

# Séries temporelles

Adrien Hardy\*

30 mars 2020

*Ce document, **en construction**, contient les notes du cours “Series temporelles” donné au semestre 2 du Master Mathématiques pour la finance à l’Université de Lille. Il est calibré pour un volume horaire de  $10 \times 1h30$  de cours, complété  $5 \times 3h$  de TP sur machines. Toute remarque et aide au débusquage de coquilles est bienvenu.*

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Décomposition canonique</b>	<b>2</b>
2.1	Décomposition additive . . . . .	2
2.2	Décomposition en pratique (ce que fait R) . . . . .	3
2.3	Autres outils : Opérateurs de différence et régression paramétrique. . . . .	4
<b>3</b>	<b>Processus stationnaires</b>	<b>6</b>
3.1	Stationnarité . . . . .	6
3.2	Auto-corrélation empirique et tests de blancheur . . . . .	9
3.3	Séries temporelles linéaires . . . . .	11
3.4	Modèles ARMA . . . . .	13
3.4.1	Moving average . . . . .	13
3.4.2	Autoregressive . . . . .	13
3.4.3	Autoregressive + Moving average . . . . .	14
3.5	Interlude mathématique : Projection orthogonale . . . . .	15
3.6	Decomposition de Wold . . . . .	15
<b>4</b>	<b>Prédiction linéaire</b>	<b>17</b>
4.1	Prédicteur linéaire optimal . . . . .	17
4.2	Equation de Yule-Walker . . . . .	18
4.3	Fonctions d’autocorrélation partielle . . . . .	20
4.4	Prédicteurs à horizon $h$ (Exercice) . . . . .	20
4.5	Calcul effectif de l’inverse $\Gamma_m^{-1}$ . . . . .	20

---

\*Laboratoire Paul Painlevé, Université de Lille, Cité Scientifique, 59655 Villeneuve d’Ascq Cedex, France. Email: [adrien.hardy@univ-lille.fr](mailto:adrien.hardy@univ-lille.fr)

<b>5</b>	<b>Prédiction par lissage</b>	<b>25</b>
5.1	Lissage exponentiel simple . . . . .	25
5.2	Lissage exponentiel double . . . . .	27
<b>6</b>	<b>Modèles de volatilité</b>	<b>28</b>
6.1	Introduction . . . . .	28
6.2	Interlude mathématique : Espérance conditionnelle (rappels) . . . . .	28
6.3	Le cadre général . . . . .	30
6.4	Modèles ARCH & GARCH . . . . .	31

## 1 Introduction

Imaginez que l'on mesure la température mensuelle moyenne en France pendant 53 ans. On obtient alors une *série temporelle*  $x_1, x_2, \dots, x_{53}$  où  $x_j$  représente la température, mettons, en degrés Celsius, du  $j$ -ième mois. Une série temporelle  $x_1, \dots, x_N$  représente plus généralement  $N$  mesures d'un phénomène quantitatif à différents temps *équidistants*. Un exemple typique en mathématiques financières serait le cours d'une action ou la valeurs d'un indice économique (CAC, DAX, ...). Les objectifs sont alors de faire de la :

- *prédiction* : obtenir un prédicteur de  $x_{N+1}$ , de  $x_{N+2}$ , etc.
- *modélisation* : décrire les facteurs explicatifs via des modèles théoriques.

Pour ce faire, l'idée générale est de partir du postulat que  $x_1, \dots, x_n$  est une réalisation d'un processus stochastique  $(X_t)_{t \in \mathbb{Z}}$  et d'étudier mathématiquement ces processus. Le but de ce cours est d'étudier différents modèles classiques de tels processus et d'appliquer ces modèles à de vraies données à l'aide de R.

## 2 Décomposition canonique

Ayant sous la main un jeu de données  $x_1, \dots, x_n$ , on commence ♠ toujours par tracer son *chronogramme*, c'est-à-dire la graphe de  $t \mapsto x_t$  pour avoir une idée générale de la série temporelle. On voit souvent apparaître des comportements déterministes, par exemple une tendance croissante des températures en cas de réchauffement climatique et une composante périodique due aux saisons, auxquelles se rajoute un comportement "aléatoire", qui semble peu prévisible, souvent dû à la complexité des systèmes étudiés.

### 2.1 Décomposition additive

Mathématiquement on pourrait décomposer  $X_t$  de la façon suivante (décomposition additive) :

$$X_t = m_t + s_t + Z_t$$

où le *bruit*  $Z_t$  est la composante aléatoire de la série qui satisfait  $\mathbb{E}[Z_t] = 0$  pour tout  $t \in \mathbb{Z}$ . La *tendance*  $m_t$  est déterministe et la *saisonnalité*  $s_t$  est également déterministe mais satisfait les conditions suivantes :

- il existe un entier  $T > 0$ , appelé la *période*, tel que  $s_{t+T} = s_t$  pour tout  $t \in \mathbb{Z}$ .
- $s_1 + s_2 + \dots + s_T = 0$  (la moyenne sur une période est nulle).

---

♠. Enfin presque, on doit souvent effectuer un prétraitement du jeu de données, comme importer/fusionner/transformer des données (avec R, SQL, etc) et gérer le problème des valeurs manquantes.

Cette deuxième condition est là pour fixer un degré de liberté dans la définition de  $m_t$  et  $s_t$  : sans cette condition, pour toute constante  $c$  on voit que les couples  $(m_t, s_t)$  et  $(m_t+c, s_t-c)$  satisfont tous les deux les conditions imposés sur la tendance et la saisonnalité et il y aurait alors ambiguïté dans la définition.

Pour une réalisation  $x_t$  on aura donc  $x_t = m_t + s_t + z_t$ , où  $z_t$  est une réalisation de  $Z_t$ .

## ILLUSTRATIONS

**Remarque 2.1.** On pourrait aussi utiliser une décomposition multiplicative :

$$X_t = m_t s_t Z_t$$

où  $m_t, s_t, Z_t$  sont comme au dessus sauf que cette fois  $\mathbb{E}[Z_t] = 1$  et  $s_1 \cdots s_T = 1$ . Cette décomposition est intéressante lorsque l'on est en présence d'une série dont l'amplitude varie fortement avec le temps, ce qui n'est pas capturé par la décomposition additive. Notez que si la série prend des valeurs positives (ce que l'on peut souvent supposer quitte à rajouter une constante assez grande), on peut ramener une décomposition multiplicative à une décomposition additive en prenant le logarithme de la série.

## ILLUSTRATIONS

### 2.2 Décomposition en pratique (ce que fait R)

En pratique, il faut commencer par identifier la période  $T$  de la composante saisonnière. C'est souvent évident suivant le jeu de données ( $T = 12$  si séries mensuelles, autour de  $T = 253$  si série journalière basée sur les jours ouvrés  $\heartsuit$ , etc). Sinon on peut obtenir  $T$  à l'oeil nu ou comme maximum du périodogramme  $\clubsuit$  de la série.

Sous R, la période doit être indiquée dès la construction de type série temporelle (`ts`, pour time series). Ensuite, on peut demander à R d'appliquer la fonction `decompose` à la série temporelle. Voilà ce que `R` retourne :

**Tendance approchée.** Si  $T = 2m + 1$  est impaire, alors on prend

$$\hat{m}_t := \frac{1}{2m + 1} \sum_{k=-m}^m x_{t+k},$$

c'est-à-dire la moyenne de la série sur une fenêtre de taille  $T$  centrée en  $x_t$ . On parle de *lissage* ou de *convolution discrète*. De façon peu rigoureuse, l'idée est que

$$\begin{aligned} \hat{m}_t &= \frac{1}{T} \sum_{k=-m}^m x_{t+k} \\ &= \underbrace{\frac{1}{2m + 1} \sum_{k=-m}^m m_{t+k}}_{\simeq m_t} + \underbrace{\frac{1}{2m + 1} \sum_{k=-m}^m s_{t+k}}_{=0} + \underbrace{\frac{1}{2m + 1} \sum_{k=-m}^m z_{t+k}}_{\simeq \mathbb{E}[Z_t]=0} \end{aligned}$$

---

$\heartsuit$ . En tous cas au USA (trading days). En effet, le NYSE et le NASDAC estiment qu'il y a en moyenne 253 jours ouvrés annuels :  $(365.25 \text{ jours moyens par année}) \times 5/7$  (proportion de jours de travail par semaine) = 260.89 - (6 jours fériés hors WE) -  $(3 \times 5/7 \text{ jours fériés à date fixe}) = 252.75 \sim 253$ .

$\clubsuit$ . C'est le module carré de la transformée de Fourier (discrète) de la série temporelle. On la trouve sur R avec la fonction "periodogram" du package TSA.

où la première approximation est raisonnable si  $t \mapsto m_t$  est à variations lentes alors que la deuxième invoque, au moins quand  $m$  est suffisamment grand, une intuition de type loi des grands nombres (ou le théorème ergodique de Birkhoff, qui sera souvent vérifié par les processus stationnaires que nous étudierons plus tard.)

Si  $T = 2m$  est impaire, on fait une construction similaire en posant

$$\hat{m}_t := \frac{1}{2m} \left( \frac{1}{2} x_{t-m} + \sum_{k=-m+1}^{m-1} x_{t+k} + \frac{1}{2} x_{t+m} \right).$$

**Saisonnalité approchée.** L'idée est que la série  $x_t - \hat{m}_t$  serait une approximation raisonnable de la saisonnalité, pourvu qu'elle soit périodique. Pour forcer cette périodicité, on pose alors

$$\hat{s}_t := \frac{1}{N_t} \sum_{k=t \bmod T} x_k - \hat{m}_k$$

où  $N_t := \#\{k \in \{1, \dots, n\} : k = t \bmod T\}$ . Par construction, on a bien  $\hat{s}_{t+T} = \hat{s}_t$  pour tout  $t \in \{1, \dots, n - T\}$ . Par exemple, si  $x_t$  est une série mensuelle est  $T = 12$ ,  $\hat{s}_1$  est la moyenne sur tous les mois de janvier  $k$  de la série  $x_k - \hat{m}_k$ . On a également, avec l'approximation  $N_t \simeq n/T$ ,

$$\sum_{t=1}^T \hat{s}_t = \frac{T}{n} \sum_{t=1}^n (x_t - \hat{m}_t) \simeq T \mathbb{E}[Z_t] = 0.$$

Cependant, cette approximation est rarement juste en pratique car la multiplication par  $T$  de l'erreur peut engendrer une erreur non-négligeable dès que  $T$  est grand.

**Bruit approché.** Finalement, on pose  $\hat{z}_t := x_t - \hat{m}_t - \hat{s}_t$ , qui sera l'approximation de la partie aléatoire de la série temporelle. En conclusion on voit qu'on peut, au moins de façon approchée, extraire la partie déterministe (tendance + saisonnalité) d'une série temporelle et donc réduire l'étude à celle de la partie aléatoire. L'hypothèse fondamentale qu'on fera dans la suite est cette partie aléatoire forme un *processus stationnaire*, et elle n'est raisonnable que si l'espérance de la série est constante.

*A partir de maintenant, on s'intéresse à des séries temporelles ayant une tendance et une saisonnalité nulle, ce qui n'est pas une perte de généralité quitte à avoir effectué au préalable une décomposition comme expliqué dans cette section.*

### 2.3 Autres outils : Opérateurs de différence et régression paramétrique.

On considère l'*opérateur de retard*  $B$  qui agit sur une suite  $(X_t)$  par  $BX_t := X_{t-1}$  et l'*opérateur de différence*  $\Delta = I - B$  défini par  $\Delta X_t = X_t - X_{t-1}$ . On considèrera aussi l'*opérateur de différence saisonnier*  $\Delta_T := I - B^T$ , ce qui donne  $\Delta_T X_t = X_t - X_{t-T}$ . Il est alors facile de vérifier (Exercice) que :

— Si  $t \mapsto m_t$  est un polynôme de degré  $d$ , alors pour tout  $t$ ,

$$\Delta^{d+1} m_t := \underbrace{\Delta \circ \dots \circ \Delta}_{d+1} m_t = 0.$$

— Si  $s_{t+T} = s_t$ , alors  $\Delta_T s_t = 0$  pour tout  $t$ .

En particulier, si on parie que la tendance est un polynôme de degré  $d$  et la saisonnalité est nulle, on peut vérifier que  $\Delta_T \circ \Delta^d x_t = 0$ . Si c'est le cas, alternativement à l'approche par lissage, on peut effectuer une régression polynomiale à  $x_t$  et obtenir le bruit  $\hat{z}_t$  comme les résidus de cette régression.

### 3 Processus stationnaires

Dans ce cours, un *processus stochastique* est une famille  $(X_t)_{t \in \mathbb{Z}}$  de variables aléatoires à valeurs dans  $\mathbb{R}$  définie sur le même espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$ . Pour un  $\omega \in \Omega$ , la fonction  $t \mapsto x_t := X_t(\omega)$  représente une réalisation de la *trajectoire* de ce processus. On observe en pratique un sous-ensemble d'une réalisation, à savoir  $x_1, \dots, x_n$ . Le but du jeu est de modéliser au mieux  $X_t$  à l'aide de l'information obtenue du jeu de données. Cela a des applications concrètes, comme par exemple construire des prédicteurs  $\hat{x}_{t+1}$  pour la réalisation  $x_{t+1}$  (non observée).

**L'espace  $L^2$ .** On notera  $L^2 := L^2(\Omega, \mathcal{F}, \mathbb{P})$ , l'espace de Hilbert des variables aléatoires de carré intégrable muni du produit scalaire  $\langle X, Y \rangle := \mathbb{E}[XY]$  et de norme  $\|X\|_{L^2} := \sqrt{\mathbb{E}[X^2]}$ . Dans  $L^2$ , on identifie deux variables aléatoires égales presque partout :  $X = Y$  dès que  $\mathbb{P}(X = Y) = 1$ . L'inégalité de Cauchy-Schwarz, très utile dans ce qui suit, qui dit que pour deux variables réelles  $X, Y \in L^2$  on a

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]} \sqrt{\mathbb{E}[Y^2]}.$$

Ici, on supposera toujours que les processus stochastiques sont dans  $L^2$ , c'est-à-dire :  $\mathbb{E}[X_t^2] < \infty$  pour tout  $t \in \mathbb{Z}$ . L'inégalité de Cauchy-Schwarz implique alors (Exercice) que l'espérance  $\mathbb{E}[X_t]$  et la covariance  $\text{Cov}(X_t, X_s) := \mathbb{E}[X_t X_s] - \mathbb{E}[X_t] \mathbb{E}[X_s]$  sont bien définies (et finies) pour  $t, s \in \mathbb{Z}$ .

#### 3.1 Stationnarité

On dira qu'un processus est (*faiblement*) *stationnaire* si, pour tout  $h \geq 1$  et tout  $t_1, \dots, t_n \in \mathbb{Z}$ , les vecteurs aléatoires

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

ont les mêmes espérances et matrices de covariance : pour tout  $t, s, h \in \mathbb{Z}$ ,

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

De façon équivalente,  $\mathbb{E}[X_t]$  ne dépend pas de  $t$  et  $\text{Cov}(X_{t+h}, X_t)$  ne dépend que de  $h$ .

Il est *fortement stationnaire* si pour tout  $h \geq 1$  et tout  $t_1, \dots, t_n \in \mathbb{Z}$ , les vecteurs aléatoires

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

ont même loi. Fortement stationnaire implique stationnaire mais la réciproque est fautive.

**Exemple 3.1** (Bruit blanc). Si  $(X_t)_{t \in \mathbb{Z}}$  est une suite de variables aléatoires de même loi telle que  $\mathbb{E}[X_t^2] < \infty$  et  $\text{Cov}(X_t, X_s) = 0$  pour tout  $t \neq s$ , alors c'est un processus stationnaire. On dit que c'est un *bruit blanc (faible)* et on écrit alors  $(X_t)_{t \in \mathbb{Z}} \sim \text{BB}(\mu, \sigma^2)$  où  $\mu = \mathbb{E}[X_0]$  et  $\sigma^2 = \text{Var}(X_0)$ . Si  $(X_t)_{t \in \mathbb{Z}}$  est une suite de variables aléatoires *i.i.d* telle que  $\mathbb{E}[X_0^2] < \infty$ , alors c'est un processus fortement stationnaire ; c'est un *bruit blanc fort*.

La *fonction d'autocovariance*  $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$  d'un processus stationnaire est définie par

$$\gamma(h) := \text{Cov}(X_{t+h}, X_t)$$

et sa *fonction d'autocorrélation*  $\rho : \mathbb{Z} \rightarrow [-1, 1]$  par

$$\rho(h) := \frac{\text{Cov}(X_{t+h}, X_t)}{\sqrt{\text{Var}(X_{t+h})\text{Var}(X_t)}} = \frac{\gamma(h)}{\gamma(0)}.$$

Ainsi,  $\rho(h)$  décrit comment deux variables de la trajectoire à distance temporelle  $h$  sont corrélées. Par exemple, si  $(X_t)_{t \in \mathbb{Z}}$  est un bruit blanc, alors  $\rho(h) = \mathbf{1}_{h=0}$ . Notez qu'on a toujours (Exercice) :

- (a)  $\rho(0) = 1$ .
- (b)  $\rho(-h) = \rho(h)$  pour tout  $h \in \mathbb{Z}$ .
- (c) Pour tout  $t_1, \dots, t_n \in \mathbb{Z}$ , la matrice

$$\Gamma_n := \left[ \gamma(t_i - t_j) \right]_{i,j=1}^n$$

est symétrique semi-définie positive.

Réciproquement, on peut montrer que si une fonction  $\rho : \mathbb{Z} \rightarrow [-1, 1]$  satisfait (a), (b) et (c) alors il existe  $\heartsuit$  un processus stochastique stationnaire de fonction d'autocorrélation  $\rho$ .

**Exemple 3.2** (Moving average). Soit  $\varepsilon_t$  un bruit blanc  $BB(0, \sigma^2)$ . On construit alors, pour un paramètre  $\theta \in \mathbb{R}$ ,

$$X_t := \varepsilon_t + \theta \varepsilon_{t-1}.$$

C'est un exemple de processus MA (moving average) à  $q = 1$  paramètres. On calcule que  $\mathbb{E}[X_t] = 0$  et

$$\text{Cov}(X_{t+h}, X_t) = \mathbb{E}[X_{t+h}X_t] = \begin{cases} (1 + \theta^2)\sigma^2 & \text{si } h = 0 \\ \theta\sigma^2 & \text{si } |h| = 1 \\ 0 & \text{sinon.} \end{cases}$$

C'est donc bien un processus stationnaire de fonction d'autocorrélation

$$\rho(h) = \begin{cases} 1 & \text{si } h = 0 \\ \frac{\theta}{1+\theta^2} & \text{si } |h| = 1 \\ 0 & \text{sinon.} \end{cases}$$

**Exemple 3.3** (Autoregressive). Soit  $(\varepsilon_t)$  un bruit blanc  $BB(0, \sigma^2)$ . On considère alors, pour un paramètre  $\varphi \in ]-1, 1[$ , le processus défini par le système d'équations

$$X_t = \varphi X_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}.$$

Il n'est pas évident qu'un tel processus existe. Pour s'en convaincre, on écrit d'abord que pour tout  $N \geq 1$ ,

$$\begin{aligned} X_t &= \varphi X_{t-1} + \varepsilon_t \\ &= \varphi^2 X_{t-2} + \varphi \varepsilon_{t-1} + \varepsilon_t \\ &\vdots \\ &= \varphi^{N+1} X_{t-N-1} + \sum_{k=0}^N \varphi^k \varepsilon_{t-k}. \end{aligned}$$

---

$\heartsuit$ . L'idée est de construire un vecteur gaussien  $(X_{-n}, \dots, X_n)$  centré de dimension  $2n$  de matrice de covariance  $\Gamma_{2n}$  (par exemple via la décomposition de Choleski) et de faire tendre  $n$  vers l'infini.

Cela nous encourage à poser,

$$X_t := \sum_{k=0}^{\infty} \varphi^k \varepsilon_{t-k}, \quad t \in \mathbb{Z}.$$

Si on peut montrer que cette série est convergente dans  $L^2$ , alors il est clair que  $X_t$  satisfait l'équation au-dessus. Montrons cette convergence : En utilisant l'inégalité de Cauchy-Schwarz on a

$$\mathbb{E}|\varepsilon_{t-k}||\varepsilon_{t-\ell}| \leq \sqrt{\mathbb{E}(\varepsilon_{t-k}^2)\mathbb{E}(\varepsilon_{t-\ell}^2)} = \sigma^2$$

et donc

$$\sum_{k,\ell=0}^{\infty} |\varphi|^{k+\ell} \mathbb{E}|\varepsilon_{t-k}||\varepsilon_{t-\ell}| \leq \sigma^2 \sum_{k,\ell=0}^{\infty} |\varphi|^{k+\ell} = \frac{\sigma^2}{(1-|\varphi|)^2} < \infty.$$

Le théorème de Fubini  $\clubsuit$  implique alors que

$$\mathbb{E}[X_t^2] = \mathbb{E} \sum_{k,\ell=0}^{\infty} \varphi^{k+\ell} \varepsilon_{t-k} \varepsilon_{t-\ell} = \sum_{k,\ell=0}^{\infty} \varphi^{k+\ell} \underbrace{\mathbb{E}(\varepsilon_{t-k} \varepsilon_{t-\ell})}_{\sigma^2 \mathbf{1}_{k=\ell}} = \frac{\sigma^2}{1-\varphi^2} < \infty.$$

Le même type d'argument montre que

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

et, si  $h \geq 0$ ,

$$\text{Cov}(X_{t+h}, X_t) = \mathbb{E} \sum_{k,\ell=0}^{\infty} \varphi^{k+\ell} \varepsilon_{t+h-k} \varepsilon_{t-\ell} = \sum_{k,\ell=0}^{\infty} \varphi^{k+\ell} \underbrace{\mathbb{E}(\varepsilon_{t+h-k} \varepsilon_{t-\ell})}_{\sigma^2 \mathbf{1}_{k=\ell+h}} = \frac{\sigma^2 \varphi^h}{1-\varphi^2} < \infty$$

et donc, pour  $h \in \mathbb{Z}$ ,

$$\gamma(h) = \frac{\sigma^2 \varphi^{|h|}}{1-\varphi^2}.$$

C'est donc un processus stationnaire, qui est un exemple de processus autorégressif (AR) à  $p = 1$  paramètres, de fonction d'auto-corrélation :

$$\rho(h) = \varphi^{|h|}.$$

**Remarque 3.4.** De façon générale, si un processus  $(Y_k)$  est tel que  $\sum_{k=0}^{\infty} Y_k \in L^2$ , alors

$$\mathbb{E} \left[ \sum_{k=0}^{\infty} Y_k \right] = \sum_{k=0}^{\infty} \mathbb{E}[Y_k].$$

---

$\clubsuit$ . On utilise ici la version suivante : Soit  $(Y_i)_{i \in I}$  est une suite de variables aléatoires indexées par un ensemble dénombrable  $I$ . Si  $\sum_{i \in I} \mathbb{E}|Y_i| < \infty$ , alors  $\sum_{i \in I} \mathbb{E}[Y_i] = \mathbb{E}[\sum_{i \in I} Y_i]$  et cette quantité est finie.



En effet, comme  $X, Y \mapsto \langle X, Y \rangle$  est continue dans  $L^2$  (Exercice), on a :

$$\begin{aligned} \mathbb{E} \left[ \sum_{k=0}^{\infty} Y_k \right] &= \langle 1, \sum_{k=0}^{\infty} Y_k \rangle \\ &= \langle 1, \lim_{N \rightarrow \infty} \sum_{k=0}^N Y_k \rangle \\ &= \lim_{N \rightarrow \infty} \langle 1, \sum_{k=0}^N Y_k \rangle \\ &= \lim_{N \rightarrow \infty} \sum_{k=0}^N \langle 1, Y_k \rangle = \sum_{k=0}^{\infty} \mathbb{E}[Y_k]. \end{aligned}$$

### 3.2 Auto-corrélation empirique et tests de blancheur

En pratique on a seulement à disposition des observation  $x_1, \dots, x_n$  (c'est-à-dire des réalisations de  $X_1, \dots, X_n$ ) pour un certain  $n \geq 1$ . On construit alors les estimateurs  $\heartsuit$

$$\hat{\gamma}_n(h) := \frac{1}{n} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x}_n)(x_t - \bar{x}_n) \quad \text{et} \quad \hat{\rho}_n(h) := \frac{\hat{\gamma}_n(h)}{\hat{\gamma}_n(0)}$$

où on a introduit la moyenne empirique

$$\bar{x}_n := \frac{1}{n} \sum_{i=1}^n x_i.$$

Sous  $\mathbb{R}$ , on obtient  $\hat{\rho}_n(h)$  via la commande `acf` (pour auto-corrélation function) et l'option `lag=h`. Pour les modèles qui nous intéressent dans la suite, on a en général la loi des grands nombres  $\spadesuit$  : pour tout  $h \in \mathbb{Z}$ ,

$$\hat{\rho}_n(h) \xrightarrow[n \rightarrow \infty]{} \rho(h).$$

De plus, si  $(X_t)_{t \in \mathbb{Z}}$  est un bruit blanc fort, en particulier  $\rho(h) = 0$  pour tout  $|h| \neq 0$ , et que  $\mathbb{E}[X_t^4] < \infty$ , alors le théorème central limite implique : pour tout  $h \geq 1$ ,

$$\sqrt{n} \begin{bmatrix} \hat{\rho}_n(1) \\ \vdots \\ \hat{\rho}_n(h) \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, I_h).$$

En particulier, sous l'hypothèse

$$H_0 := \{(X_t)_{t \in \mathbb{Z}} \text{ est un bruit blanc fort}\}$$

on a (pourquoi ?)

$$\xi_h := n \sum_{i=1}^h \hat{\rho}_n(i)^2 \xrightarrow[n \rightarrow \infty]{\text{loi}} \chi^2(h)$$

$\heartsuit$ . On trouve aussi dans la littérature  $\hat{\rho}_n(h)$  défini avec un facteur multiplicatif  $\frac{1}{n-h}$  plutôt que  $\frac{1}{n}$ . Les résultats qui suivent et les interprétations restent cependant inchangés.

$\spadesuit$ . C'est en particulier vrai dès que  $X_t$  est un processus stationnaire fort et ergodique, ce qui est le cas dès que  $\sum_{h \geq 0} |\gamma(h)| < \infty$ .

où  $\chi^2(h)$  est une variable aléatoire qui suit une loi du Chi-deux à  $h$  degrés de liberté.

Si à l'inverse il existe  $i \in \{1, \dots, h\}$  tel que  $\rho(i) \neq 0$  alors  $\xi_h \rightarrow \infty$  quand  $n \rightarrow \infty$  en probabilités. Ceci permet de construire un test statistique *asymptotique* pour  $H_0$  appelé le *test de Box-Pierce* : Avec une erreur de première espèce  $\alpha \in ]0, 1[$  fixé (typiquement  $\alpha = 0.05$ ), si  $q_\alpha := F_{\chi^2(h)}^{-1}(1 - \alpha)$  est le quantile d'ordre  $1 - \alpha$  d'une variable  $X \sim \chi^2(h)$ , en d'autres termes tel que  $\mathbb{P}(X \geq q_\alpha) = \alpha$ , alors on a sous  $H_0$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\xi_h \geq q_\alpha) = \alpha.$$

En pratique, si la valeur numérique de  $\xi_h$  est plus grande que  $q_\alpha$  alors on rejète  $H_0$  pour l'hypothèse

$$H_1 = \{\text{il existe } i \in \{1, \dots, h\} \text{ tel que } \rho(i) \neq 0\}.$$

Sinon, si  $\xi_h \leq q_\alpha$ , on n'a pas d'évidence contre  $H_0$ . Il n'est donc pas déraisonnable de faire l'hypothèse  $H_0$ . Attention : on n'a jamais prouvé que  $H_0$  était vraie ! *Attention cependant au paramètre  $h$  : il se peut que l'évidence contre  $H_0$  n'apparaisse que pour certains  $h$  assez grand. Il convient donc d'effectuer ce test pour le plus de valeurs de  $h$  possibles.*

**p-valeur** : On retrouve ce test sous R via la commande `Box.test` qui renvoie alors la *p-valeur du test* définie par

$$\text{p-val} := \inf\{\alpha > 0 : \text{on rejète } H_0\}.$$

On voit que cette quantité est bien définie car  $q_\alpha \rightarrow \infty$  quand  $\alpha \rightarrow 0$ . Si  $\text{p-val} < \alpha$  alors on peut rejeter  $H_0$  au seuil d'erreur  $\alpha$ . Plus  $\text{p-val}$  est petite, plus on a d'évidence contre  $H_0$ . Attention cependant : si  $H_0$  est vraie, alors la variable aléatoire  $\text{p-val}$  (c'est une variable aléatoire avant qu'on ait collecté les données et calculé  $\xi$ ) suit une loi uniforme sur  $[0, 1]$ . En particulier, avoir une p-valeur proche de 1 ne donne pas plus d'évidence pour  $H_0$  que quelconque autre valeur supérieure au seuil  $\alpha$  fixé.

**Remarque 3.5.** *Pour obtenir les résultats du test pour plusieurs  $h$  simultanément, on peut utiliser la commande `Box.test.2` du package `caschrono` de R.*

**Remarque 3.6.** *Un test alternatif est celui de Ljung-Box (aussi disponible dans les options de `Box.test`) basé sur le fait que, sous  $H_0$ ,*

$$\xi_h^{LB} := h(h+2) \sum_{i=1}^h \frac{\hat{\rho}_n(i)^2}{n-i} \xrightarrow[n \rightarrow \infty]{\text{loi}} \chi^2(h).$$

*Cette convergence vers  $\chi^2(h)$  semble (numériquement) plus rapide que celle de  $\xi_h^{LB}$  ; mais c'est essentiellement du tuning.*

**Remarque 3.7** (Les pointillés en bleu dans la fonction `acf` de R ?). *On vu au dessus que, sous  $H_0$ , on a  $\sqrt{n}\hat{\rho}_n(h) \rightarrow \mathcal{N}(0, 1)$  en loi quand  $n \rightarrow \infty$  pour tout  $h \geq 0$  fixé. Ainsi, si on prend  $q_{5\%} := F_{|\mathcal{N}(0,1)|}^{-1}(0.95) = F_{\mathcal{N}(0,1)}^{-1}(0.975) \simeq 1.96$  qui satisfait  $\mathbb{P}(X \in [-q_{5\%}, q_{5\%}]) = 5\%$  avec  $X \sim \mathcal{N}(0, 1)$  alors on a sous  $H_0$  :*

$$\mathbb{P}\left(\hat{\rho}_n(h) \in \left[-\frac{q_{5\%}}{\sqrt{n}}, \frac{q_{5\%}}{\sqrt{n}}\right]\right) \xrightarrow[n \rightarrow \infty]{} 5\%.$$

C'est justement cet intervalle  $[-\frac{95\%}{\sqrt{n}}, \frac{95\%}{\sqrt{n}}]$  qui est représenté verticalement par les pointillés bleu dans la sortie graphique de l'acf de  $R$ . Notez qu'il ne dépend pas de  $h$ . Une valeur  $\hat{\rho}_n(h)$  n'a donc asymptotiquement que 5% de chances de sortir de ces démarcations si  $H_0$  est vrai.

**Remarque 3.8.** Il est cependant difficile de tester la stationnarité à la seule vue du jeu de données  $x_1, \dots, x_n$ . C'est souvent une hypothèse de travail que l'on fait sans plus de vérifications mais il faut garder en tête que cela peut être une source d'erreurs.

Que faire si on rejète  $H_0$ ? Il faut des modèles plus sophistiqués de bruit.

### 3.3 Séries temporelles linéaires

On dit qu'une suite  $(\alpha_k)_{k \in \mathbb{Z}}$  est  $\ell^p$  si

$$\sum_{k \in \mathbb{Z}} |\alpha_k|^p < \infty.$$

Remarquons que  $\ell^1 \Rightarrow \ell^2$  mais que l'implication réciproque est fautive.

On dit que  $(X_t)$  est *linéaire* si il existe  $(\varepsilon_t) \sim \text{BB}(0, \sigma^2)$ , une suite  $(\alpha_k)_{k \in \mathbb{Z}} \in \ell^1$  et une constante  $m \in \mathbb{R}$  tels que (dans  $L^2$ )

$$X_t = \sum_{k \in \mathbb{Z}} \alpha_k \varepsilon_{t-k} + m. \quad (3.1)$$

Un série linéaire est automatiquement stationnaire et on a (Exercice) :

$$\mathbb{E}[X_t] = m \quad \text{Cov}(X_{t+h}, X_t) = \sigma^2 \sum_{k \in \mathbb{Z}} \alpha_k \alpha_{k+h}.$$

**Remarque 3.9.** Plus généralement, si  $Z_t$  est un processus stationnaire et  $(\alpha_k)_{k \in \mathbb{Z}}$  est  $\ell^1$ , alors

$$X_t := \sum_{k \in \mathbb{Z}} \alpha_k Z_{t-k}$$

définit un processus stationnaire de fonction d'autocorrélation (Exercice  $\heartsuit$ )

$$\gamma_X(h) := \sum_{k, \ell \in \mathbb{Z}} \alpha_k \alpha_\ell \gamma_Z(h + k - \ell).$$

Une telle série est dite :

- *causale* si  $\alpha_k = 0$  dès que  $k < 0$ ;  $X_t$  ne dépend pas de  $\varepsilon_{t+1}, \varepsilon_{t+2}, \dots$ , c'est-à-dire du "futur" de  $\varepsilon_t$ .
- *inversible* si  $(\varepsilon_t)$  est un processus *causal* de  $(X_t)$  : il existe  $(\beta_k) \in \ell^1$  tel que

$$\varepsilon_t = \sum_{k=0}^{\infty} \beta_k X_{t-k}.$$

**Exemple 3.10.** Retour à l'AR(1),  $X_t = \varphi X_{t-1} + \varepsilon_t$  avec  $(\varepsilon_t) \sim \text{BB}(0, \sigma^2)$  et  $\varphi \in \mathbb{R}$ .

---

$\heartsuit$ . Montrer cela quand  $\alpha_k$  est nul sauf pour un nombre fini de termes, puis prenez une limite pour traiter le cas général.

- Quand  $|\varphi| < 1$ , nous avons montré qu'il existait une solution stationnaire dans  $L^2$  donnée par

$$X_t = \sum_{k=0}^{\infty} \varphi^k \varepsilon_{t-k},$$

qui est donc causale. Comme par définition on a  $\varepsilon_t = X_t - \varphi X_{t-1}$ , le processus est également inversible.

- Quand  $|\varphi| > 1$ , on peut écrire la relation  $X_{t-1} = \frac{1}{\varphi} X_t - \frac{1}{\varphi} \varepsilon_t$ , et le même calcul que précédemment montre qu'il existe une solution stationnaire dans  $L^2$  donnée par

$$X_t = - \sum_{k=1}^{\infty} \left(\frac{1}{\varphi}\right)^k \varepsilon_{t+k-1}$$

de fonction d'auto-corrélation  $\rho(h) = (1/|\varphi|)^{|h|}$ . Ce processus n'est donc pas causal.

- Si  $|\varphi| = 1$ , supposons qu'une solution  $(X_t)$  stationnaire existe dans  $L^2$ . On a pour tout  $N \geq 1$ ,

$$X_t = \sum_{k=0}^N \varphi^k \varepsilon_{t-k} + \varphi^{N+1} X_{t-N-1}$$

et donc

$$\left\| X_t - \varphi^{N+1} X_{t-N-1} \right\|_{L^2}^2 = \left\| \sum_{k=0}^N \varphi^k \varepsilon_{t-k} \right\|_{L^2}^2 = \sigma^2 \sum_{k=0}^N \underbrace{\varphi^{2k}}_{=1} \xrightarrow{N \rightarrow \infty} \infty.$$

D'un autre côté, l'inégalité triangulaire nous donne

$$\left\| X_t - \varphi^{N+1} X_{t-N-1} \right\|_{L^2}^2 \leq \left( \left\| X_t \right\|_{L^2} + \underbrace{|\varphi|^{N+1}}_{=1} \left\| X_{t-N-1} \right\|_{L^2} \right)^2 = 4\mathbb{E}[X_0]^2 < \infty$$

ce qui nous mène à une contradiction : il n'existe pas de solution stationnaire dans  $L^2$  quand  $|\varphi| = 1$ .

**Notation avec l'opérateur de retard.** Si  $\alpha = (\alpha_k) \in \ell^1$ , on note

$$\alpha(z) := \sum_{k \in \mathbb{Z}} \alpha_k z^k$$

qui converge pour tout  $z \in \mathbb{C}$  tels que  $|z| = 1$  car  $(\alpha_k) \in \ell^1$  par hypothèse; en effet  $|\alpha(z)| \leq \sum_{k \in \mathbb{Z}} |\alpha_k| < \infty$ . Si on note formellement

$$\alpha(B) = \sum_{k \in \mathbb{Z}} \alpha_k B^k,$$

alors (3.1) s'écrit de façon concise  $X_t = \alpha(B)\varepsilon_t$ . Si on veut montrer que  $(X_t)$  est inversible, on cherche alors  $\beta = (\beta_k) \in \ell^1$  tel que  $\varepsilon_t = \beta(B)\varepsilon_t = \beta(B)\alpha(B)X_t$  et donc  $\beta(B)$  est l'opérateur inverse de  $\alpha(B)$  si il existe :  $\beta(B)\alpha(B) = I$ . Si on note  $\beta * \alpha$  la suite (Exercice : vérifier qu'elle est  $\ell^1$  !)

$$(\beta * \alpha)_k := \sum_{\ell \in \mathbb{Z}} \beta_\ell \alpha_{k-\ell}$$

on a alors  $\beta(B)\alpha(B) = (\beta * \alpha)(B)$  et trouver  $(\beta_k)$  revient à résoudre  $(\beta * \alpha)_k = \mathbf{1}_{k=0}$ .

### 3.4 Modèles ARMA

Dans cette section, on notera respectivement le disque et le cercle unité par

$$\mathbb{D} := \{z \in \mathbb{C} : |z| \leq 1\}, \quad \mathbb{S} := \{z \in \mathbb{C} : |z| = 1\}.$$

#### 3.4.1 Moving average

Les modèles  $MA(q)$  représentent la classe de processus stochastiques  $(X_t)$  définis par

$$X_t := \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

où  $(\varepsilon_t) \sim \text{BB}(0, \sigma^2)$  et qui dépend de  $q + 1$  paramètres réels  $\theta_1, \dots, \theta_q, \sigma^2$  libres. Ce sont des processus linéaires, donc stationnaires, qui sont clairement causals. On a  $\mathbb{E}[X_t] = 0$  et  $\rho(h) = 0$  dès que  $|h| > q$  (Exercice); on dit que c'est *un processus de portée  $q$* . Si on considère le polynôme

$$\Theta(z) := 1 + \theta_1 z + \cdots + \theta_q z^q,$$

on peut alors écrire

$$X_t = \Theta(B)\varepsilon_t.$$

On peut alors montrer le résultat suivant :

**Théorème 3.11.** *Un processus  $MA(q)$  est inversible  $\Leftrightarrow \Theta(z) \neq 0$  pour tout  $z \in \mathbb{D}$ .*

L'idée de la preuve est qu'inverser  $\Theta(B)$  revient à trouver une suite  $(\beta_k) \in \ell^1$  telle que

$$\frac{1}{\Theta(z)} = \sum_{k \in \mathbb{Z}} \beta_k z^k =: \beta(z)$$

pour tout  $z \in \mathbb{S}$ , ce qui est possible si et seulement  $\Theta(z) \neq 0$  pour  $z \in \mathbb{S}$ , et  $\beta(z)$  puisqu'on aura alors  $\Theta(B)\beta(B) = I$ . De plus,  $\Theta(z) \neq 0$  sur tout le disque  $\mathbb{D}$  si et seulement si  $\beta_k = 0$  pour tout  $k < 0$ .

#### 3.4.2 Autoregressive

Les modèles  $AR(p)$  représentent la classe de processus stochastiques  $(X_t)$  définis par comme les solutions *stationnaires* du système d'équations

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + \varepsilon_t \quad (\text{AR}(p))$$

où  $(\varepsilon_t) \sim \text{BB}(0, \sigma^2)$  et  $\varphi_1, \dots, \varphi_p, \sigma$  sont des paramètres réels libres, *pourvu que de telles solutions existent*.

Si on considère le polynôme

$$\Phi(z) := 1 - \varphi_1 z - \cdots - \varphi_p z^p,$$

alors on a

$$\Phi(B)X_t = \varepsilon_t.$$

Notez que si il existe,  $(X_t)$  est clairement inversible. On peut alors montrer le résultat suivant :

**Théorème 3.12.** — (AR(p)) admet une unique solution  $(X_t) \Leftrightarrow \Phi(z) \neq 0$  pour  $z \in \mathbb{S}$ .  
 —  $(X_t)$  est causal  $\Leftrightarrow \Phi(z) \neq 0$  pour  $z \in \mathbb{D}$ . On a alors

$$X_t = \sum_{k=0}^{\infty} \beta_k \varepsilon_{t-k}$$

où  $(\beta_k) \in \ell^1$  satisfait

$$\frac{1}{\Phi(z)} = \sum_{k=0}^{\infty} \beta_k z^k.$$

**Exemple 3.13.** Pour le modèle  $AR(1)$  on a  $\Phi(z) = 1 - \varphi z$  et on voit qu'une solution stationnaire existe ssi  $|\varphi| \neq 1$ , et que le processus est causal si et seulement si  $|\varphi| < 1$ , auquel cas on trouve que  $\beta_k = \varphi^k$  grâce à l'identité

$$\frac{1}{1 - \varphi z} = \sum_{k=0}^{\infty} \varphi^k z^k.$$

### 3.4.3 Autoregressive + Moving average

Les modèles  $ARMA(p, q)$  représentent la classe de processus stochastiques  $(X_t)$  définis comme solutions stationnaires du systèmes d'équations

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (\text{ARMA}(p, q))$$

où  $(\varepsilon_t) \sim \text{BB}(0, \sigma^2)$  et  $\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \sigma$  sont des paramètres réels libres, pourvu que de telles solutions existent. Avec les polynômes  $\Theta(z)$  et  $\Phi(z)$  définis comme au dessus, on a donc

$$\Phi(B)X_t = \Theta(B)\varepsilon_t.$$

On peut alors finalement montrer le résultat suivant :

**Théorème 3.14.** — (ARMA(p, q)) admet une unique solution stationnaire  $(X_t)$   
 $\Leftrightarrow \Phi(z) \neq 0$  pour  $z \in \mathbb{S}$ .  
 —  $(X_t)$  est causal  $\Leftrightarrow \Phi(z) \neq 0$  pour  $z \in \mathbb{D}$ . On peut alors montrer qu'il existe  $c > 0$  et  $r \in ]0, 1[$  tels que  $|\gamma(h)| \leq cr^{|h|}$  pour tout  $h \in \mathbb{Z}$  : la fonction d'autocorrélation est à décroissance exponentielle.  
 —  $(X_t)$  est inversible  $\Leftrightarrow \Theta(z) \neq 0$  pour  $z \in \mathbb{D}$ .

**Remarque 3.15.** Pour les modèles  $ARMA(p, q)$  tels que  $\Phi(z) \neq 0$  pour tout  $z \in \mathbb{D}$ , on a alors  $\mathbb{E}[X_t] = 0$ . On peut décentrer le processus en considérant, pour une constante  $C \in \mathbb{R}$ , l'équation

$$\Phi(B)X_t = \Theta(B)\varepsilon_t + C$$

à la place de (ARMA(p, q)). Les mêmes théorèmes s'appliquent, sauf que maintenant

$$\mathbb{E}[X_t] = \frac{C}{1 - \varphi_1 - \dots - \varphi_p} \quad (\text{Exercice}).$$

### 3.5 Interlude mathématique : Projection orthogonale

Si  $\mathcal{H}$  est un sous-espace vectoriel *fermé* de  $L^2$ , alors pour tout  $X \in L^2$  il existe une unique v.a. de  $\mathcal{H}$  appelée la *projection orthogonale de  $X$  sur  $\mathcal{H}$*  et notée  $P_{\mathcal{H}}(X)$  qui satisfait :

$$\|P_{\mathcal{H}}(X) - X\|_{L^2} = \min_{Y \in \mathcal{H}} \|Y - X\|_{L^2}.$$

On a aussi la caractérisation utile :  $P_{\mathcal{H}}(X)$  est l'unique élément de  $L^2$  tel que

- $P_{\mathcal{H}}(X) \in \mathcal{H}$
- $X - P_{\mathcal{H}}(X)$  est orthogonal à  $\mathcal{H}$ .

Ce résultat est toujours vrai si  $\mathcal{H}$  est seulement un sous-ensemble convexe et fermé de  $L^2$  (ou d'un espace de Hilbert quelconque).

Si  $\mathcal{H}$  est un sous-espace de dimension finie de  $L^2$  muni d'une base *orthonormée*  $e_1, \dots, e_m$ , alors on a :

$$P_{\mathcal{H}}(X) = \sum_{j=1}^d \langle X, e_j \rangle e_j.$$

Si cette base est seulement *orthogonale*, on l'orthonormalise en prenant  $\frac{e_j}{\|e_j\|}$  et on a alors

$$P_{\mathcal{H}}(X) = \sum_{j=1}^d \langle X, e_j \rangle \frac{e_j}{\|e_j\|^2}.$$

### 3.6 Decomposition de Wold

Réciproquement, il s'avère que presque tout processus stationnaire est presque un processus linéaire (sauf que les coefficients  $(\alpha_k)$  peuvent appartenir à  $\ell^2$  seulement et non  $\ell^1$ ), où "presque" veut dire à un processus déterministe près.

Dans la suite, étant donné un processus stochastique  $(X_t)$  on considèrera souvent le sous-espace  $\mathcal{H}_t$  de  $L^2 := L^2(\Omega, \mathbb{P})$  donné par la fermeture de  $\text{Vect}(X_t, X_{t-1}, \dots)$ . En d'autres termes,  $\mathcal{H}_t$  est le sous-espace vectoriel des variables aléatoires de la forme

$$\sum_{k=0}^{\infty} \alpha_k X_{t-k}$$

où  $(\alpha_k) \in \ell^2$ . On note alors

$$X_t^* := P_{\mathcal{H}_{t-1}}(X_t) = \arg \min_{Y \in \mathcal{H}_{t-1}} \|X_t - Y\|_{L^2}$$

qui est la prediction optimale (au sens  $L^2$ ) de  $X_t$  par le passé  $X_{t-1}, X_{t-2}, \dots$ . On appelle  $X_t - X_t^*$  l'*innovation* au temps  $t$ . On dit que  $X_t$  est *déterministe* si  $X_t = X_t^*$  pour tout  $t$ , c'est-à-dire d'innovation nulle.

Par exemple, si  $Z$  est une variable aléatoire  $L^2$ , alors  $X_t := Z$  et  $X_t := (-1)^t Z$  sont déterministes.

**Théorème 3.16.** *Si  $(X_t)$  est stationnaire non-déterministe alors il existe :*

- $(D_t)$  *déterministe*
- $(\varepsilon_t) \sim BB(0, \sigma^2)$  *non-corrélé à  $(D_t)$  défini sur le même espace probabilités.*
- *Une suite de réels  $(\alpha_k) \in \ell^2$  avec  $\alpha_0 = 1$*

tels que

$$X_t = D_t + \sum_{k=0}^{\infty} \alpha_k \varepsilon_{t-k}$$

dans  $L^2$ . Cette décomposition est unique. Si de plus  $(\alpha_k) \in \ell^1$  alors cette décomposition a lieu dans  $L^1$  et  $\mathbb{E}[X_t] = \mathbb{E}[D_t]$ .



## 4 Prédiction linéaire

On suppose ici que  $(X_t)_{t \in \mathbb{Z}}$  est un processus stationnaire  $L^2$  qui est centré et non-dégénéré : Pour tout  $t \in \mathbb{Z}$ ,

$$\mathbb{E}[X_t] = 0, \quad \gamma(0) = \mathbb{E}[X_t^2] > 0.$$

### 4.1 Prédicteur linéaire optimal

Le problème posé est le suivant : étant donné des observations  $x_{t-1}, \dots, x_{t-m}$  pour un entier  $m \geq 1$ , comment approcher au mieux  $x_t$  ? On pose alors

$$\mathcal{H}_{t-1,m} := \text{Vect}(X_{t-1}, \dots, X_{t-m}),$$

qui est un sous-espace de dimension fini des variables aléatoires  $L^2$ . Ainsi  $X \in \mathcal{H}_{t-1,m} \Leftrightarrow$  il existe des coefficients  $c_1, \dots, c_m \in \mathbb{R}$  tels que  $X = c_1 X_{t-1} + \dots + c_m X_{t-m}$ . Notez que

$$\begin{aligned} \dim \mathcal{H}_{t-1,m} = m &\Leftrightarrow \{ \text{si } c_1 X_{t-1} + \dots + c_m X_{t-m} = 0 \text{ alors } c_1 = \dots = c_m = 0 \} \\ &\Leftrightarrow \text{la matrice } \left[ \langle X_{t-i}, X_{t-j} \rangle \right]_{i,j=1}^m \text{ est inversible} \quad (\text{Exercice}) \\ &\Leftrightarrow \text{la matrice } \Gamma_m := \left[ \gamma(i-j) \right]_{i,j=1}^m \text{ est inversible.} \end{aligned}$$

**On supposera que cette condition est satisfaite pour tout  $m \geq 1$ .** On peut montrer que c'est le cas dès que  $\gamma(h) \rightarrow 0$  quand  $h \rightarrow \infty$ . C'est donc le cas pour le modèle ARMA(p,q) causal.

On définit alors le *prédicteur linéaire optimal*  $X_t^{*,m}$  comme la projection orthogonale de  $X_t$  sur  $\mathcal{H}_{t-1,m}$ . On note aussi  $E_{t,m} := X_t - X_t^{*,m}$  l'*erreur de prédiction* et  $\sigma_m^2 := \|E_{t,m}\|^2$  sa variance, qu'on appelle l'*erreur de prédiction quadratique moyenne*.

Ainsi,  $X_t^{*,m}$  est l'*unique* variable  $L^2$  caractérisée par :

$$X_t^{*,m} \in \mathcal{H}_{t-1,m} \quad \text{et} \quad X_t - X_t^{*,m} \text{ est orthogonal à } \mathcal{H}_{t-1,m} \quad (4.1)$$

Une caractérisation alternative est la suivante :

- Il existe  $\varphi_{1,m}, \dots, \varphi_{m,m} \in \mathbb{R}$  tels que  $X_t^{*,m} = \varphi_{1,m} X_{t-1} + \dots + \varphi_{m,m} X_{t-m}$
- $\varphi_{1,m}, \dots, \varphi_{m,m}$  minimisent la fonction (on pose  $c_0 := -1$ )

$$\begin{aligned} (c_1, \dots, c_m) &\mapsto \|X_t - (c_1 X_{t-1} + \dots + c_m X_{t-m})\|^2 \\ &= \left\| \sum_{k=0}^m c_k X_{t-k} \right\|^2 \\ &= \sum_{k,\ell=0}^m c_k c_\ell \gamma(k-\ell). \end{aligned}$$

**Exemple 4.1** (AR(p) causal). Soit  $(X_t)$  une solution stationnaire de

$$X_t = \varphi_1 X_{t-1} + \dots + X_{t-p} + \varepsilon_t, \quad (\varepsilon_t) \sim BB(0, \sigma^2)$$

que l'on suppose causal ; on a vu dans la section précédente que cela est vrai si et seulement si  $1 - \varphi_1 z - \dots - \varphi_p z^p \neq 0$  pour tout  $z \in \mathbb{D}$ . Ainsi, il existe une suite  $(\alpha_k) \in \ell^1$  telle que

$$X_t = \sum_{k=0}^{\infty} \alpha_k \varepsilon_{t-k}$$

et donc  $\mathbb{E}[\varepsilon_t X_{t-h}] = 0$  pour tout  $h \geq 1$ . Du coup, on a montré que pour tout  $m \geq 1$ ,

$$\varepsilon_t = X_t - \underbrace{(\varphi_1 X_{t-1} + \dots + X_{t-p})}_{\in \mathcal{H}_{t-1,p}} \text{ est orthogonal à } \mathcal{H}_{t-1,m}$$

et donc, par unicité du projeté orthogonal  $X_t^{*,m}$  et sa caractérisation (4.1), on obtient  $X_t^{*,m} = \varphi_1 X_{t-1} + \dots + X_{t-p}$  pour tout  $m \geq p$ . En particulier, pour tout  $m \geq p$ ,

$$\varphi_{k,m} = \begin{cases} \varphi_k & \text{si } 1 \leq k \leq p \\ 0 & \text{si } p < k \leq m \end{cases}.$$

## 4.2 Equation de Yule-Walker

Comment calculer  $\varphi_{k,m}$  de façon générale ? Quand  $m = 1$ , en utilisant que  $X_{t-1}/\|X_{t-1}\|$  est une base orthonormée de l'espace  $\mathcal{H}_{t-1,1}$  de dimension un on obtient :

$$X_t^{*,1} = P_{\mathcal{H}_{t-1,1}}(X_t) = \frac{\langle X_t, X_{t-1} \rangle}{\|X_{t-1}\|^2} X_{t-1} = \frac{\gamma(1)}{\gamma(0)} X_{t-1}$$

et donc on a toujours  $\varphi_{1,1} = \gamma(1)/\gamma(0) = \rho(1)$ . De plus,

$$\sigma_1^2 = \|X_t - X_t^{*,1}\|^2 = \gamma(0) - \frac{\gamma(1)^2}{\gamma(0)}.$$

Plus généralement, on peut exprimer  $\varphi_{k,m}$  et  $\sigma_m$  en fonction de  $\gamma(h)$  ; rappelons que la matrice (de covariance de  $X_{t-1}, \dots, X_{t-m}$ )

$$\Gamma_m := \left[ \gamma(i-j) \right]_{i,j=1}^m$$

est supposée inversible dans toute cette section.

**Théorème 4.2** (Equations de Yule-Walker). *Pour tout  $m \geq 1$ , on a la relation :*

$$\begin{bmatrix} \varphi_{1,m} \\ \vdots \\ \varphi_{m,m} \end{bmatrix} = \Gamma_m^{-1} \begin{bmatrix} \gamma(1) \\ \vdots \\ \gamma(m) \end{bmatrix}.$$

De plus, on a :

$$\sigma_m^2 = \gamma(0) - \begin{bmatrix} \varphi_{1,m} \\ \vdots \\ \varphi_{m,m} \end{bmatrix}^t \begin{bmatrix} \gamma(1) \\ \vdots \\ \gamma(m) \end{bmatrix} = \gamma(0) - \begin{bmatrix} \gamma(1) \\ \vdots \\ \gamma(m) \end{bmatrix}^t \Gamma_m^{-1} \begin{bmatrix} \gamma(1) \\ \vdots \\ \gamma(m) \end{bmatrix}.$$

Ainsi, pourvu que l'on réussisse à inverser la matrice  $\Gamma_p$ , on peut exprimer les coefficients  $\varphi_{k,m}$  des prédicteurs en termes de  $\gamma(h)$ . D'un autre côté, nous avons vu qu'étant donné une série temporelle  $x_1, \dots, x_n$  on a pour tout  $h$  un estimateur raisonnable  $\hat{\gamma}_n(h)$  de  $\gamma(h)$ . En remplaçant  $\gamma(h)$  par  $\hat{\gamma}_n(h)$  dans les équations de Yule-Walker on obtient alors des estimateurs  $\hat{\varphi}_{k,m}$  de  $\varphi_{k,m}$ , et donc un estimateur

$$\hat{X}_t^{*,m} := \hat{\varphi}_{1,m}x_{t-1} + \dots + \hat{\varphi}_{m,m}x_{t-m}$$

du prédicteur  $X_t^{*,m}$  ! De même, on obtient un estimateur  $\hat{\sigma}_m^2$  de  $\sigma_m^2$ . Ces estimateurs sont appelés les *estimateurs de Yule-Walker*.

*Démonstration.* Comme  $X_t - X_t^{*,m}$  est orthogonal à  $\mathcal{H}_{t-1,m} = \text{Vect}(X_{t-1}, \dots, x_{t-m})$ , on obtient pour tout  $1 \leq j \leq m$  que  $\langle X_t - X_t^{*,m}, X_{t-j} \rangle = 0$ . Ceci s'écrit également, pour tout  $1 \leq j \leq m$  :

$$\begin{aligned} \langle X_t, X_{t-j} \rangle &= \langle X_t^{*,m}, X_{t-j} \rangle \\ \Leftrightarrow \gamma(j) &= \sum_{k=1}^m \varphi_{k,m} \underbrace{\langle X_{t-k}, X_{t-j} \rangle}_{\gamma(j-k)} \end{aligned}$$

ce qui donne matriciellement

$$\begin{bmatrix} \gamma(1) \\ \vdots \\ \gamma(m) \end{bmatrix} = \Gamma_m \begin{bmatrix} \varphi_{1,m} \\ \vdots \\ \varphi_{m,m} \end{bmatrix} \quad (4.2)$$

et l'identité recherchée est obtenu en multipliant par  $\Gamma_m^{-1}$  de chaque côté.

Ensuite, en utilisant que  $X_t - X_t^{*,m}$  est orthogonal à  $X_t^{*,m} \in \mathcal{H}_{t-1,m}$ , on obtient

$$\begin{aligned} \sigma_m^2 &= \|X_t - X_t^{*,m}\|^2 \\ &= \langle X_t - X_t^{*,m}, X_t \rangle \\ &= \gamma(0) - \langle X_t^{*,m}, X_t \rangle \\ &= \gamma(0) - \sum_{k=1}^m \varphi_{k,m} \langle X_{t-k}, X_t \rangle \\ &= \gamma(0) - \begin{bmatrix} \varphi_{1,m} \\ \vdots \\ \varphi_{m,m} \end{bmatrix}^t \begin{bmatrix} \gamma(1) \\ \vdots \\ \gamma(m) \end{bmatrix} \end{aligned}$$

et la seconde identité est obtenue grâce à (4.2) et  ${}^t(\Gamma_m^{-1}) = ({}^t\Gamma_m)^{-1}$ .  $\square$

**Exemple 4.3** (AR(1)). On avait calculé pour le modèle AR(1),  $X_t = \varphi X_{t-1} + \varepsilon_t$  avec  $(\varepsilon_t) \sim BB(0, \sigma^2)$ , que

$$\gamma(h) = \frac{\sigma^2 \varphi^{|h|}}{1 - \varphi^2}.$$

Si le modèle est supposé causal, on obtient ainsi pour tout  $m \geq 1$ ,

$$\sigma_m^2 = \gamma(0) - \begin{bmatrix} \varphi \\ 0 \\ \vdots \\ 0 \end{bmatrix}^t \begin{bmatrix} \gamma(1) \\ \vdots \\ \gamma(m) \end{bmatrix} = \sigma^2.$$

### 4.3 Fonctions d'autocorrélation partielle

Rappelons que l'ordre  $q$  d'un modèle MA( $q$ ) peut-être détecté grâce à la fonction d'autocorrélation car  $\rho(h) = 0$  pour tout  $h > q$  alors que  $\rho(q) \neq 0$  (Exercice).

La *fonction d'autocorrélation partielle*  $\kappa : \mathbb{N}^* \rightarrow \mathbb{R}$  (PACF en anglais et sous R) est définie par

$$\kappa(h) := \varphi_{h,h}$$

et joue un rôle similaire dans la détection du paramètre  $p$  des modèles AR( $p$ ). En effet, comme on l'a vu précédemment,  $\kappa(h) = 0$  pour tout  $h > p$  alors que  $\kappa(p) = \varphi_p \neq 0$ .

La *fonction d'autocorrélation partielle empirique*  $\hat{\kappa}_n(h) := \hat{\varphi}_{h,h}$  est disponible sous R via la commande `pacf`. On peut montrer que, si  $(X_t)$  est solution stationnaire d'un modèle AR causal telle que  $(\varepsilon_t)$  soit un BB(0,  $\sigma$ ) *fort*, alors

$$\sqrt{n} \kappa_n(h) \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, 1)$$

pour tout  $h > p$ , ce qui permet de délimiter un intervalle de confiance à 5% comme on l'avait fait pour l'acf.

### 4.4 Prédicteurs à horizon $h$ (Exercice)

On pourrait chercher à obtenir un prédicteur pour  $X_{t+h}$  comme fonction de  $X_{t-1}, \dots, X_{t-m}$  pour un  $h \geq 0$ . La même méthode qu'au dessus est applicable : On définit  $X_{t+h}^{*,m}$  comme la projection orthogonal de  $X_{t+h}$  sur  $\mathcal{H}_{t-1,m}$  et on cherche à déterminer les coefficients  $\varphi_{k,m}^{(h)}$  où  $1 \leq h \leq m$  tels que

$$X_{t+h}^{*,m} = \varphi_{1,m}^{(h)} X_{t-1} + \dots + \varphi_{m,m}^{(h)} X_{t-m}.$$

En reprenant la preuve des équations de Yule-Walker, on obtient alors pour tout  $m \geq 1$  :

$$\begin{bmatrix} \varphi_{1,m}^{(h)} \\ \vdots \\ \varphi_{m,m}^{(h)} \end{bmatrix} = \Gamma_m^{-1} \begin{bmatrix} \gamma(1+h) \\ \vdots \\ \gamma(m+h) \end{bmatrix}$$

et

$$(\sigma_m^{(h)})^2 := \|X_{t+h} - X_{t+h}^{*,m}\|^2 = \gamma(0) - \begin{bmatrix} \varphi_{1,m}^{(h)} \\ \vdots \\ \varphi_{m,m}^{(h)} \end{bmatrix} \begin{bmatrix} \gamma(1+h) \\ \vdots \\ \gamma(m+h) \end{bmatrix}.$$

### 4.5 Calcul effectif de l'inverse $\Gamma_m^{-1}$

**Décomposition de Cholesky.**  $\Gamma_m$  est clairement symétrique. Remarquons que c'est une matrice semi-définie positive : pour tout  $v \in \mathbb{R}^m$ , si on note  $U$  le vecteur d'entrées  $U_i := X_i - \mathbb{E}[X_i]$ , alors  $\Gamma_m = \mathbb{E}[U^t U]$  et donc

$${}^t v \Gamma_m v = \mathbb{E} \|{}^t U v\|^2 \geq 0.$$

Comme  $\Gamma_m$  est supposée inversible, elle est donc symétrique définie-positive.

La factorization de Cholesky, qui s'applique à n'importe quelle matrice symétrique définie-positive, assure qu'il existe une matrice  $m \times m$  diagonale  $D$  de coefficients diagonaux strictement positifs et une matrice  $m \times m$   $L$  triangulaire inférieure avec des uns sur la diagonale telles que

$$\Gamma_m = LD^tL.$$

Attention, en général pas de lien avec la décomposition spectrale. Maintenant, comme  $\Gamma_m^{-1} = ({}^tL)^{-1}D^{-1}L^{-1} = {}^t(L)^{-1}D^{-1}L^{-1}$ , il ne reste plus qu'à obtenir  $L^{-1}$  par récurrence simple (en la cherchant sous la forme d'une matrice triangulaire inférieure avec des uns sur la diagonale  $\spadesuit$ ) et  $D^{-1}$  qui est égal à  $\text{diag}(D_{11}^{-1}, \dots, D_{mm}^{-1})$ . En écrivant le produit précédent en coordonnées, on trouve la relation de récurrence :

$$\begin{aligned} D_{11} &:= (\Gamma_m)_{11} \\ D_{jj} &:= (\Gamma_m)_{jj} - \sum_{r=1}^{j-1} L_{jr}^2 D_{rr}, \quad 2 \leq j \leq p, \\ L_{ij} &:= \frac{1}{D_{jj}} \left( (\Gamma_m)_{ij} - \sum_{r=1}^{j-1} L_{ir} D_{rr} L_{jr} \right), \quad 1 \leq i < j \leq p. \end{aligned}$$

La complexité d'un tel algorithme est  $\mathcal{O}(m^3)$ . On peut montrer que comme  $\Gamma_m$  est une matrice de Toeplitz, c'est-à-dire constante sur les diagonales, on peut en fait s'en sortir avec  $\mathcal{O}(m^2)$  opérations.

**Algorithme de Durbin-Levinson.** L'algorithme plus communément utilisé utilise fortement la structure de Toeplitz de la matrice de covariance  $\Gamma_m$ , c'est-à-dire le fait que  $(X_t)$  est stationnaire. Il a un coût computationnel de  $\mathcal{O}(m^2)$ .

**Théorème 4.4** (Algorithme de Durbin-Levinson). *Les coefficients  $(\varphi_{k,m})$  et  $(\sigma_m^2)$  sont calculés en partant de*

$$\varphi_{1,1} = \frac{\gamma(1)}{\gamma(0)}, \quad \sigma_0^2 := \gamma(0)$$

et puis en utilisant les équations de récurrence, pour  $m \geq 1$  :

$$\varphi_{m+1,m+1} = \frac{1}{\sigma_m^2} \left( \gamma(m+1) - \sum_{k=1}^m \varphi_{k,m} \gamma(m-k+1) \right)$$

et

$$\begin{bmatrix} \varphi_{1,m+1} \\ \vdots \\ \varphi_{m,m+1} \end{bmatrix} = \begin{bmatrix} \varphi_{1,m} \\ \vdots \\ \varphi_{m,m} \end{bmatrix} - \varphi_{m+1,m+1} \begin{bmatrix} \varphi_{m,m} \\ \vdots \\ \varphi_{1,m} \end{bmatrix}$$

ainsi que

$$\sigma_{m+1}^2 = \sigma_m^2 (1 - \varphi_{m+1,m+1}^2).$$

---

$\spadesuit$ . Si  $m$  n'est pas trop grand, on aussi utiliser l'observation suivante : si on écrit  $L = I + \Delta$  alors  $\Delta^m = 0$  et donc  $L^{-1} = (I + \Delta)^{-1} = \sum_{k=0}^{\infty} \Delta^k = \sum_{k=0}^{m-1} \Delta^k$ .

*Démonstration.* On écrit

$$\mathcal{H}_{t-1,m+1} = \text{Vect}(X_{t-1}, \dots, X_{t-m-1}) = \underbrace{\text{Vect}(X_{t-1}, \dots, X_{t-m})}_{\mathcal{H}_{t-1,m}} \oplus E$$

où le sous-espace  $E$  est orthogonal à  $\mathcal{H}_{t-1,m}$ . Nécessairement  $\dim E = 1$  et on peut prendre  $E = \text{Vect}(Y)$  avec

$$Y := X_{t-m-1} - P_{\mathcal{H}_{t-1,m}}(X_{t-m-1}).$$

Cette décomposition en sous-espace orthogonaux entraîne l'identité clef  $\heartsuit$  :

$$\begin{aligned} X_t^{*,m+1} &= P_{\mathcal{H}_{t-1,m+1}}(X_t) \\ &= P_{\mathcal{H}_{t-1,m}}(X_t) + P_E(X_t) \\ &= X_t^{*,m} + \frac{\langle X_t, Y \rangle}{\|Y\|^2} Y \\ &= X_t^{*,m} + \kappa_{m+1}(X_{t-m-1} - P_{\mathcal{H}_{t-1,m}}(X_{t-m-1})), \quad \kappa_{m+1} := \frac{\langle X_t, Y \rangle}{\|Y\|^2}. \end{aligned} \quad (4.3)$$

*Equation récurrente pour  $\varphi_{k,m}$*  : la même preuve que celle des équations de Yule-Walker montre que

$$P_{\mathcal{H}_{t-1,m}}(X_{t-m-1}) = \sum_{k=1}^m \alpha_k X_{t-k}$$

avec la relation

$$\Gamma_m \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{bmatrix} = \begin{bmatrix} \gamma(m) \\ \vdots \\ \gamma(1) \end{bmatrix}.$$

Mais comme on a la symétrie  $(\Gamma_m)_{ij} = (\Gamma_m)_{m-i,m-j}$ , on obtient avec les équations de Yule-Walker

$$\Gamma_m \begin{bmatrix} \alpha_m \\ \vdots \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} \gamma(1) \\ \vdots \\ \gamma(m) \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} \alpha_m \\ \vdots \\ \alpha_1 \end{bmatrix} = \Gamma_m^{-1} \begin{bmatrix} \gamma(1) \\ \vdots \\ \gamma(m) \end{bmatrix} = \begin{bmatrix} \varphi_{1,m} \\ \vdots \\ \varphi_{m,m} \end{bmatrix},$$

c'est-à-dire  $\alpha_k = \varphi_{m-k+1,m}$  et finalement,

$$P_{\mathcal{H}_{t-1,m}}(X_{t-m-1}) = \sum_{k=1}^m \varphi_{m-k+1,m} X_{t-k}.$$

En réinjectant cette expression dans (4.3) et en identifiant les coefficients (là on utilise que la projection orthogonale est unique), on obtient les équations de récurrence :

$$\varphi_{k,m+1} = \begin{cases} \varphi_{k,m} - \kappa_{m+1} \varphi_{m-k+1,m} & \text{pour } 1 \leq k \leq m \\ \kappa_{m+1} & \text{si } k = m+1. \end{cases}$$

---

$\heartsuit$ . (Exercice) Montrer que si  $E$  et  $F$  sont deux sous-espace orthogonaux de  $L^2$  et qu'on note  $E \oplus F := \{X + Y : X \in E, Y \in F\}$  alors  $P_{E \oplus F} = P_E + P_F$ .

Formule pour  $\kappa_m = \varphi_{m+1,m+1}$  : On rappelle qu'on a trouvé,

$$Y = X_{t-m+1} - \sum_{k=1}^m \varphi_{m-k+1,m} X_{t-k} = X_{t-m+1} - \sum_{k=1}^m \varphi_{k,m} X_{t-m+k-1}.$$

On calcule

$$\langle X_t, Y \rangle = \gamma(m+1) - \sum_{k=1}^m \varphi_{k,m} \gamma(m-k+1)$$

et, en utilisant que  $Y$  est orthogonal à  $\mathcal{H}_{t-1,m}$  et les équations de Yule-Walker,

$$\|Y\|^2 = \langle Y, X_{t-m+1} \rangle = \gamma(0) - \sum_{k=1}^m \varphi_{k,m} \gamma(k) = \sigma_m^2.$$

Equation récurrente pour  $\sigma_m^2$  :

$$\begin{aligned} \sigma_{m+1}^2 &= \|X_t - X_t^{*,m+1}\|^2 \\ &= \|(X_t - X_t^{*,m}) + \kappa_{m+1} Y\|^2 \\ &= \sigma_m^2 - 2\kappa_{m+1} \langle X_t, Y \rangle + \kappa_{m+1}^2 \|Y\|^2 \\ &= \sigma_m^2 - \kappa_{m+1}^2 \|Y\|^2 \\ &= \sigma_m^2 (1 - \kappa_{m+1}^2). \end{aligned}$$

□

**Résumé :** Comme on peut calculer les fonctions d'autocorrélation empiriques  $\hat{\rho}_n(h)$  pour tout  $h$  depuis les données  $x_1, \dots, x_n$ , la formule de l'algorithme de Durbin-Levinson permet alors de calculer itérativement les  $\hat{\varphi}_{k,m}$  et  $\hat{\sigma}_m^2$ . Si l'on s'est convaincu que la série correspond bien à un modèle  $AR(p)$  avec  $p$  donné, en prenant  $m = p$  on obtient

- des estimateurs des coefficients  $\varphi_1, \dots, \varphi_p$  donnés par  $\hat{\varphi}_{1,p}, \dots, \hat{\varphi}_{p,p}$ .
- une prédiction pour  $x_{n+1}$  (et même  $x_{n+h}$ ) donnée par

$$\hat{x}_{n+1}^* := \hat{\varphi}_{1,p} x_n + \dots + \hat{\varphi}_{p,p} x_{n-p}$$

- une valeurs approchée du risque quadratique moyen de la prédiction :  $\hat{\sigma}_p^2$ .

Il convient d'étudier les résidus de la modélisation avec le modèle  $AR(p)$  et de voir s'ils passent le test de blancheur.

Si on s'intéresse plutôt à un modèle  $MA(q)$  voir  $ARMA(p,q)$ , il existe des formules plus compliquées mais explicites entre les coefficients  $\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_m$  et les  $\varphi_{k,m}$ , qu'on pourra trouver dans [BrDa, Chapitre 8.3 et 8.4]. On peut donc utiliser la même approche pour tester le modèle  $ARMA(p,q)$  et obtenir une prédiction  $\hat{x}_{n+1}^*$  de  $x_{n+1}$  (ou  $x_{n+h}$ ).

Si l'on ne connaît pas a priori les paramètres  $(p, q)$  du modèle  $ARMA$ , il existe des approches pour les choisir mais elles demandent de plus fortes hypothèses de gaussianité. L'idée est que si  $(\varepsilon_t)$  est un bruit blanc gaussien alors on peut calculer explicitement la vraisemblance du modèle. Comme on peut le faire avec n'importe quel modèle paramétré, on pénalise la log-vraisemblance à l'aide d'un critère (AIC, AICc, BIC) de manière à maximiser une fonction du type

paramètres  $\mapsto$  log-Vraisemblance(paramètres) + Pénalisation(nombre de paramètres)

où la pénalisation dépend du critère choisi, mais est toujours une fonction décroissante en le nombre de paramètres. On évalue cette fonction sur un ensemble fini de paramètres et on choisit le modèle ARMA(p,q) qui rend la valeur maximale. L'idée est qu'en augmentant le nombre de paramètres on fera mécaniquement croître la log-vraisemblance seule, mais que la pénalisation rajoutée privilégiera les modèles parcimonieux, c'est-à-dire avec le moins de paramètres possibles.



## 5 Prédiction par lissage

Que faire si aucun des modèles paramétriques ci-dessus ne semble adapté pour modéliser le jeu de données? On peut toujours faire du lissage (très utilisé par l'industrie), même si aucune garantie théorique n'est apportée (en général).

Etant observé  $x_1, \dots, x_n$ , on note  $\hat{x}_n(h)$  la prédiction à horizon  $h$ , c'est-à-dire notre candidat pour l'inconnu  $x_{n+h}$  connaissant  $x_1, \dots, x_n$ .

### 5.1 Lissage exponentiel simple

**Modèle : “ $x_t = c + \varepsilon_t$ ”.** On se donne un paramètre  $0 < \beta < 1$ . Le principe est de faire une régression de  $(x_1, \dots, x_n)$  par une constante mais en donnant un poids exponentiel  $\beta^{n-j}$  à  $x_j$  pour pénaliser les observations plus éloignées dans le passé. En d'autres termes on considère la solution du problème des moindres carrés :

$$\text{minimiser } c \mapsto \sum_{j=0}^{n-1} \beta^j (x_{n-j} - c)^2.$$

La solution est donnée par (Exercice)

$$\frac{1 - \beta}{1 - \beta^n} \sum_{j=0}^{n-1} \beta^j x_{n-j}. \quad (5.1)$$

En utilisant l'approximation  $\beta^n \simeq 0$  dans le dénominateur, qui est raisonnable pour  $n$  assez grand, on pose alors pour tout  $h \geq 1$ ,

$$\hat{x}_n(h) := (1 - \beta) \sum_{j=0}^{n-1} \beta^j x_{n-j}.$$

Notez que ce prédicteur ne dépend pas de  $h$ .

**Formules de mise à jour :** De  $\hat{x}_n(h) = (1 - \beta)(x_n + \beta x_{n-1} + \beta^2 x_{n-2} + \dots + \beta^{n-1} x_1)$ , on voit qu'on peut calculer  $\hat{x}_n(h)$  simplement en fonction du prédicteur  $\hat{x}_{n-1}(h)$  basé sur  $x_1, \dots, x_{n-1}$  et de la nouvelle donnée  $x_n$  :

$$\hat{x}_n(h) = \beta \hat{x}_{n-1}(h) + (1 - \beta)x_n.$$

**Interprétation du paramètre  $\beta$  :** Quand  $\beta \rightarrow 0$ , on a  $\hat{x}_n(h) \simeq x_n$  et le prédicteur ne prend en compte que la dernière valeur. Quand  $\beta \rightarrow 1$ , on a

$$\hat{x}_n(h) \simeq (1 - \beta)(x_1 + \dots + x_n)$$

et le prédicteur donne un poids égal à toutes les données. Bref, le paramètre  $\beta$  encode l'importance que l'on donne au passé :

- $\beta \sim 0$  : court terme
- $\beta \sim 1/2$  : moyen terme
- $\beta \sim 1$  : long terme

**Comment choisir  $\beta$  ?** Typiquement, on choisira  $\beta \in ]0, 1/3]$  pour une prédiction tenant compte du passé proche,  $\beta \in [2/3, 1[$  pour tenir compte de tout le passé, et  $\beta \in ]1/3, 2/3]$  pour une version intermédiaire.

Sinon, on peut aussi prendre le  $\beta$  qui minimise, par exemple, les erreurs de prédiction à horizon un des données que l'on connaît déjà, c'est-à-dire le  $\hat{\beta}$  qui minimise

$$\beta \mapsto \sum_{k=1}^{n-1} (x_{k+1} - \hat{x}_k[1])^2.$$

**Remarque :** Il n'existe à priori pas de relation simple entre l'erreur de prévision quadratique moyenne  $\mathbb{E}[(X_{t+h} - \hat{X}_n(h))^2]$  et le paramètre  $\beta$ , comme on peut le vérifier sur un modèle AR(1) où tous les calculs sont explicites.

(Exercice) Plus précisément, en faisant l'approximation pour  $|a| < 1$ ,

$$\sum_{j=0}^{n-1} a^j \simeq \frac{1}{1-a},$$

montrer qu'on a, si  $X_t = \varphi X_{t-1} + \varepsilon_t$  avec  $|\varphi| < 1$  et  $(\varepsilon_t) \sim \text{BB}(0, 1)$ ,

$$\mathbb{E}[(X_{n+h} - \hat{X}_n[h])^2] \simeq E(h, \beta, \varphi) := \frac{1}{1-\varphi^2} \left( 1 - \frac{2(1-\beta)\varphi^h}{1-\beta\varphi} + \frac{(1-\beta)(1+\beta\varphi)}{(1+\beta)(1-\beta\varphi)} \right).$$

*Aide :* En séparant la somme en  $k < \ell$ ,  $k = \ell$ , et  $k > \ell$ , on montrera que

$$\sum_{k, \ell=0}^{\infty} \varphi^{|k-\ell|} \beta^{k+\ell} = \frac{1+\beta\varphi}{(1-\beta^2)(1-\beta\varphi)}.$$

Si on étudie (sur  $\mathbb{R}$ ), par exemple en prenant  $h = 1$ , la fonction  $\beta \mapsto E(1, \beta, \varphi)$ , alors on voit que cette fonction est décroissante sur  $]0, 1[$  si  $\varphi = 1/3$ , mais qu'elle n'est plus monotone si  $\varphi = 2/3$  (elle est décroissante puis croissante), comme illustré en Figure 5.1.

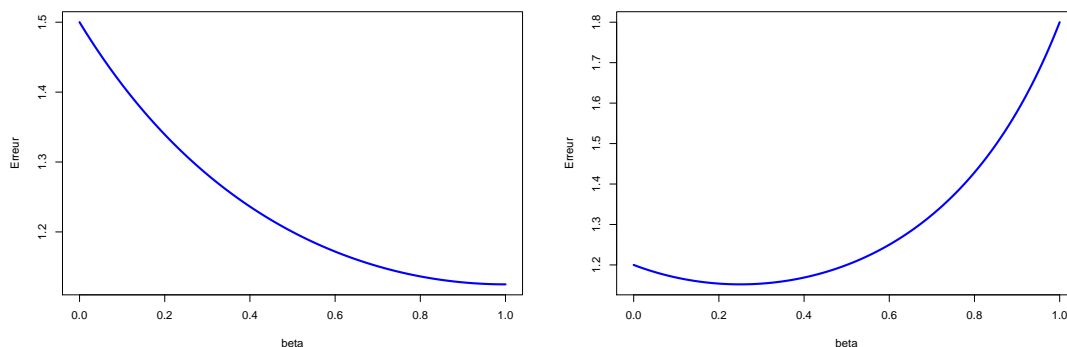


FIGURE 1 –  $E(1, \beta, \varphi)$  en fonction de  $\beta$  pour  $\varphi = 1/3$  (gauche) et  $\varphi = 2/3$  (droite).

## 5.2 Lissage exponentiel double

**Modèle :** “ $x_t = a + (t-n)b + \varepsilon_t$ .” Si le lissage simple ajuste une droite de pente constante au temps  $t = n$ , le lissage double ajuste une droite de pente quelconque. Plus précisément, on se donne un paramètre  $0 < \beta < 1$  et on cherche à résoudre le problème des moindres carrés :

$$\text{minimiser } (a, b) \mapsto \sum_{j=0}^{n-1} \beta^j (x_{n-j} - (aj + b))^2.$$

Après calcul et des approximations similaires  $\heartsuit$  à (5.1), on trouve alors

$$\hat{x}_n[h] = A_n h + B_n$$

avec

$$A_n := \frac{1 - \beta}{\beta} (S_1(n) - S_2(n)), \quad B_n := 2S_1(n) - S_2(n)$$

où

$$S_1(n) := (1 - \beta) \sum_{j=0}^{n-1} \beta^j x_{n-j}, \quad S_2(n) := (1 - \beta) \sum_{j=0}^{n-1} \beta^j S_1(n - j).$$

On a alors les formules de mise à jour :

$$\begin{aligned} A_n &= A_{n-1} + (1 - \beta)^2 (x_n - \hat{x}_{n-1}[1]) \\ B_n &= B_{n-1} + A_{n-1} + (1 - \beta^2) (x_n - \hat{x}_{n-1}[1]). \end{aligned}$$

En utilisant que  $\hat{x}_n[1] = A_n + B_n$  et en travaillant un peu, on peut écrire ces formules comme :

$$\begin{aligned} A_n &= \frac{1 + \beta}{2} A_{n-1} + \frac{1 - \beta}{2} (\hat{x}_n[1] - \hat{x}_{n-1}[1]) \\ B_n &= \beta^2 \hat{x}_{n-1}[1] + (1 - \beta^2) x_n \end{aligned}$$

que l'on interprète comme des combinaisons convexes.

**Une généralisation : Lissage de Holt-Winters.** Le lissage de Holt-Winters est basé sur la même idée que le lissage double sauf que l'on généralise les dernières équations en s'autorisant des coefficients différents : On se donne deux paramètres  $0 < \alpha, \gamma < 1$  et on prend encore

$$\hat{x}_n[h] = A_n h + B_n$$

avec cette fois la formule généralisée :

$$\begin{aligned} A_n &= \gamma A_{n-1} + (1 - \gamma) (\hat{x}_n[1] - \hat{x}_{n-1}[1]) \\ B_n &= \alpha \hat{x}_{n-1}[1] + (1 - \alpha) x_n. \end{aligned}$$

Le lissage exponentiel double est donc un cas particulier de Holt-Winters avec :

$$\alpha = \beta^2, \quad \gamma = \frac{1 + \beta}{2}$$

**NB :** On peut aussi intégrer une composante périodique dans le lissage, basée sur le modèle “ $x_t = a(t - n) + b + s_t + \varepsilon_t$ ” avec  $s_t$  périodique.

---

$\heartsuit$ . Plus précisément, on fait les approximations :  $\sum_{j=0}^{n-1} \beta^j \simeq (1 - \beta)^{-1}$ ,  $\sum_{j=0}^{n-1} j\beta^j \simeq \beta(1 - \beta)^{-2}$ , et  $\sum_{j=0}^{n-1} j^2\beta^j \simeq \beta(1 + \beta)(1 - \beta)^{-3}$ . (Exercice) : Montrer que ces approximations sont exactes quand  $n \rightarrow \infty$ .

## 6 Modèles de volatilité

### 6.1 Introduction

En mathématiques financières, la *volatilité* d'un actif<sup>?</sup> est une mesure des fluctuations de sa valeur au cours du temps, bien que plusieurs définitions coexistent. Certains indices de volatilité sont devenus des instruments financiers, comme l'indice VIX<sup>♠</sup>. En général, la volatilité n'est pas observable directement et se définit au travers d'une modélisation mathématique. Ici, ce sera *l'écart-type conditionnel des log-rendements*.

**Faits stylisés :** Quand on considère une série temporelle  $(X_t)$  qui représente les log-rendements d'un actif financier, on observe souvent les comportements suivants.

1. Typiquement, les réalisations semblent peu ou pas corrélées (ACF  $\simeq$  bruit blanc) sans pour autant être indépendantes (bruit blanc faible mais pas fort). *Cependant*, les amplitudes  $|X_t|$  s'avèrent fortement corrélées, et de façon équivalente  $X_t^2$ . Aussi, les fortes variations sont souvent suivies de fortes variations : on dit qu'il y a des *clusters de volatilité*.
2. Il y a généralement une asymétrie entre l'effet des valeurs passées négatives et l'effet des valeurs passées positives sur la volatilité : les baisses tendent à engendrer une augmentation de la volatilité supérieure à celle engendrée par des hausses de même ampleur (panique sur les marchés) ; c'est l'*effet de levier*.
3. Aussi,  $X_t$  a souvent une distribution *leptokurtique*. Plus précisément, on définit le *kurtosis* d'une variable  $X$  par

$$\kappa(X) := \frac{\mathbb{E}[(X - \mathbb{E}[X])^4]}{\text{Var}(X)^2}.$$

C'est une mesure de l'étalement de la distribution de  $X$ . On prend pour référence le cas où  $X$  est une variable gaussienne  $\mathcal{N}(\mu, \sigma^2)$ , auquel cas on a  $\kappa(X) = 3$  (Exercice). On dit qu'une variable aléatoire  $X$  est *leptokurtique* si  $\kappa(X_t) > 3$ . Cela veut dire que sa queue de distribution  $\mathbb{P}(|X| > t)$  tend vers zéro quand  $t \rightarrow \infty$  moins rapidement que dans le cas gaussien ; on parle de *queue lourde*. Ainsi, une variable leptokurtique voit ses réalisations prendre des valeurs extrêmes (on parle d'*outliers* en anglais) plus souvent qu'une variable gaussienne.

### 6.2 Interlude mathématique : Espérance conditionnelle (rappels)

Etant donné un espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$ , une variable aléatoire  $X \in L^2 = L^2(\Omega, \mathcal{F}, \mathbb{P})$ <sup>♣</sup> et  $\mathcal{G}$  une sous-tribu de  $\mathcal{F}$ , on définit l'*espérance conditionnelle*  $\mathbb{E}[X|\mathcal{G}]$  comme la projec-

---

?. En finance un *actif (financier)* est un produit financier au sens général : action d'une entreprise, matière première (or, pétrole, etc), dette publique, etc.

♠. L'indice "de la peur" calculé par le Chicago Board Option Exchange sur le marché américain.

♣. On pourrait aussi considérer une variable aléatoire  $X \in L^1$  et étendre la définition d'espérance conditionnelle à ces variables mais on n'en aura pas besoin dans la suite.

tion orthogonale de  $X$  sur le sous-espace de  $L^2$  des variables aléatoires  $\mathcal{G}$ -mesurables  $\heartsuit$ . En d'autres termes,  $\mathbb{E}[X|\mathcal{G}]$  est la meilleure approximation (dans  $L^2$ ) de  $X$  par une variable aléatoire  $\mathcal{G}$ -mesurable. On rappelle que si  $\mathcal{G} = \sigma(X_i : i \in I)$  où  $(X_i)_{i \in I}$  est une collection dénombrable de variables aléatoires, alors une variable aléatoire  $Z$  est  $\mathcal{G}$ -mesurable si et seulement si il existe une fonction  $f$  mesurable telle que  $Z = f(X_i : i \in I)$ .

*L'espérance conditionnelle est donc caractérisée par les propriétés suivantes :*

- (1)  $\mathbb{E}[X|\mathcal{G}]$  est  $\mathcal{G}$ -mesurable.
- (2) Si  $Y \in L^2$  est  $\mathcal{G}$ -mesurable, alors  $\mathbb{E}[XY] = \mathbb{E}[Y\mathbb{E}[X|\mathcal{G}]]$ .

En effet, la projection orthogonale est caractérisée par le fait que  $X - \mathbb{E}[X|\mathcal{G}]$  est orthogonale à l'espace des variables aléatoires  $L^2$   $\mathcal{G}$ -mesurables, ce qui donne (2).

Quelques propriétés utiles pour des variables aléatoires  $X, Y \in L^2$  :

- On a toujours  $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$ .
- Si  $Y$  est  $\mathcal{G}$ -mesurable, alors  $\mathbb{E}[XY|\mathcal{G}] = Y\mathbb{E}[X|\mathcal{G}]$ .
- Si  $X$  est indépendante  $?$  de  $\mathcal{G}$ , alors  $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$ .

Aussi, si on a des tribus emboîtées  $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ , alors

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}].$$

On définit également

$$\text{Var}[X|\mathcal{G}] := \mathbb{E}[X^2|\mathcal{G}] - \mathbb{E}[X|\mathcal{G}]^2 = \mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}])^2|\mathcal{G}]$$

et

$$\text{Cov}(X, Y|\mathcal{G}) := \mathbb{E}[XY|\mathcal{G}] - \mathbb{E}[X|\mathcal{G}]\mathbb{E}[Y|\mathcal{G}]$$

**Exercice :** Monter la formule

$$\text{Var}[X] = \mathbb{E}(\text{Var}[X|\mathcal{G}]) + \text{Var}(\mathbb{E}[X|\mathcal{G}]). \quad (6.1)$$

Une observation importante est la suivante.

**Remarque 6.1.** *On peut aussi faire le lien avec le prédicteur linéaire optimal dans le cas gaussien : Rappelons que  $\mathcal{H}_{t-1, m} := \text{Vect}(X_{t-1}, \dots, X_{t-m})$  et supposons de plus cette fois que  $(X_{t-m}, \dots, X_t)$  est un vecteur gaussien  $\clubsuit$  centré. On a alors*

$$X_t^{*, m} := P_{\mathcal{H}_{t-1, m}}(X_t) = \mathbb{E}[X_t | X_{t-1}, \dots, X_{t-m}]$$

où l'on a noté  $\mathbb{E}[X_t | X_{t-1}, \dots, X_{t-m}] := \mathbb{E}[X_t | \sigma(X_{t-1}, \dots, X_{t-m})]$ . En conclusion, dans ce cadre gaussien, le prédicteur linéaire optimal  $X_t^{*, m}$  n'est pas seulement la meilleure approximation (dans  $L^2$ ) de  $X_t$  par une fonction linéaire de  $X_{t-1}, \dots, X_{t-m}$ , mais la meilleure approximation par une fonction mesurable de ces variables. On ne peut pas faire faire mieux !

---

$\heartsuit$ .  $Y : \Omega \rightarrow \mathbb{R}$  est par définition  $\mathcal{G}$ -mesurable si  $Y^{-1}(A) \in \mathcal{G}$  pour tout Borelien  $A \subset \mathbb{R}$ .

$?$ . Cela veut dire que  $X$  est indépendante des variables  $\mathbf{1}_A$  pour tout  $A \in \mathcal{G}$ . Si  $\mathcal{G} = \sigma(Y_i : i \in I)$  où  $(Y_i)_{i \in I}$  est une collection quelconque de variables aléatoires, cela revient à supposer que  $X$  est indépendante des  $(Y_i)_{i \in I}$ .

$\clubsuit$ . C'est ce qui arrive par exemple lorsque  $(X_t)$  est un ARMA associé à un bruit blanc gaussien.

Donnons-en une preuve : on a par définition de  $X_t^{*,m}$  que

$$\text{Cov}(X_t - X_t^{*,m}, X_{t-k}) = \mathbb{E}[(X_t - X_t^{*,m})X_{t-k}] = 0$$

pour tout  $1 \leq k \leq m$  et, comme  $(X_t - X_t^{*,m}, X_{t-1}, \dots, X_{t-m})$  est un vecteur gaussien (car combinaison linéaire de vecteurs gaussiens), cela montre que  $X_t - X_t^{*,m}$  est indépendant des  $X_{t-k}$  pour  $1 \leq k \leq m$ . Ainsi,

$$\mathbb{E}[X_t - X_t^{*,m} | X_{t-1}, \dots, X_{t-m}] = \mathbb{E}[X_t - X_t^{*,m}] = 0$$

c'est-à-dire

$$\mathbb{E}[X_t | X_{t-1}, \dots, X_{t-m}] = \mathbb{E}[X_t^{*,m} | X_{t-1}, \dots, X_{t-m}].$$

Finalement, comme  $X_t^{*,m}$  est  $\sigma(X_{t-1}, \dots, X_{t-m})$  mesurable (c'est une combinaison linéaire de ces variables), on a  $\mathbb{E}[X_t^{*,m} | X_{t-1}, \dots, X_{t-m}] = X_t^{*,m}$  et cela conclut la preuve.

### 6.3 Le cadre général

On rappelle qu'on considère une suite de variables aléatoires  $(X_t)_{t \in \mathbb{Z}}$  définies sur le même espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$ , et on suppose que  $X_t \in L^2$  pour tout  $t \in \mathbb{Z}$ . On note  $\mathcal{F}_t := \sigma(X_s : s \leq t)$  la tribu engendrée par les variables aléatoires  $X_t, X_{t-1}, \dots$ . C'est donc une sous-algèbre de  $\mathcal{F}$ . On définit alors la *volatilité*  $\sigma_t$  de  $X_t$  comme l'écart-type conditionnel de  $X_t$  sachant  $\mathcal{F}_{t-1}$ , c'est-à-dire

$$\sigma_t := \sqrt{\text{Var}[X_t | \mathcal{F}_{t-1}]}, \quad t \in \mathbb{Z}.$$

Par définition de l'espérance conditionnelle la variable aléatoire  $X_t - \mathbb{E}[X_t | \mathcal{F}_{t-1}]$  est orthogonale à  $\mathcal{F}_{t-1}$ ; on l'appelle l'*innovation* au temps  $t$ . Quitte à étudier  $X_t - \mathbb{E}[X_t | \mathcal{F}_{t-1}]$ , après avoir modélisé la série  $\mathbb{E}[X_t | \mathcal{F}_{t-1}]$  de façon pertinente (via un modèle ARMA ou autre), on fait l'hypothèse de travail que

$$\mathbb{E}[X_t | \mathcal{F}_{t-1}] = 0, \quad t \in \mathbb{Z}.$$

**Modèle général :** En fait, dans la suite on considère des séries temporelles de la forme

$$X_t = \sigma_t \varepsilon_t \tag{6.2}$$

où l'on suppose que :

- $(\varepsilon_t)_{t \in \mathbb{Z}} \sim \text{BB}(0, 1)$  fort et  $(\varepsilon_t)_{t \in \mathbb{Z}} \in L^2$
- $\varepsilon_t$  est indépendante de  $\mathcal{F}_{t-1}$  pour tout  $t \in \mathbb{Z}$ ,
- $(\sigma_t)_{t \in \mathbb{Z}} \in L^2$  et  $\sigma_t$  est  $\mathcal{F}_{t-1}$ -mesurable pour tout  $t \in \mathbb{Z}$ .

**Attention :** Ces séries temporelles n'ont aucune raison d'être stationnaires a priori.

**Proposition 6.2.** *On a les propriétés suivantes :*

- $\mathbb{E}[X_t | \mathcal{F}_{t-1}] = 0$ , et en particulier  $\mathbb{E}[X_t] = 0$ , pour tout  $t \in \mathbb{Z}$ .
- $\text{Var}[X_t | \mathcal{F}_{t-1}] = \sigma_t^2$ ; en particulier la volatilité de  $X_t$  est bien donnée par  $\sigma_t$ .
- $\text{Cov}(X_t, X_{t+h}) = 0$  pour tout  $h \neq 0$ .

En particulier on voit que  $(X_t)$  est stationnaire si et seulement si  $\text{Var}[X_t]$  ne dépend pas de  $t$ , et dans ce cas,  $(X_t)$  est un bruit blanc (faible). Notez que l'identité (6.1) entraîne que  $\text{Var}[X_t] = \mathbb{E}(\text{Var}[X_t|\mathcal{F}_{t-1}]) = \mathbb{E}[\sigma_t^2]$ .

*Démonstration.* D'abord, pour tout  $t \in \mathbb{Z}$ ,

$$\mathbb{E}[X_t|\mathcal{F}_{t-1}] = \mathbb{E}[\sigma_t \varepsilon_t|\mathcal{F}_{t-1}] = \sigma_t \mathbb{E}[\varepsilon_t|\mathcal{F}_{t-1}] = \sigma_t \mathbb{E}[\varepsilon_t] = 0.$$

En particulier,  $\mathbb{E}[X_t] = \mathbb{E}[\mathbb{E}[X_t|\mathcal{F}_{t-1}]] = 0$ . Ensuite, on a

$$\text{Var}[X_t|\mathcal{F}_{t-1}] = \mathbb{E}[X_t^2|\mathcal{F}_{t-1}] = \sigma_t^2 \mathbb{E}[\varepsilon_t^2] = \sigma_t^2.$$

Aussi, pour tout  $h > 0$ , en utilisant que  $\mathcal{F}_{t-1} \subset \mathcal{F}_{t+h-1}$  on obtient

$$\begin{aligned} \mathbb{E}[X_t X_{t+h}|\mathcal{F}_{t-1}] &= \mathbb{E}[\mathbb{E}[X_t X_{t+h}|\mathcal{F}_{t+h-1}]|\mathcal{F}_{t-1}] \\ &= \mathbb{E}[X_t \mathbb{E}[X_{t+h}|\mathcal{F}_{t+h-1}]|\mathcal{F}_{t-1}] \\ &= \mathbb{E}[X_t \sigma_{t+h} \underbrace{\mathbb{E}[\varepsilon_{t+h}]_{=0}}_{=0}|\mathcal{F}_{t-1}] = 0 \end{aligned}$$

et donc

$$\text{Cov}(X_t, X_{t+h}) = \mathbb{E}[X_t X_{t+h}] = \mathbb{E}[\mathbb{E}[X_t X_{t+h}|\mathcal{F}_{t-1}]] = 0.$$

Quand  $h < 0$ , le même calcul montre que  $\text{Cov}(X_t, X_{t+h}) = 0$  en calculant  $\mathbb{E}[X_t X_{t+h}|\mathcal{F}_{t+h-1}]$  grace à l'inclusion maintenant inversée  $\mathcal{F}_{t+h-1} \subset \mathcal{F}_{t-1}$ .  $\square$

## 6.4 Modèles ARCH & GARCH

Terminologie : *ARCH* = *Auto-Regressive Conditional Heteroscedasticity*  $\spadesuit$ .

On dit que  $(X_t)$  est un *ARCH*( $p$ ) si  $X_t = \sigma_t \varepsilon_t$  satisfait les hypothèses du modèle (6.2), est stationnaire, et

$$\sigma_t^2 = \omega + \alpha_1 X_{t-1}^2 + \dots + \alpha_p X_{t-p}^2$$

avec  $\omega > 0$  et  $\alpha_1, \dots, \alpha_p \geq 0$  tels que  $\alpha_1 + \dots + \alpha_p < 1$ . Ces conditions sur les paramètres assurent l'existence et l'unicité d'un tel processus stationnaire. Notez que  $\sigma_t$  est  $\mathcal{F}_{t-1}$ -mesurable par définition.

Par construction, la volatilité  $\sigma_t$  de  $X_t$  est d'autant plus grande que l'amplitude des valeurs du passé  $X_{t-1}, \dots, X_{t-p}$  l'est. Comme on le verra en TP, cela permet de reproduire les "clusters" de volatilité des faits stylisés décrits au début du chapitre.

**Exercice 6.3.** Montrer que pour  $(X_t)$  un *ARCH*( $p$ ), on a

$$\text{Var}(X_t) = \frac{\omega}{1 - \alpha_1 - \dots - \alpha_p}.$$

**Exercice 6.4.** Montrer que pour  $(X_t)$  un *ARCH*(1), on a l'égalité  $\heartsuit$  dans  $L^1$

$$X_t^2 = \omega \sum_{k=0}^{\infty} \alpha_1^k \prod_{j=0}^k \varepsilon_{t-j}^2.$$

$\spadesuit$ . On parle d'*hétéroscédasticité* lorsque la variance d'une suite de variables aléatoire dépend de leur indice. On parle d'*homoscédasticité* dans le cas contraire.

$\heartsuit$ . Rappelons qu'on a égalité entre variables aléatoires  $X = \sum_{k=0}^{\infty} X_k$  dans  $L^p \Leftrightarrow \mathbb{E}|X - \sum_{k=0}^N X_k|^p \rightarrow 0$  quand  $N \rightarrow \infty$ .

Revenons sur les faits stylisés décrits en introduction, en particulier l'aspect leptokurtique. Soit  $(X_t)$  un ARCH(1), avec  $\sigma_t^2 = \omega + \alpha X_{t-1}^2$  pour  $\omega > 0$  et  $0 < \alpha < 1$ . Supposons de plus que  $\mathbb{E}[X_t^4] < \infty$  (qui ne dépend pas de  $t$ ); on verra dans la suite ce que cela implique. On calcule alors

$$\begin{aligned}\mathbb{E}[X_t^4] &= \mathbb{E}[\varepsilon_t^4 \sigma_t^4] \\ &= \mathbb{E}[\varepsilon_t^4] \mathbb{E}[\sigma_t^4] = \mathbb{E}[\varepsilon_t^4] \mathbb{E}[(\omega + \alpha X_{t-1}^2)^2] \\ &= \mathbb{E}[\varepsilon_t^4] (\omega^2 + \frac{2\alpha\omega^2}{1-\alpha} + \alpha^2 \mathbb{E}[X_t^4]) \\ &= \mathbb{E}[\varepsilon_t^4] (\omega^2 \frac{1+\alpha}{1-\alpha} + \alpha^2 \mathbb{E}[X_t^4])\end{aligned}$$

ce qui donne

$$\mathbb{E}[X_t^4] = \frac{\omega^2(1+\alpha)\mathbb{E}[\varepsilon_t^4]}{(1-\alpha)(1-\alpha^2\mathbb{E}[\varepsilon_t^4])}.$$

Ainsi, pour avoir  $\mathbb{E}[X_t^4] < \infty$  fini, il faut supposer que  $\mathbb{E}[\varepsilon_t^4] < \infty$  et que  $\alpha < 1/\sqrt{\mathbb{E}[\varepsilon_t^4]}$ .

Si par exemple, on prend  $(\varepsilon_t)$  une suite i.i.d de variables  $\mathcal{N}(0, 1)$ , on a  $\mathbb{E}[\varepsilon_t^4] = 3$  et

$$\mathbb{E}[X_t^4] = \frac{3\omega^2(1+\alpha)}{(1-\alpha)(1-3\alpha^2)} \quad \text{et} \quad \kappa(X_t) = 3 \frac{1-\alpha^2}{1-3\alpha^2} > 3,$$

ce qui montre bien qu'un tel ARCH(1) est leptokurtique dès que  $\alpha < 1/\sqrt{3}$ .

Finalement, on a la propriété suivante qui fait le lien avec le modèle AR(1).

**Proposition 6.5.** *Si  $(X_t)$  est un ARCH(1) et  $\mathbb{E}[X_t^4] < \infty$  alors  $(X_t^2)$  est un AR(1) : Il existe un bruit blanc  $(\eta_t) \sim BB(0, \eta^2)$  avec  $\eta^2 > 0$  tel que*

$$X_t^2 = \omega + \alpha X_{t-1}^2 + \eta_t, \quad t \in \mathbb{Z}.$$

*Démonstration.* Si on pose  $\eta_t := X_t^2 - \sigma_t^2$ , en utilisant que  $\sigma_t^2 = \omega + \alpha X_{t-1}^2$  il suffit de montrer que  $\eta_t$  est un bruit blanc centré pour prouver la proposition. Comme  $\eta_t = \sigma_t^2(\varepsilon_t^2 - 1)$  on obtient, par indépendance de  $\sigma_t$  et  $\varepsilon_t$ ,

$$\mathbb{E}[\eta_t] = \mathbb{E}[\sigma_t^2](\mathbb{E}[\varepsilon_t^2] - 1) = 0.$$

Ensuite, comme  $X_t \in L^4$  nécessairement  $\mathbb{E}[\varepsilon_t^4] < \infty$  (et ne dépend pas de  $t$ ) et, comme  $\mathbb{E}[\sigma_t^4]$  ne dépend pas de  $t$ , on voit que

$$\text{Var}[\eta_t] = \mathbb{E}[\eta_t^2] = \mathbb{E}[\sigma_t^4] \mathbb{E}[(\varepsilon_t^2 - 1)^2] = \mathbb{E}[\sigma_t^4](\mathbb{E}[\varepsilon_t^4] - 1)$$

ne dépend pas de  $t$ . Finalement, pour  $h < 0$ , on a

$$\begin{aligned}\text{Cov}(\eta_t, \eta_{t+h}) &= \mathbb{E}[\eta_t \eta_{t+h}] \\ &= \mathbb{E}[\mathbb{E}[\eta_t \eta_{t+h} | \mathcal{F}_{t-1}]] \\ &= \mathbb{E}[\sigma_t^2 \sigma_{t+h}^2 (\varepsilon_{t+h}^2 - 1) \mathbb{E}[(\varepsilon_t^2 - 1) | \mathcal{F}_{t-1}]] \\ &= \mathbb{E}[\sigma_t^2 \sigma_{t+h}^2 (\varepsilon_{t+h}^2 - 1) \underbrace{\mathbb{E}[(\varepsilon_t^2 - 1)]}_{=0}] = 0,\end{aligned}$$

où l'on a utilisé que  $\varepsilon_{t+h} = X_{t+h}/\sigma_{t+h}$  est  $\mathcal{F}_{t-1}$ -mesurable. Le cas où  $h > 0$  s'obtient de la même façon mais en conditionnant par  $\mathcal{F}_{t+h-1}$  à la place de  $\mathcal{F}_{t-1}$ .  $\square$



En conclusion, sous des hypothèses raisonnables, le modèle ARCH(1) est un modèle dont la volatilité est leptokurtique et tel que  $(X_t^2)$  suit un modèle AR(1) de coefficient “ $\varphi_1$ ” =  $\alpha > 0$ , ce qui permet en particulier peut reproduire les clusters de volatilité. Par contre l’asymétrie de l’effet levier n’est pas modélisée.

**Le cas  $p$  général.** Plus généralement, on a le résultat suivant, que l’on admet :

**Proposition 6.6.** *Si  $(X_t)$  est un ARCH(p) alors  $(X_t) \sim BB(0, \sigma^2)$  où*

$$\sigma^2 = \frac{\omega}{1 - \sum_{i=1}^p \alpha_i}.$$

*De plus, si  $\mathbb{E}(X_t^4) < \infty$  alors  $(X_t^2)$  est un AR(p) :*

$$X_t^2 = \omega + \sum_{j=1}^p \alpha_j X_{t-j}^2 + \eta_t,$$

*avec  $(\eta_t) \sim BB(0, \eta^2)$  où  $\eta^2 > 0$ .*

En bref, si  $(X_t) \sim \text{ARCH}(p)$  alors  $(X_t)$  est une bruit blanc (faible) leptokurtique et  $(X_t^2) \sim \text{AR}(p)$ .

**Détecter l’effet ARCH et prédire la volatilité.** En pratique, pour détecter un “effet ARCH” dans une série temporelle  $x_1, \dots, x_n$  qui ressemble à un bruit blanc, par exemple à la vue de son ACF, on considère les données au carré  $x_1^2, \dots, x_n^2$  et on leur fait passer un test de type Box-Pierce. Si  $H_0 = \{\text{bruit blanc fort}\}$  est rejeté, on peut suspecter un effet ARCH. On peut alors tester l’hypothèse  $H_0 = \{\alpha_1 = \dots = \alpha_p = 0\}$  dans le modèle linéaire  $X_t = \omega + \alpha_1 X_{t-1}^2 + \dots + \alpha_p X_{t-p}^2$  (test de Fisher ; F-test). Le rejet d’ $H_0$  rend le modèle ARCH pertinent, et on peut donc modéliser  $(X_t)$  par un ARCH(p) et obtenir une estimation de ses coefficients  $\alpha_1, \dots, \alpha_p$ , par exemple à l’aide de l’algorithme de Durbin-Levinson que l’on a vu précédemment. On peut finalement faire de la prédiction sur la volatilité  $\sigma_t$  car, par définition,  $\sigma_t = \sqrt{\alpha_1 X_{t-1}^2 + \dots + \alpha_p X_{t-p}^2}$ .

**Le modèle ARCH généralisé : GARCH.** Une version généralisée du modèle ARCH est le GARCH, pour Generalized ARCH. On dit que  $X_t = \varepsilon_t \sigma_t$  est un *GARCH(p, q)* si c’est un processus stationnaire qui satisfait les hypothèses (6.2) tel que

$$\sigma_t^2 = \omega + \sum_{j=1}^p \alpha_j X_{t-j}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

où les paramètres satisfont

$$\omega, \alpha_p, \beta_q > 0, \quad \alpha_j, \beta_j \geq 0, \quad \sum_{j=1}^p \alpha_j + \sum_{j=1}^q \beta_j < 1.$$

On a alors le résultat suivant (admis) :

**Proposition 6.7.** Si  $(X_t)$  est un GARCH(p,q) alors  $(X_t) \sim BB(0, \sigma^2)$  où

$$\sigma^2 = \frac{\omega}{1 - \sum_{j=1}^p \alpha_j - \sum_{j=1}^q \beta_j}.$$

De plus, si  $\mathbb{E}(X_t^4) < \infty$  alors  $(X_t^2)$  est un ARMA(m,q) avec  $m := \max(p, q)$  :

$$X_t^2 = \omega + \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) X_{t-j}^2 + \eta_j - \sum_{j=1}^q \beta_j \eta_{t-j}$$

avec  $(\eta_t) \sim BB(0, \eta^2)$  où  $\eta^2 > 0$ .

Il faut donc faire attention au fait que si  $(X_t) \sim \text{GARCH}(p,q)$  alors  $(X_t^2)$  n'est *pas forcément* un ARMA(p,q) mais un ARMA(max(p,q),q). Expliquons ce fait : si on pose  $\eta_t := X_t^2 - \sigma_t^2$  avec un peu d'effort on montre, comme pour l'ARCH(1), que  $(\eta_t)$  est un bruit blanc. On a alors

$$\begin{aligned} X_t^2 &= \sigma_t^2 + \eta_t \\ &= \omega + \sum_{j=1}^p \alpha_j X_{t-j}^2 + \sum_{j=1}^q \beta_j \underbrace{\sigma_{t-j}^2}_{=X_{t-j}^2 + \eta_{t-j}} - \eta_j \\ &= \omega + \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) X_{t-j}^2 + \eta_j - \sum_{j=1}^q \beta_j \eta_{t-j}, \end{aligned}$$

qui correspond bien à un modèle ARMA(max(p,q),q) pour  $(X_t^2)$  avec coefficients “ $\varphi_j$ ” =  $\alpha_j + \beta_j$  et “ $\theta_j$ ” =  $-\beta_j$ .

**Remarque 6.8.** Les conditions sur les paramètres  $\alpha_j, \beta_j$  pour que  $\mathbb{E}(X_t^4) < \infty$  ne sont malheureusement pas explicites en général.

Si ces modèles offrent plus de liberté de paramètres que l'ARCH(1) et modélisent également les clusters de volatilité et la condition de leptokurtocité, ils ne modélisent pas non plus l'effet levier.

**Modéliser l'effet levier ?** Il existe plusieurs variations autour des modèles ARCH pour modéliser l'asymétrie propre à l'effet levier, comme le TGARCH (Treshold GARCH), l'APARCH (Asymetric Power ARCH), ou l'EGARCH (Exponential ARCH), où l'on introduit de force l'asymétrie désirée.

## Références

- [BrDa] P. J. Brockwell, R. A. Davis, *Time Series Theory and Methods*, Springer Series in Statistics, Second edition.
- [Zak] J.-M. Zakoïan, *Modèles ARCH : une revue de la littérature*, Journal de la société statistique de Paris, tome 133, no 1-2 (1992), p. 40-57. [Lien](#)