépend des propriétés du polynôme  $\mathcal{R}$ .

**Définition 4.1.4** On dit que le processus ARMA( $p, q$ ) défini sur  $\mathbb{Z}$  est un « processus causal inversible » si son polynôme autorégressif  $\mathcal{A}$  ainsi que son polynôme moyenne mobile  $\mathcal{B}$  sont causaux.

Il va de soi qu'un ARMA causal inversible **est stationnaire** sur  $\mathbb{Z}$ . Un tel processus admet à la fois des représentations MA( $\infty$ ) et AR( $\infty$ ).

#### 4.1.4 Estimation des paramètres

On se donne une trajectoire, c'est-à-dire une réalisation de  $(X_t)$  sur un intervalle des temps fini,  $t \in \{1, \dots, n\}$  par exemple. L'expression du modèle ARMA fait intervenir un obstacle majeur en vue de l'estimation de ses paramètres : tout comme pour le processus MA, il n'est pas possible d'exprimer le bruit du modèle en fonction de variables observables. En effet, il est certes possible d'écrire

$$\varepsilon_t = X_t - \mu - a_1 X_{t-1} - \dots - a_p X_{t-p} - b_1 \varepsilon_{t-1} - \dots - b_q \varepsilon_{t-q}$$

mais les variables intervenant à droite de l'égalité, à l'exception de  $X_t, \dots, X_{t-p}$ , ne sont pas observables. Il s'ensuit qu'un modèle ARMA **ne peut pas être estimé par moindres carrés**. C'est pourquoi l'on utilise généralement le maximum de vraisemblance comme méthode d'estimation.

**Proposition 4.1.1** Supposons que les polynômes  $\mathcal{A}$  et  $\mathcal{B}$  possèdent toutes leurs racines à l'extérieur du cercle unité et donc qu'ils sont causaux (pour tout  $z \in \mathbb{C}$  tel que  $|z| \leq 1$ ,  $\mathcal{A}(z) \neq 0$  et  $\mathcal{B}(z) \neq 0$ ). Soient  $\hat{\mu}, \hat{a}_1, \dots, \hat{a}_p, \hat{b}_1, \dots, \hat{b}_q, \hat{\sigma}^2$  les estimateurs du maximum de vraisemblance dans le modèle ARMA( $p, q$ ) donné, pour tout  $t \in \{1, \dots, n\}$ , par

$$\mathcal{A}(L) X_t = \mu + \mathcal{B}(L) \varepsilon_t$$

où  $a_p \neq 0, b_q \neq 0$  et  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2$ . On suppose que les valeurs initiales  $X_{-p}, \dots, X_0, \varepsilon_{-q}, \dots, \varepsilon_0$  sont arbitrairement choisies. Alors,  $\hat{\mu}, \hat{a}_1, \dots, \hat{a}_p, \hat{b}_1, \dots, \hat{b}_q$  et  $\hat{\sigma}^2$  sont consistants.

*Démonstration.* Admise. □

Dans la pratique, cela passe par l'approximation de la vraisemblance du modèle par une loi normale. On obtient une fonction de vraisemblance très complexe et non linéaire (en raison des autocorrélations et donc du fait que les variables  $(X_t)$  ne sont pas indépendantes), que les logiciels de calcul (R, S+, Matlab, etc.) maximisent numériquement. **Les estimateurs du maximum de vraisemblance des paramètres d'un modèle ARMA n'ont donc pas d'expression explicite, mais ils sont approximables par les logiciels de calcul.** Notons pour conclure que l'espérance  $m$  peut être estimée indépendamment des autres paramètres par la moyenne empirique d'un échantillon observé,

$$\hat{m} = \bar{X}_n = \frac{1}{n} \sum_{t=1}^n X_t.$$

Les autres paramètres sont alors estimés en recentrant le processus, c'est-à-dire en ôtant aux observations la valeur  $\bar{X}_n$ . Attention : on se rappelle que  $m \neq \mu$  pour les AR et les ARMA ! Il s'agira alors d'estimer  $\mu$ , le cas échéant, par

$$\hat{\mu} = (1 - \hat{a}_1 - \dots - \hat{a}_p) \hat{m}.$$

**Remarque 4.1.4** *Il est possible d'établir des équations de Yule-Walker pour les processus ARMA, tout comme pour les AR mais d'une façon légèrement modifiée : il s'agit non plus de travailler sur les covariances d'ordre 0 à p du processus, mais sur les covariances d'ordre q à p + q. On obtient ainsi un système de p + 1 équations linéaires à p + 1 inconnues permettant d'estimer de façon consistante les paramètres  $a_1, \dots, a_p, \sigma^2$ .*

### 4.1.5 Application à la prévision

Les quantités non observables  $\varepsilon_{t-q}, \dots, \varepsilon_{t-1}$  nous obligent, tout comme pour le modèle MA, à passer par un algorithme de prévision. Ce n'était pas le cas pour les AR comme nous l'avons vu dans le Chapitre III, où tout était beaucoup plus simple. L'algorithme que nous proposons pour les ARMA (et qui n'est pas unique) est le suivant.

1. On considère une série chronologique  $(X_t)$  observée sur  $t \in \{1, \dots, n\}$ , modélisée par un modèle ARMA( $p, q$ ) et dont les paramètres  $\mu, a_1, \dots, a_p, b_1, \dots, b_q, \sigma^2$  ont été estimés par maximum de vraisemblance à l'aide d'un logiciel de calcul.
2. On note  $r = \max(p, q)$  l'instant séparant les valeurs initiales des valeurs engendrées par le processus.
3. On initialise les  $r$  premières valeurs du processus résiduel  $(\tilde{\varepsilon}_t)$  de manière arbitraire, de sorte que  $\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_r \in L^2(\Omega, \mathcal{A}, \mathbb{P})$  pour que le processus reste du second ordre.
4. Pour toutes les valeurs observées du processus situées dans l'intervalle de temps  $t \in \{r + 1, \dots, n\}$ ,
  - (a) on reconstruit

$$\tilde{X}_t = \hat{\mu} + \hat{a}_1 X_{t-1} + \dots + \hat{a}_p X_{t-p} + \hat{b}_1 \tilde{\varepsilon}_{t-1} + \dots + \hat{b}_q \tilde{\varepsilon}_{t-q},$$

- (b) puis on évalue le résidu de modélisation correspondant,

$$\tilde{\varepsilon}_t = X_t - \tilde{X}_t = X_t - \hat{\mu} - \hat{a}_1 X_{t-1} - \dots - \hat{a}_p X_{t-p} - \hat{b}_1 \tilde{\varepsilon}_{t-1} - \dots - \hat{b}_q \tilde{\varepsilon}_{t-q}.$$

5. Une fois le processus reconstruit  $(\tilde{X}_t)$  et les résidus de modélisation  $(\tilde{\varepsilon}_t)$  évalués jusqu'au dernier instant connu  $n$ , on prédit le premier instant inconnu par

$$\tilde{X}_{n+1} = \hat{\mu} + \hat{a}_1 X_n + \dots + \hat{a}_p X_{n-p+1} + \hat{b}_1 \tilde{\varepsilon}_n + \dots + \hat{b}_q \tilde{\varepsilon}_{n-q+1}.$$

Nous verrons plusieurs exemples de processus ARMA lors de la séance de TP finale, au cours de laquelle nous modéliserons l'évolution de certaines actions en bourse afin de faire de la prédiction.

## 4.2 Introduction à la non stationnarité

La propriété la plus importante d'un estimateur est sa consistance : il est primordial, pour bien appliquer un modèle (et donc modéliser correctement le phénomène étudié), d'estimer ses paramètres avec le plus de pertinence possible. Or, comme nous l'avons vu aux Chapitres II et III ainsi que dans la première section du Chapitre IV, les estimateurs utilisés au sein des modèles MA, AR ou ARMA nécessitent la causalité des polynômes associés ( $\mathcal{A}$  et/ou  $\mathcal{B}$ ) pour s'avérer consistants. En d'autres termes (et même si ce n'est pas toujours vrai), **le processus doit être stationnarisé pour être correctement estimé**. On comprend alors la nécessité de tester la stationnarité d'un processus observé avant de tenter de le modéliser.

### 4.2.1 Le processus ARIMA

Nous ne considérerons pas ici le cas dit « explosif » où le processus ARMA possède une racine autorégressive à l'intérieur strict du cercle unité, c'est-à-dire qu'il existe  $z_0 \in \mathbb{C}$  tel que  $|z_0| < 1$  et  $\mathcal{A}(z_0) = 0$ . En effet, on peut montrer qu'un tel processus explose et se dirige très rapidement vers l'infini, ce qui ne présente pour nous aucun intérêt pratique. Ainsi, les deux cas que nous étudions sont les suivants :

1. Toutes les racines de  $\mathcal{A}$  sont à l'extérieur strict du cercle unité.
2. Il existe au moins une racine de  $\mathcal{A}$  sur le cercle unité.

Le cas (1) résume la définition de la causalité de  $\mathcal{A}$  : nous savons alors que le processus ARMA engendré est stationnaire sur  $\mathbb{Z}$ . Le cas (2) en revanche marque une frontière entre la stationnarité et la non stationnarité. De fait, s'il existe  $z_0 \in \mathbb{C}$  tel que  $|z_0| = 1$  et  $\mathcal{A}(z_0) = 0$ , alors le polynôme  $\mathcal{A}$  n'est pas causal et le processus ARMA engendré **n'est pas stationnaire** : il n'admet pas d'écriture  $MA(\infty)$ . En séparant les racines situées à l'extérieur du cercle unité des racines situées sur le cercle unité, on sépare les termes engendrant la non stationnarité des termes engendrant la stationnarité, et l'on construit ce que l'on appelle un *processus ARIMA*.

**Définition 4.2.1** *On dit que la série chronologique  $(X_t)$  définie sur  $\mathbb{Z}$  est un « processus autorégressif moyenne mobile intégré d'ordre  $(p, d, q)$  » si elle est définie, pour tout  $t \in \mathbb{Z}$ , par*

$$(1 - L)^d \mathcal{A}(L) X_t = \mu + \mathcal{B}(L) \varepsilon_t$$

où  $(\varepsilon_t)$  est un bruit blanc de variance  $\sigma^2$ ,  $a_p \neq 0$  et  $b_q \neq 0$ .

La notation usuelle des autorégressifs moyennes mobiles intégrés est ARIMA, de l'anglais *autoregressive integrated moving average*. Pour signifier que le processus

$(X_t)$  est un autorégressif moyenne mobile intégré d'ordre  $(p, d, q)$ , nous noterons

$$(X_t) \sim \text{ARIMA}(p, d, q).$$

Littéralement, cela signifie que le processus possède  $d$  racines valant précisément 1, et qu'il est nécessaire de différencier le processus  $d$  fois pour se ramener à un  $\text{ARMA}(p, q)$  stationnaire. Il est donc clair qu'un processus  $\text{ARIMA}(p, d, q)$  **n'est pas stationnaire** dès que  $d > 0$ .

**Remarque 4.2.1** *Un  $\text{ARIMA}(p, 0, q)$  est un  $\text{ARMA}(p, q)$ .*

À titre d'exemple, le processus défini, pour tout  $t \in \mathbb{Z}$ , par

$$X_t = \frac{3}{2}X_{t-1} - \frac{1}{2}X_{t-2} + \varepsilon_t - \frac{1}{4}\varepsilon_{t-1}$$

où  $(\varepsilon_t)$  est un bruit blanc, est un  $\text{ARMA}(2,1)$  non stationnaire, car on voit que son polynôme autorégressif possède 1 et 2 comme racines. Il se factorise donc sous la forme

$$(1 - L)\left(1 - \frac{L}{2}\right)X_t = \left(1 - \frac{L}{4}\right)\varepsilon_t$$

et l'on en déduit que  $(X_t)$  est en fait un  $\text{ARIMA}(1,1,1)$  non stationnaire. Finalement, stationnariser  $(X_t)$  revient à poser, pour tout  $t \in \mathbb{Z}$ ,

$$Y_t = (1 - L)X_t = \Delta X_t \quad \text{et donc} \quad \left(1 - \frac{L}{2}\right)Y_t = \left(1 - \frac{L}{4}\right)\varepsilon_t$$

car l'on voit bien alors que  $(Y_t)$  est un  $\text{ARMA}(1,1)$  stationnaire.

**Corollaire 4.2.1** *Soit  $(X_t)$  un processus  $\text{ARIMA}(p, d, q)$  avec  $d > 0$ . Alors,  $(X_t)$  n'est pas stationnaire sur  $\mathbb{Z}$  et le processus  $(Y_t)$  défini, pour tout  $t \in \mathbb{Z}$ , par*

$$Y_t = \Delta^d X_t$$

*est un processus  $\text{ARMA}(p, q)$  stationnaire.*

**Démonstration.** La démonstration est évidente en factorisant  $(X_t)$  par  $(1 - L)^d$ .  
□

## 4.2.2 Détecter la stationnarité : le test ADF

Étant donnée une trajectoire  $(X_t)$  sur l'intervalle des temps discret  $t \in \{1, \dots, n\}$ , détecter la stationnarité revient à tester

$$\mathcal{H}_0 : \text{“(}X_t\text{) est non stationnaire”} \quad \text{contre} \quad \mathcal{H}_1 : \text{“(}X_t\text{) est stationnaire”}$$

puisque l'on se rappelle (voir le Chapitre III de modélisation statistique) qu'une procédure de test statistique ne permet que d'accepter  $\mathcal{H}_1$ , le cas échéant, mais pas

d'accepter  $\mathcal{H}_0$ . Le test le plus connu à cet égard est le **test de Dickey-Fuller augmenté**, que l'on note traditionnellement « test ADF ». Pour résumer grossièrement (car tout cela est très complexe), la procédure ADF teste si l'estimateur de l'une des racines du polynôme autorégressif est significativement proche de 1. Si tel est le cas, on ne rejette pas  $\mathcal{H}_0$  : on admet qu'il existe une racine sur le cercle unité et le processus est un ARIMA non stationnaire. On dit encore qu'il possède **une racine unitaire**. Le cas contraire, on se place sous  $\mathcal{H}_1$  : on admet que toutes les racines sont à l'extérieur strict du cercle unité, le processus est un ARMA stationnaire.

### 4.2.3 Détecter la non stationnarité : le test KPSS

Pour renforcer la conclusion tirée du test ADF, il convient de mettre en pratique un test complémentaire de détection de la non stationnarité. Cela revient cette fois à tester

$$\mathcal{H}_0 : "(X_t) \text{ est stationnaire}" \quad \text{contre} \quad \mathcal{H}_1 : "(X_t) \text{ est non stationnaire}."$$

Le test le plus connu à cet égard est le **test de Kwiatkowski-Phillips-Schmidt-Shin**, que l'on note traditionnellement « test KPSS ». Pour résumer encore plus grossièrement que pour le test ADF, il s'agit ici de tester si le résidu de la modélisation ARMA se comporte comme un bruit blanc (auquel cas on se place sous  $\mathcal{H}_0$  : on admet que  $(X_t)$  est un ARMA stationnaire), ou bien s'il se comporte comme une marche aléatoire (auquel cas on se place sous  $\mathcal{H}_1$  : on admet que  $(X_t)$  est un ARIMA non stationnaire). En effet, on se souvient que, pour tout  $t \in \mathbb{N}^*$ , une marche aléatoire  $(S_t)$  s'écrit

$$S_t = S_{t-1} + \eta_t$$

où  $(\eta_t)$  est un bruit blanc de variance  $\sigma^2$  indépendant de  $S_0$ . Il vient ainsi

$$S_t = S_0 + \sum_{k=1}^t \eta_k \quad \text{et} \quad \mathbb{V}(S_t) = \mathbb{V}(S_0) + \sigma^2 t.$$

Cela confirme qu'un résidu se comportant comme une marche aléatoire engendre une non stationnarité. De plus,

$$\Delta S_t = (1 - L)S_t = S_t - S_{t-1} = \eta_t.$$

Nous voyons ainsi que, sous l'alternative où le résidu se comporte comme une marche aléatoire, il existe une racine unitaire qu'une simple différenciation permet d'éliminer pour se ramener à un bruit blanc  $(\eta_t)$ . Les choses sont bien plus complexes au sein du test KPSS, nous avons simplement donné ici le principe sous-jacent.

### 4.2.4 En résumé...

Concluons ce chapitre en proposant un protocole expérimental permettant de décider si un processus observé est stationnaire.

1. On considère une série chronologique  $(X_t)$  observée sur  $t \in \{1, \dots, n\}$  que l'on souhaite modéliser par un ARMA.
2. On effectue le test ADF ainsi que le test KPSS sur  $(X_t)$ .
  - (a) On retient  $\mathcal{H}_0(\text{ADF})$  et  $\mathcal{H}_1(\text{KPSS})$ .
    - i. On admet que le processus n'est pas stationnaire et qu'il possède une racine unitaire.
    - ii. On crée le processus  $(Y_t = \Delta X_t)$  et l'on recommence la procédure appliquée cette fois au processus  $(Y_t)$ .
  - (b) On retient  $\mathcal{H}_1(\text{ADF})$  et  $\mathcal{H}_0(\text{KPSS})$ .
    - i. On admet que le processus est stationnaire et qu'il ne possède pas de racine unitaire.
  - (c) Les 2 tests se contredisent.
    - i. On en cherche les raisons "à la main" (voir ci-dessous).
3. Une fois le processus stationnarisé, on le traite comme un ARMA et l'on estime ses paramètres.

Le nombre de passages par l'étape (2.a) permet d'évaluer le paramètre  $d$  dans la modélisation ARIMA : c'est le nombre de racines unitaires à éliminer par différenciation. Par ailleurs, l'étape (2.c) peut se rencontrer en pratique. Lorsque les 2 tests se contredisent (de manière informelle, l'un suggère la stationnarité pendant que l'autre suggère la non stationnarité), les raisons peuvent en être multiples : il existe une racine supérieure à 1 mais très proche de 1 (ce qui peut tromper la procédure ADF), le processus n'est pas observé sur un intervalle de temps assez long (le comportement asymptotique propre aux statistiques de tests n'est pas atteint), l'influence des valeurs initiales n'est pas estompée, etc. Dans ce cas, une solution est d'estimer les paramètres avec  $d = 0$  puis avec  $d = 1$ , et de comparer l'ARMA et l'ARIMA obtenus sur la base d'un critère de modélisation (log-vraisemblance par exemple, AIC, SSR, etc.)

**Définition 4.2.2** On appelle « critère d'information d'Akaike » et l'on note « AIC » le critère de modélisation bayésien défini par

$$\text{AIC} = -2 \log \mathcal{L} + 2k$$

où  $\mathcal{L}$  est la vraisemblance de la modélisation et  $k$  est le nombre de paramètres estimés.

**Définition 4.2.3** On appelle « critère d'information bayésien » et l'on note « BIC » le critère de modélisation bayésien défini par

$$\text{BIC} = -2 \log \mathcal{L} + k \log n$$

où  $\mathcal{L}$  est la vraisemblance de la modélisation,  $k$  est le nombre de paramètres estimés et  $n$  est la taille de l'échantillon.

En pratique, on préfère souvent utiliser les critères AIC et BIC (que l'on note parfois aussi « SBC » pour *Schwarz bayesian criterion*) pour comparer deux modélisations : celle minimisant un tel critère maximise la log-vraisemblance tout en étant pénalisée par le nombre de paramètres à estimer, source de variabilité. Ainsi, **les critères AIC et BIC satisfont le principe de parcimonie.**