

Université des Frères Mentouri Constantine 1  
Faculté des Sciences Exactes  
Département de Mathématiques

Cours Master 2 - SPA

**STATISTIQUE DES PROCESSUS  
ALÉATOIRES**

**Dr. Meghlaoui DAKHMOUCHE**

**Année universitaire : 2017/2018**

# Table des matières

<b>1</b>	<b>Processus stochastiques et séries chronologiques</b>	<b>4</b>
1.1	Séries chronologiques . . . . .	5
1.2	Processus stochastiques . . . . .	5
1.2.1	Processus stochastiques stationnaires . . . . .	6
1.3	Modèles mathématiques stochastiques . . . . .	9
1.3.1	Modèles stochastiques stationnaires . . . . .	10
1.3.2	Modèles autorégressifs . . . . .	13
1.3.3	Modèles moyennes mobiles . . . . .	14
1.3.4	Modèles mixtes autorégressif-moyennes mobiles . . . .	15
1.4	Modèles nonstationnaires . . . . .	16
<b>2</b>	<b>Fonction d'autocorrélation et spectre des processus stationnaires</b>	<b>18</b>
2.1	Propriétés des autocorrélations des processus stationnaires . .	18
2.1.1	Fonctions d'autocovariance et d'autocorrélation . . . .	18
2.1.2	Spectre et densité spectrale . . . . .	23
2.1.3	Lien entre le spectre empirique et l'estimateur de la fonction d'autocovariance . . . . .	28
<b>3</b>	<b>Modèles stationnaires linéaires</b>	<b>30</b>
3.1	Processus linéaire général . . . . .	30
3.1.1	Deux formes équivalentes pour un processus linéaire . .	30
3.1.2	Fonction génératrice des autocovariances d'un processus linéaire . . . . .	33
3.1.3	Conditions de stationnarité et d'inversibilité pour un processus linéaire . . . . .	35
3.1.4	Processus autorégressif et moyenne mobile . . . . .	38
3.2	Processus autoregressifs . . . . .	41

3.2.1	Conditions de stationnarité pour un processus autorégressif	41
3.2.2	Fonction d'autocorrélation et spectre d'un processus autorégressif . . . . .	43
3.2.3	Processus autorégressif du premier ordre . . . . .	45
3.2.4	Processus autorégressif du second ordre . . . . .	47
3.2.5	Fonction d'autocorrélation partielle . . . . .	49
3.2.6	Estimation de la fonction d'autocorrélation partielle . .	51
3.2.7	Erreurs standards des estimateurs des autocorrélations partielles . . . . .	52
3.2.8	Calculs par R . . . . .	52
3.3	Processus moyenne mobile . . . . .	53
3.3.1	Conditions d'inversibilité pour les processus moyenne mobile . . . . .	53
3.3.2	Fonction d'autocorrélation et spectre d'un processus moyenne mobile . . . . .	54
3.3.3	Processus moyenne mobile du premier ordre . . . . .	56
3.3.4	Processus moyen mobile du second ordre . . . . .	58
3.3.5	Dualité entre les processus autorégressifs et les processus moyenne mobile . . . . .	59
3.4	Processus mixte autorégressif-moyenne mobile . . . . .	61
3.4.1	Propriétés de stationnarité et d'inversibilité . . . . .	61
3.4.2	Fonction d'autocorrélation et spectre des processus mixtes	62
3.4.3	Processus autorégressif du premier ordre-moyenne mobile du premier ordre . . . . .	64
<b>4</b>	<b>Modèles linéaires non stationnaires</b>	<b>67</b>
4.1	Processus autorégressif-moyenne mobile intégré . . . . .	67
4.1.1	Processus autorégressif de premier ordre non stationnaire	67
4.1.2	Modèle général pour un processus non stationnaire présentant une homogénéité . . . . .	68
4.1.3	Forme générale d'un modèle ARIMA . . . . .	73
4.2	Deux formes explicites pour le modèle ARIMA . . . . .	76
4.2.1	Modèle sous forme d'équation aux différences . . . . .	76
4.2.2	Modèle sous forme d'innovations aléatoires . . . . .	77
<b>5</b>	<b>Modèles conditionnels hétéroscédastiques</b>	<b>85</b>
5.1	Le modèle <i>ARCH</i> . . . . .	86

5.1.1	Quelques propriétés du modèle <i>ARCH</i> . . . . .	88
5.1.2	Erreurs de prévision pour le modèle ARCH . . . . .	90
5.2	Le modèle <i>GARCH</i> . . . . .	92

# Introduction

Le cours que nous présentons est orienté sur les aspects les plus sensibles dans l'étude des processus aléatoires et des séries chronologiques. Il est abordé sous l'aspect probabiliste et complété par son aspect statistique orienté vers la modélisation des phénomènes aléatoires dépendants du temps. Il existe principalement deux types de paramétrisation des processus aléatoires. Le premier type que l'on intitule représentation externe des processus, est un résumé des travaux de Box et Jenkins, Brockwell et Davis, et Hamilton principalement. Cette forme d'expression des modèles stochastiques est basée sur le théorème de décomposition de Wold et sur l'idée de G.U. Yule qui stipule qu'un phénomène dont les réalisations successives sont fortement dépendantes entre elles, peut être considéré comme étant le résultat du passage à travers un filtre linéaire d'une suite de variables aléatoires indépendantes et centrées. Dans notre exposé on adoptera les notations de Box et Jenkins pour décrire les modèles.

Le deuxième type de paramétrisation appelé représentation interne d'un processus, est en fait la représentation de Kalman-Bucy ou représentation Etat-Espace. Cette technique a été développée dans le cadre du contrôle des systèmes linéaires. Elle est résumée par un système de deux équations, que l'on appellera aussi filtre de Kalman-Bucy.

Une série chronologique est une suite d'observations d'un processus aléatoire prises séquentiellement dans le temps. Une caractéristique intrinsèque d'une série chronologique est que les observations adjacentes sont dépendantes. La nature de cette dépendance entre les observations d'une série chronologique est d'un intérêt pratique considérable. L'analyse des séries chronologiques est l'ensemble des techniques d'analyse de cette dépendance. Ceci nécessite le développement de modèles stochastiques pour les séries chronologiques et l'utilisation de tels modèles dans des domaines d'application importants.

Les méthodes discutées sont appropriées pour les systèmes discrets (données échantillonnées), où l'observation du processus a lieu à des intervalles de temps équidistants. Nous illustrons l'utilisation de ces séries chronologiques dans la prévision des valeurs futures d'une série chronologique à partir des valeurs présente et passées.

# Chapitre 1

## Processus stochastiques et séries chronologiques

Un modèle décrivant la structure de probabilité d'une suite d'observations dans le temps est appelé processus stochastique. Une série chronologique de  $N$  observations successives  $\mathbf{X}' = (X_1, X_2, \dots, X_N)$ , est considérée comme une réalisation d'un échantillon, à partir d'une population infinie de tels échantillons, qui auraient pu être générés par le processus aléatoire.

Un objectif majeur de la statistique des processus est de déduire les propriétés et le type de processus aléatoire à partir de celles de la série chronologique. Par exemple, pour faire une prévision, on doit déduire la distribution de probabilité d'une observation future de la population, à partir d'un ensemble de valeurs passées. Pour ce faire, nous avons besoin de méthodes de représentation des processus stochastiques, et nous avons également besoin d'un ensemble de modèles stochastiques capables de décrire des situations qui se produisent dans la pratique.

Une classe importante de processus stochastiques que l'on étudiera est la classe des processus stationnaires. Ils sont supposés être dans une certaine forme d'équilibre statistique, et en particulier, de varier dans le temps de manière stable autour d'une moyenne constante. Les dispositifs utiles pour décrire le comportement des processus stationnaires sont la fonction d'autocorrélation et le spectre.

## 1.1 Séries chronologiques

Une série chronologique est un ensemble d'observations générées séquentiellement dans le temps. Les observations d'une série chronologique réalisées aux instants  $t_1, t_2, \dots, t_N$  peuvent être désignées par  $X(t_1), X(t_2), \dots, X(t_N)$ . Dans ce cours, nous considérons seulement les séries chronologiques où les observations sont réalisées à intervalle fixe  $h$ . Nous noterons  $X_1, X_2, \dots, X_N$  pour désigner les observations faites à des intervalles de temps équidistants  $t_0 + h, t_0 + 2h, \dots, t_0 + Nh$ . Si nous adoptons  $t_0$  comme origine et  $h$  comme unité de temps, nous pouvons considérer  $X_t$  comme l'observation à l'instant  $t$ . Les séries chronologiques discrètes peuvent se présenter de deux façons, par échantillonnage d'un processus aléatoire ou par accumulation d'une variable stochastique sur une période donnée.

## 1.2 Processus stochastiques

Un phénomène statistique évoluant dans le temps suivant une loi de probabilité s'appelle un processus stochastique et on le notera  $X_t$ . Si le temps  $t$  varie dans un ensemble continu on dira que le processus est continu, si  $t$  varie dans un ensemble discret on dira que le processus est discret. Nous nous référerons souvent à  $X_t$  simplement comme un processus, en omettant le mot "stochastique". La série chronologique à analyser peut alors être considérée comme une réalisation particulière, produite par le mécanisme de probabilité sous-jacent, du système étudié.

Une caractéristique centrale du développement de modèles de séries chronologiques est l'hypothèse d'une certaine forme d'équilibre statistique. Une hypothèse particulièrement utile de ce genre est celle de la stationnarité. Habituellement, une série chronologique stationnaire peut être décrite par sa moyenne, sa variance et sa fonction d'autocorrélation ou de façon équivalente par sa moyenne, sa variance et sa fonction de densité spectrale. Dans ce qui suit, nous considérons les propriétés de ces fonctions et, en particulier, les propriétés de la fonction d'autocorrélation, qui seront largement utilisées dans le développement de modèles pour des séries chronologiques.



### 1.2.1 Processus stochastiques stationnaires

Une classe très spéciale de processus stochastiques, appelés processus stationnaires, repose sur l'hypothèse que le processus est dans un état particulier d'équilibre statistique. Un processus stochastique est dit strictement stationnaire si ses propriétés ne sont pas affectées par un changement d'origine temporelle, c'est-à-dire si la distribution de probabilité conjointe associée à  $m$  observations  $X_{t_1+k}, X_{t_2+k}, \dots, X_{t_m+k}$ , réalisées aux instants  $t_1, t_2, \dots, t_m$ , est la même que celle associée à  $m$  observations  $X_{t_1+k}, X_{t_2+k}, \dots, X_{t_m+k}$ , réalisées aux instants  $t_1 + k, t_2 + k, \dots, t_m + k$ . Ainsi, pour qu'un processus discret soit strictement stationnaire, la distribution conjointe de tout ensemble d'observations ne doit pas être affectée en décalant tous les temps d'observation vers l'avant ou vers l'arrière de n'importe quelle quantité entière  $k$ .

#### Moyenne et variance d'un processus stationnaire

Lorsque  $m = 1$ , l'hypothèse de stationnarité implique que la distribution de probabilité  $P_{X_t}$  est la même à tout instant  $t$ . Si  $P_{X_t}$  admet une densité  $f(x)$ , le processus stochastique admet donc une moyenne constante

$$\mu = E(X_t) = \int_{-\infty}^{+\infty} x f(x) dx \quad (1.1)$$

qui définit le niveau autour duquel il fluctue, et une variance constante

$$\sigma_{X_t}^2 = E(X_t - \mu)^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx \quad (1.2)$$

qui mesure sa dispersion autour de ce niveau, et dans ce cas on dit que le processus est homoscédastique.

Puisque la distribution de probabilité  $f(x)$  est la même pour tous les temps  $t$ , sa forme peut être déduite en traçant l'histogramme des observations  $X_1, X_2, \dots, X_N$  formant la série chronologique observée. De plus, la moyenne  $\mu$  du processus stochastique peut être estimée par la moyenne de la série chronologique observée

$$\bar{X} = \frac{1}{N} \sum_{t=1}^N X_t \quad (1.3)$$

et la variance  $\sigma_{X_t}^2$  du processus stochastique peut être estimée par la variance de l'échantillon de la série chronologique observée

$$\hat{\sigma}_{X_t}^2 = \frac{1}{N} \sum_{t=1}^N (X_t - \bar{X})^2 \quad (1.4)$$

### Coefficients d'autocovariance et d'autocorrélation

L'hypothèse de stationnarié implique également que la distribution de probabilité conjointe  $P_{(X_{t_1}, X_{t_2})}$  est la même pour tous les instants  $t_1, t_2$  séparés par un intervalle constant. En particulier, il s'ensuit que la covariance entre les valeurs  $X_t$  et  $X_{t+k}$  à des intervalles de temps  $k$  constants (ou retard  $k$ ), doit être la même pour tout  $t$  sous l'hypothèse de stationnarité. Cette covariance est appelée *autocovariance* à décalage ou à horizon  $k$  et est définie par

$$\gamma_k = Cov(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)] \quad (1.5)$$

De même, l'autocorrélation à horizon  $k$  est définie telle que

$$\begin{aligned} \rho_k &= \frac{E\{(X_t - E(X_t))(X_{t+k} - E(X_{t+k}))\}}{\sqrt{(X_t - E(X_t))^2 (X_{t+k} - E(X_{t+k}))^2}} \\ &= \frac{E\{(X_t - E(X_t))(X_{t+k} - E(X_{t+k}))\}}{\sigma_{X_t}^2} \end{aligned}$$

puisque pour un processus stationnaire la variance  $\sigma_{X_t}^2 = \gamma_0$  est la même à l'instant  $t + k$  qu'à l'instant  $t$ .

Ainsi, l'autocorrélation à horizon  $k$ , c'est-à-dire la corrélation entre  $X_t$  et  $X_{t+k}$ , est

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad (1.6)$$

ce qui implique, en particulier, que  $\rho_0 = 1$ .

Il s'ensuit également pour un processus stationnaire que la nature de la distribution de probabilité conjointe  $P_{(X_t, X_{t+k})}$  des valeurs séparées par  $k$  intervalles de temps peut être devinée en traçant un diagramme de dispersion utilisant des couples de valeurs  $(X_t, X_{t+k})$  de la série chronologique, séparées par un intervalle constant.

## Propriétés de la matrice d'autocovariance

La matrice de covariance associée à un processus stationnaire pour les observations  $(X_1, X_2, \dots, X_n)$  faites à  $n$  instants successifs est définie telle que

$$\begin{aligned} \mathbf{\Gamma}_n &= \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \dots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \dots & \gamma_{n-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \dots & \gamma_{n-3} \\ \dots & \dots & \dots & \dots & \dots \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \dots & \gamma_0 \end{pmatrix} \\ &= \sigma_X^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{n-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \dots & 1 \end{pmatrix} = \sigma_X^2 \mathbf{P}_n \end{aligned} \quad (1.7)$$

Une matrice de covariance  $\mathbf{\Gamma}_n$  de cette forme, symétrique avec les éléments sur n'importe quelle diagonale constants, est appelée *matrice d'autocovariance* et la matrice de corrélation correspondante  $\mathbf{P}_n$  est appelée *matrice d'autocorrélation*. Les matrices  $\mathbf{\Gamma}_n$  et  $\mathbf{P}_n$  sont définies positives.

## Ergodicité

La propriété d'ergodicité lie les moyennes statistiques (effectuées sur l'espace des réalisations sous-jacent à la définition des variables aléatoires qui constituent le processus) et les moyennes temporelles (effectuées sur les fonctions du temps qui sont les réalisations du processus).

**Définition 1** *Un processus aléatoire  $X_t$  est ergodique si ses moments peuvent être obtenus comme des moyennes à partir d'une seule de ses réalisations.*

*En particulier :*

- *Il est ergodique du premier ordre si*

$$\frac{1}{T} \int_0^T X_t dt \xrightarrow{T \rightarrow \infty} m_X = E(X_t)$$

- *Il est ergodique du second ordre si*

$$\frac{1}{T} \int_0^T X_t X_{t+k} dt \xrightarrow{T \rightarrow \infty} \gamma_X(k) = Cov(X_t, X_{t+k})$$

Si l'on examine la première limite, on se rend compte que le membre gauche ne dépend pas du temps, et donc pour que cette équation puisse être vérifiée, il faut que le processus ait une moyenne constante. On peut donc affirmer que pour qu'un processus soit ergodique, il doit nécessairement être stationnaire, mais l'affirmation contraire est fausse.

## Prévision des séries chronologiques

L'utilisation à l'instant  $t$  des observations disponibles d'une série chronologique pour prévoir sa valeur à l'instant  $t + \ell$  peut servir de base pour (1) la planification économique et commerciale, (2) la planification de la production, (3) l'inventaire et le contrôle de la production, et (4) contrôle et optimisation des processus industriels, etc. Les prévisions sont généralement nécessaires sur une période connue sous le nom d'*horizon* (lead time), qui varie avec chaque problème.

Nous supposons que les observations sont disponibles à des intervalles équidistants. Par exemple, dans un problème de prévision des ventes, les volumes des ventes  $X_t$  du mois en cours  $t$  et les volumes des ventes  $X_{t-1}, X_{t-2}, X_{t-3}, \dots$  au cours des mois précédents pourraient être utilisés pour prévoir les ventes aux horizons  $\ell = 1, 2, 3, \dots, 12$  mois à venir. Notons  $\hat{X}_t(\ell)$  la prévision faite à l'origine  $t$  des ventes  $X_{t+\ell}$  à l'instant futur  $t + \ell$ , c'est-à-dire à l'horizon  $\ell$ . La fonction  $\hat{X}_t(\ell)$ , qui fournit les prévisions d'origine  $t$  et à horizon  $\ell = 1, 2, \dots$  sera appelée *fonction de prévision* d'origine  $t$ . L'objectif est d'obtenir une fonction de prévision telle que le carré moyen des écarts  $X_{t+\ell} - \hat{X}_t(\ell)$  entre les valeurs réelles et prévues soit aussi faible que possible pour tout  $\ell$ . Cette méthode de prévision est développée en supposant que la série chronologique  $X_t$  évolue dans un modèle stochastique de forme connue.

Dans la suite nous allons introduire une classe de modèles pour les séries chronologiques que l'on appellera modèles *ARIMA* (Autorégressive Integrated Moving Average models).

## 1.3 Modèles mathématiques stochastiques

L'idée d'utiliser un modèle mathématique pour décrire le comportement d'un phénomène physique est bien établie. En particulier, il est parfois possible de construire un modèle basé sur des lois physiques, ce qui nous permet de calculer la valeur d'une quantité dépendante du temps presque exacte-

ment à n'importe quel instant. Ainsi, nous pourrions calculer la trajectoire d'un missile lancé dans une direction connue avec une vitesse connue. Si un calcul exact était possible, un tel modèle serait entièrement déterministe. Cependant, il est probable qu'aucun phénomène n'est totalement déterministe, car des facteurs inconnus peuvent intervenir, comme la vitesse changeante du vent qui peut projeter le missile légèrement hors trajectoire. Dans de nombreux problèmes, nous devons considérer un phénomène dépendant du temps, comme les ventes mensuelles de journaux, dans lequel il existe de nombreux facteurs inconnus et pour lesquels il n'est pas possible de proposer un modèle déterministe permettant un calcul exact du comportement futur du phénomène. Néanmoins, il peut être possible de construire un modèle qui peut être utilisé pour calculer la probabilité qu'une valeur future soit comprise entre deux limites spécifiées. Un tel modèle est appelé modèle probabiliste ou modèle stochastique.

### 1.3.1 Modèles stochastiques stationnaires

Une classe importante de modèles stochastiques pour la description de séries chronologiques est l'ensemble des modèles stationnaires. Les modèles stationnaires supposent que le processus reste en équilibre statistique avec des propriétés probabilistes qui ne changent pas avec le temps. En particulier, il varie autour d'un niveau moyen constant et avec une variance constante. Cependant, les prévisions sont d'une importance particulière dans l'industrie, l'économie, la finance, etc., où de nombreuses séries chronologiques sont souvent mieux représentées par des modèles non stationnaires. Il n'est donc pas étonnant que bon nombre des méthodes de prévision économique initialement proposées par Holt (1957, 1963), Winters (1960), et Brown (1962), utilisant le lissage exponentiel, paraissent appropriées pour un type particulier de processus non stationnaire.

Le modèle stochastique pour lequel le lissage exponentiel fournit l'erreur quadratique moyenne minimale (Muth, 1960), est un élément d'une classe plus vaste de processus non stationnaires appelés processus autorégressif-moyenne mobile intégré (ARIMA), que l'on étudiera plus loin. Cette classe de processus fournit une gamme de modèles stationnaires et non stationnaires à l'aide desquels on peut représenter de façon adéquate plusieurs types de séries chronologiques rencontrées dans la pratique.

## Stationnarité faible

Nous avons vu que pour qu'un processus soit strictement stationnaire, toute la structure de probabilité ne doit dépendre que des différences entre les instants d'observation. Une exigence moins restrictive, appelée *stationnarité faible* d'ordre  $r$ , est que les moments jusqu'à un certain ordre  $r$  dépendent uniquement des différences entre les instants d'observation. Par exemple, l'existence d'une moyenne fixe  $\mu$  et d'une matrice d'autocovariance  $\mathbf{\Gamma}_n$  de la forme (1.7) est suffisante pour assurer la stationnarité jusqu'au second ordre. Autrement dit, un processus  $\{X_t\}$  est faiblement stationnaire (d'ordre 2), ou du second ordre, si la moyenne  $E(X_t) = \mu$  est une constante pour tout  $t$  et les autocovariances  $Cov(X_t, X_{t+k}) = \gamma_k$  ne dépendent que des différences entre les instants d'observation ou décalage  $k$  pour tout  $t$ . Ainsi, une stationnarité de second ordre et une hypothèse de normalité sont suffisantes pour produire une stationnarité stricte.

## Processus gaussiens

Si la distribution de probabilité conjointe des observations d'un phénomène aléatoire est une distribution normale multivariée, le processus est dit processus normal ou gaussien. Puisque la distribution normale multivariée est entièrement caractérisée par ses moments du premier et du deuxième ordre, l'existence d'une moyenne fixe  $\mu$  et d'une matrice d'autocovariance  $\mathbf{\Gamma}_n$  de la forme (1.7) pour tout  $n$  serait suffisant pour assurer la stationnarité d'un processus gaussien.

## Processus de bruit blanc

L'exemple le plus fondamental d'un processus stationnaire est une suite de variables aléatoires *indépendantes et identiquement distribuées*, notées  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_t, \dots$ , que nous supposons aussi de moyenne nulle et de variance  $\sigma_\varepsilon^2$ . Ce processus est strictement stationnaire et est appelé processus de *bruit blanc*. Parce que l'indépendance implique que les  $\varepsilon_t$  ne sont pas corrélés, sa fonction d'autocovariance est simplement

$$\gamma_k = \begin{cases} \sigma_\varepsilon^2 & \text{si } k = 0 \\ 0 & \text{si } k \neq 0 \end{cases}$$

Si l'on se concentre uniquement sur les propriétés du second ordre. Alors une suite de variables aléatoires  $\varepsilon_t$  non corrélées, ayant une moyenne nulle

et une variance commune  $\sigma_\varepsilon^2$ , a la même fonction d'autocovariance  $\gamma_k$  que ci-dessus, et est faiblement (de second ordre) stationnaire. Un tel processus peut également être considéré comme un processus de bruit blanc (au sens faible).

### Quelques opérateurs simples

Nous employons largement l'opérateur  $B$  (*backward shift operator*), qui est défini tel que  $BX_t = X_{t-1}$ ; par conséquent  $B^k X_t = X_{t-k}$ . L'opération inverse est effectuée à l'aide de l'opérateur  $F = B^{-1}$  (*forward shift operator*), défini tel que  $FX_t = X_{t+1}$ ; et par conséquent  $F^k X_t = X_{t+k}$ . Un autre opérateur important est l'opérateur de différence  $\nabla$  (*backward difference operator*), défini par  $\nabla X_t = X_t - X_{t-1}$ . Et que l'on peut exprimer en fonction de l'opérateur  $B$ , puisque

$$\nabla X_t = X_t - X_{t-1} = (1 - B) X_t$$

### Modèle de filtre linéaire

Les modèles stochastiques que nous employons sont basés sur une idée due à Yule (1927). Une série chronologique observable  $X_t$  dans laquelle les valeurs successives sont fortement dépendantes peut souvent être considérée comme générée à partir d'une série de "chocs" indépendants  $\varepsilon_t$ . Ces chocs sont des tirages aléatoires d'une distribution fixe, souvent supposés normaux et ayant une moyenne nulle et une variance  $\sigma_\varepsilon^2$ . Une telle suite de variables aléatoires indépendantes  $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots$  est appelée processus de bruit blanc. Le processus de bruit blanc  $\varepsilon_t$  est supposé se transformer en un processus  $X_t$  par ce qu'on appelle un *filtre linéaire*. L'opération de filtrage linéaire se résume simplement à une somme pondérée des chocs aléatoires présent et passés  $\varepsilon_t$ , de sorte que

$$\begin{aligned} X_t &= \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots \\ &= \mu + \psi(B) \varepsilon_t \end{aligned} \tag{1.8}$$

En général,  $\mu$  est un paramètre qui détermine le "niveau" du processus, et

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$$

est l'opérateur linéaire qui transforme  $\varepsilon_t$  en  $X_t$  et s'appelle la *fonction de transfert* du filtre.

La représentation du modèle (1.8) peut procurer une gamme flexible de modèles de dépendance entre les valeurs du processus  $\{X_t\}_{t \geq 1}$  exprimées en termes de chocs aléatoires indépendants (inobservables)  $\varepsilon_t$ .

La suite  $\psi_1, \psi_2, \dots$ , formée par les poids peut, théoriquement, être finie ou infinie. Si cette séquence est finie, ou infinie et absolument sommable dans le sens où  $\sum_{j=0}^{\infty} |\psi_j| < \infty$ , le filtre est dit être stable et le processus  $X_t$  est stationnaire. Le paramètre  $\mu$  est alors la moyenne autour de laquelle le processus varie. Sinon,  $X_t$  est non stationnaire et  $\mu$  n'a pas de signification spécifique, sauf comme point de référence pour le niveau du processus.

### 1.3.2 Modèles autorégressifs

Un modèle stochastique qui peut être extrêmement utile dans la représentation de certaines séries chronologiques est le modèle *autorégressif*. Dans ce modèle, la valeur présente du processus est exprimée comme une combinaison linéaire finie des *valeurs passées du processus* et d'un bruit blanc  $\varepsilon_t$ . Notons les valeurs d'un processus observées à intervalles réguliers et aux instants  $t, t-1, t-2, \dots$  par  $X_t, X_{t-1}, X_{t-2}, \dots$ , etc. Supposons que le processus  $X_t$  soit centé. Alors

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t \quad (1.9)$$

est appelé un *processus autorégressif* (AR) d'ordre  $p$ . La raison de ce nom est qu'un modèle linéaire en général

$$x = \phi_1 y_1 + \phi_2 y_2 + \dots + \phi_p y_p + \varepsilon$$

reliant une variable "dépendante"  $x$  à un ensemble de variables "indépendantes"  $y_1, y_2, \dots, y_p$ , plus un terme d'erreur aléatoire  $\varepsilon$ , est appelé modèle de *régression*, et  $x$  est dit "régressé" sur  $y_1, y_2, \dots, y_p$ . Dans (1.9) la variable  $X_t$  est régressée sur ses propres valeurs passées. Le modèle est donc *autorégressif*. Si nous définissons un *opérateur autorégressif* d'ordre  $p$  en fonction de l'opérateur de recul  $B$  par

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

le modèle autorégressif (1.9) peut être écrit succinctement tel que

$$\phi(B) X_t = \varepsilon_t$$

Le modèle contient  $p+2$  paramètres inconnus  $\mu, \phi_1, \phi_2, \dots, \phi_p, \sigma_\varepsilon^2$ , qui dans la pratique doivent être estimés à partir des données. Le paramètre additionnel



$\sigma_\varepsilon^2$  est la variance du processus de bruit blanc  $\varepsilon_t$ .

Il nest facile de voir que le modèle autorégressif est un cas particulier du modèle de filtre linéaire de (1.8). En effet, nous pouvons éliminer  $X_{t-1}$  du côté droit de l'équation (1.9) en substituant

$$X_{t-1} = \phi_1 X_{t-2} + \phi_2 X_{t-3} + \dots + \phi_p X_{t-p-1} + \varepsilon_{t-1}$$

De même, nous pouvons substituer  $X_{t-2}$ , et ainsi de suite, pour obtenir finalement une série infinie en fonction de  $\varepsilon$ 's. Considérons, en particulier, le processus autorégressif du premier ordre ( $AR(1)$ ),  $X_t = \phi X_{t-1} + \varepsilon_t$ . Aprs  $k$  substitutions successives de  $X_{t-j} = \phi X_{t-j-1} + \varepsilon_{t-j}$ ,  $j = 1, 2, \dots, k$  dans la partie droite, nous obtenons

$$X_t = \phi^{k+1} X_{t-k-1} + \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \dots + \phi^k \varepsilon_{t-k}$$

A la limite, quand  $k \rightarrow \infty$  ceci nous conduit à la représentation en série convergente  $X_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$  avec  $\psi_j = \phi^j$ ,  $j \geq 1$ , à condition que  $|\phi| < 1$ . Symboliquement, dans le cas d'un processus autorégressif général, nous avons

$$\phi(B) X_t = \varepsilon_t$$

ce qui est équivalent à

$$X_t = \phi^{-1}(B) \varepsilon_t = \psi(B) \varepsilon_t$$

avec  $\psi(B) = \phi^{-1}(B) = \sum_{j=0}^{\infty} \psi_j B^j$ .

Les processus autorégressifs peuvent être stationnaires ou non stationnaires. Pour que le processus soit stationnaire, les coefficients  $\phi$  doivent être tels que les poids  $\psi_1, \psi_2, \dots$  dans  $\psi(B) = \phi^{-1}(B)$  forment une série convergente. La condition nécessaire pour la stationnarité est que l'opérateur autorégressif,  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ , considéré comme un polynôme d'indéterminée  $B$  de degré  $p$ , doit avoir toutes ses racines supérieures à 1 en valeur absolue; c'est-à-dire que toutes les racines doivent se trouver à l'extérieur du cercle unité. Pour le processus  $AR$  du premier ordre,  $X_t = \phi X_{t-1} + \varepsilon_t$  cette exigence se réduit à la condition  $|\phi| < 1$ .

### 1.3.3 Modèles moyennes mobiles

Le modèle autorégressif (1.9) exprime la valeur présente  $X_t$  du processus en tant que somme pondérée *fini* de  $p$  valeurs antérieurs  $X_{t-1}, X_{t-2}, \dots, X_{t-p}$

du processus, plus un bruit blanc  $\varepsilon_t$ . De manière équivalente, comme nous venons de le voir, elle exprime  $X_t$  comme une somme pondérée *infinie* des  $\varepsilon$ . Un autre type de modèle, d'une grande importance pratique dans la représentation des séries chronologiques, est le processus fini *moyenne mobile*. Dans ce cas, nous prenons  $X_t$ , comme combinaison linéaire d'un nombre fini  $q$  de valeurs du processus bruit blanc  $\varepsilon$ . Ainsi,

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (1.10)$$

est appelé processus *moyenne mobile* (*MA*) d'ordre  $q$ . Le nom "moyenne mobile" est quelque peu trompeur parce que les poids du processus bruit blanc  $\varepsilon$ ,  $1, -\theta_1, -\theta_2, \dots, -\theta_q$ , n'ont pas besoin d'être ni positifs ni de somme égale à 1. Cependant, cette dénomination est couramment utilisée et nous l'employons donc. Si nous définissons un *opérateur moyennes mobiles* d'ordre  $q$  par

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

le modèle de la moyenne mobile peut être écrit de façon succincte

$$X_t = \theta(B) \varepsilon_t$$

Il contient  $q + 2$  paramètres inconnus  $\mu, \theta_1, \theta_2, \dots, \theta_q, \sigma_\varepsilon^2$ , qui en pratique doivent être estimés à partir des données.

### 1.3.4 Modèles mixtes autorégressif-moyennes mobiles

Pour obtenir une plus grande flexibilité dans l'ajustement des séries chronologiques, il est parfois avantageux d'inclure à la fois des termes autorégressifs et des termes moyennes mobiles dans le modèle. Cela conduit à un modèle autorégressif-mobile moyen (*ARMA*) :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (1.11)$$

ou bien

$$\phi(B) X_t = \theta(B) \varepsilon_t$$

Le modèle utilise  $p + q + 2$  paramètres inconnus  $\mu, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma_\varepsilon^2$ , estimés à partir des données. Ce modèle peut aussi être écrit sous la forme du filtre linéaire (1.8) comme  $X_t = \phi^{-1}(B) \theta(B) \varepsilon_t$ , avec  $\psi(B) = \phi^{-1}(B) \theta(B)$ . En pratique, il est souvent vrai qu'une représentation adéquate des séries

chronologiques stationnaires réelles peut être obtenue avec des modèles autorégressifs, en moyenne mobile ou mixtes, dans lesquels  $p$  et  $q$  ne dépassent pas 2. Nous discutons des classes de modèles autorégressifs, de moyennes mobiles et mixtes de façon beaucoup plus détaillée dans la suite.

## 1.4 Modèles nonstationnaires

Beaucoup de séries rencontrées dans l'industrie, l'économie ou dans les affaires (par exemple, les cours des actions et les chiffres des ventes) présentent un comportement non stationnaire et, en particulier, ne varient pas autour d'une moyenne fixe. De telles séries peuvent néanmoins présenter un comportement homogène dans le temps. En particulier, bien que le niveau général de fluctuation puisse être différent à différents instants, le comportement général de la série peut être similaire dans le temps. Nous montrons dans la suite que ce genre de comportements peuvent souvent être représenté par un modèle fonction d'un opérateur autorégressif généralisé  $\varphi(B)$ , dans lequel un ou plusieurs des zéros du polynôme  $\varphi(B)$  [c'est-à-dire qu'une ou plusieurs des racines de l'équation  $\varphi(B) = 0$ ] se trouvent sur le cercle unité. En particulier, s'il y a  $d$  racines unité et que toutes les autres racines se trouvent en dehors du cercle unité, l'opérateur  $\varphi(B)$  peut être écrit tel que

$$\varphi(B) = \phi(B)(1 - B)^d$$

où  $\phi(B)$  est un opérateur autorégressif stationnaire.

Ainsi, un modèle qui peut représenter un comportement non stationnaire homogène est de la forme

$$\varphi(B) X_t = \phi(B)(1 - B)^d X_t = \theta(B) \varepsilon_t$$

ou bien,

$$\phi(B) W_t = \theta(B) \varepsilon_t \tag{1.12}$$

où

$$W_t = (1 - B)^d X_t = \nabla^d X_t \tag{1.13}$$

Ainsi, un comportement non stationnaire homogène peut parfois être représenté par un modèle qui fait que la  $d^{\text{ème}}$  différence du processus soit stationnaire. En pratique,  $d$  est généralement 0, 1, ou au plus 2, avec  $d = 0$  correspondant au comportement stationnaire. Le processus défini par (1.12) et (1.13) fournit un modèle robuste pour décrire des séries chronologiques stationnaires et

non stationnaires et est appelé processeur  $ARIMA(p, d, q)$  ou *autoregressive integrated moving average process* d'ordre  $(p, d, q)$ . Il est défini tel que

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (1.14)$$

with  $W_t = \nabla^d X_t$ .

La raison de l'utilisation du mot "intégré" (qui devrait peut-être être "sommé") dans le titre  $ARIMA$  est la suivante. La relation inverse de (1.13) est  $X_t = S^d W_t$ , où  $S = \nabla^{-1} = (1 - B)^{-1} = 1 + B + B^2 \dots$ , est le *opérateur de sommation* défini par

$$SW_t = \sum_{j=0}^{\infty} W_{t-j} = W_t + W_{t-1} + W_{t-2} + \dots$$

Ainsi, le processus général  $ARIMA$  peut être généré en additionnant ou en "intégrant" le processus  $ARMA$  stationnaire  $W_t$   $d$  fois. Un peu plus loin, nous décrirons comment une forme spéciale du modèle (1.14) peut être utilisée pour représenter des séries chronologiques saisonnières.

On introduira aussi les modèles conditionnellement hétéroscédastiques tels que les modèles  $ARCH$  et  $GARCH$ . Ces modèles supposent que la variance conditionnelle d'une observation compte tenu de son passé varie avec le temps et qu'elle est utile pour modéliser la volatilité variant dans le temps dans les séries chronologiques économiques et financières, en particulier.

## Chapitre 2

# Fonction d'autocorrélation et spectre des processus stationnaires

Une caractéristique centrale du développement de modèles pour les séries chronologiques est l'hypothèse d'une certaine forme d'équilibre statistique. Une hypothèse particulièrement utile de ce genre est celle de la stationnarité. Habituellement, un processus stochastique stationnaire peut être utilement décrit par sa moyenne, sa variance et sa fonction d'autocorrélation ou, de manière équivalente, par sa moyenne, sa variance et sa densité spectrale. Dans ce chapitre, nous considérons les propriétés de ces fonctions et, en particulier, les propriétés de la fonction d'autocorrélation, qui seront largement utilisées dans le développement de modèles pour les séries chronologiques réelles.

### 2.1 Propriétés des autocorrélations des processus stationnaires

#### 2.1.1 Fonctions d'autocovariance et d'autocorrélation

On a vu plus haut que le coefficient d'autocovariance à horizon  $k$ ,  $\gamma_k$ , mesure la covariance entre deux valeurs  $X_t$  et  $X_{t+k}$  espacées de  $k$  unités de temps. La quantité  $\gamma_k$  considérée comme fonction du décalage  $k$  est appelée *fonction d'autocovariance* du processus stochastique, et est notée  $\{\gamma_k\}$ . De même, le coefficient d'autocorrélation  $\rho_k$  considéré comme fonction du

décalage  $k$  est appelé *fonction d'autocorrélation* du processus, et est notée  $\{\rho_k\}$ . Notez que la fonction d'autocorrélation est sans dimension, c'est-à-dire indépendante de l'échelle de mesure de la série chronologique. Du fait que  $\gamma_k = \sigma_X^2 \rho_k$ , la connaissance de la fonction d'autocorrélation  $\{\rho_k\}$  et de la variance  $\sigma_X^2$  est équivalent à la connaissance de la fonction d'autocovariance  $\{\gamma_k\}$ . Puisque  $\rho_k = \rho_{-k}$ , la fonction d'autocorrélation est nécessairement symétrique par rapport à zéro. Ainsi, en pratique il suffit de tracer la moitié positive de cette fonction. Le graphe de la fonction d'autocorrélation est parfois appelée le *corrélogramme*.

D'après ce qui a été mentionné plus haut, un processus stationnaire normal  $X_t$  est complètement caractérisé par sa moyenne  $\mu$  et sa fonction d'autocovariance  $\{\gamma_k\}$ , ou de manière équivalente par sa moyenne  $\mu$ , la variance  $\sigma_X^2$ , et la fonction d'autocorrélation  $\{\rho_k\}$ .

### Estimation des fonctions d'autocovariance et d'autocorrélation

Jusqu'à présent, nous n'avons considéré que la fonction d'autocorrélation théorique pour décrire un processus stochastique. En pratique, nous disposons d'une série chronologique  $X_1, X_2, \dots, X_N$  qui est un échantillon de taille  $N$  du processus observé. A partir de cet échantillon nous pouvons obtenir les estimations de la moyenne et des autocorrélations. La moyenne  $\mu = E(X_t)$  est estimée par la moyenne de l'échantillon  $\bar{X} = \frac{1}{N} \sum_{t=1}^N X_t$ .

Il est facile de constater que  $E(\bar{X}) = \mu$ , i.e. que  $\bar{X}$  est un estimateur sans biais de  $\mu$ . Pour mesurer la précision de l'estimateur  $\bar{X}$ , il suffit alors de calculer sa variance

$$Var(\bar{X}) = \frac{1}{N^2} \sum_{t=1}^N \sum_{s=1}^N \gamma_{t-s} = \frac{\gamma_0}{N} \left[ 1 + 2 \sum_{k=1}^{N-1} \left( 1 - \frac{k}{N} \right) \rho_k \right]$$

Une approximation sur grand échantillon pour la variance est donnée par

$$Var(\bar{X}) = \frac{1}{N^2} \sum_{t=1}^N \sum_{s=1}^N \gamma_{t-s} \approx \frac{\gamma_0}{N} \left( 1 + 2 \sum_{k=1}^{\infty} \rho_k \right)$$

De telle sorte que

$$NVar(\bar{X}) \xrightarrow{N \rightarrow \infty} \gamma_0 \left( 1 + 2 \sum_{k=1}^{\infty} \rho_k \right)$$

en admettant que  $\sum_{k=-\infty}^{+\infty} |\rho_k| < \infty$ .

Notons que le premier facteur dans  $Var(\bar{X})$ ,  $\frac{\gamma_0}{N}$ , est l'expression familière de la variance de  $\bar{X}$  obtenu à partir d'un échantillon aléatoire *iid* de taille  $N$ . Mais l'existence d'autocorrélation entre les valeurs  $X_t$  peut affecter sensiblement la précision de  $\bar{X}$ . Par exemple, dans le cas où un processus stationnaire possède des autocorrélations  $\rho_k = \phi^{|k|} < 1$ , l'approximation sur grand échantillon pour la variance de  $\bar{X}$  devient  $Var(\bar{X}) = \frac{\gamma_0}{N} \frac{1+\phi}{1-\phi}$ , et le deuxième facteur  $\frac{1+\phi}{1-\phi}$  peut évidemment différer sensiblement de 1.

Un certain nombre d'estimateurs de la fonction d'autocorrélation ont été suggérées dans la littérature, et leurs propriétés sont discutées par Jenkins et Watts (1968), entre autres. On peut cependant affirmer que l'estimation la plus satisfaisante de l'autocorrélation  $\rho_k$  est

$$r_k = \hat{\rho}_k = \frac{c_k}{c_0} \quad (2.1)$$

où

$$c_k = \hat{\gamma}_k = \frac{1}{N} \sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) \quad k = 0, 1, \dots, K \quad (2.2)$$

est l'estimation de l'autocovariance  $\gamma_k$  et  $\bar{X}$  est la moyenne de l'échantillon de la série chronologique.

Les valeurs  $r_k$  dans (2.1) sont appelées "*fonction d'autocorrélation empirique*". Pour obtenir une estimation utile de la fonction d'autocorrélation, en pratique, on a généralement besoin d'au moins 50 observations, et les autocorrélations estimées  $r_k$  seraient calculées pour  $k = 0, 1, \dots, K$ , où  $K$  n'est pas plus grand que  $\frac{N}{4}$  en général.

### Ecart-types des estimateurs des autocorrélations

Pour identifier un modèle pour une série chronologique, en utilisant les méthodes que l'on décrira plus loin, il est utile de disposer d'un test pour vérifier si les  $\rho_k$  sont effectivement nuls au-delà d'un certain horizon. Pour ce faire, on peut utiliser l'expression de la variance de  $r_k$  pour un processus gaussien stationnaire proposée par Bartlett (1946) :

$$Var(r_k) = \frac{1}{N} \sum_{\nu=-\infty}^{\infty} (\rho_\nu^2 + \rho_{\nu+k}\rho_{\nu-k} - 4\rho_k\rho_\nu\rho_{\nu-k} + 2\rho_k^2\rho_\nu^2) \quad (2.3)$$

Par exemple, si  $\rho_k = \phi^{|k|}$  ( $-1 < \phi < 1$ ), i.e. la fonction d'autocorrélation tend vers zéro exponentiellement, alors (2.3) donne

$$Var(r_k) \simeq \frac{1}{N} \left[ \frac{(1 + \phi^2)(1 - \phi^{2k})}{1 - \phi^2} - 2k\phi^{2k} \right] \quad (2.4)$$

et en particulier

$$Var(r_1) \simeq \frac{1}{N} (1 - \phi^2)$$

Pour tout processus pour lequel les autocorrélations  $\rho_\nu$  sont nulles pour  $\nu > q$ , tous les termes sauf le premier apparaissant du côté droit de (2.3) sont nuls quand  $k > q$ . Ainsi, la variance de l'estimateur  $r_k$  des autocorrélations à horizon  $k > q$  est donnée par l'approximation de Bartlett telle que

$$Var(r_k) \simeq \frac{1}{N} \left( 1 + 2 \sum_{\nu=1}^q \rho_\nu^2 \right) \quad k > q \quad (2.5)$$

Pour utiliser ce résultat en pratique, les autocorrélations estimées  $r_k$  ( $k = 1, 2, \dots, q$ ) sont substituées aux autocorrélations théoriques  $\rho_k$ . La racine carrée de (2.5) est considérée comme l'écart-type à horizon lointain.

On montrera plus loin que le processus en moyenne mobile (MA) a une structure de corrélation telle que l'approximation (2.5) s'applique à ce processus. Des expressions similaires pour la covariance approchée entre les autocorrélations estimées  $r_k$  et  $r_{k+s}$  à deux horizons différents  $k$  et  $k+s$  ont également été données par Bartlett (1946). En particulier, l'approximation à horizon lointain se réduit à

$$Cov(r_k, r_{k+s}) \simeq \frac{1}{N} \sum_{\nu=-q}^q \rho_\nu \rho_{\nu+s} \quad (2.6)$$

Ce résultat montre qu'il faut être prudent dans l'interprétation des autocorrélations individuelles, car des covariances assez grandes peuvent exister entre les valeurs voisines. Cet effet peut parfois déformer l'aspect visuel de la fonction d'autocorrélation de l'échantillon, qui peut ne pas chuter à zéro rapidement.

Un cas d'intérêt particulier se produit pour  $q = 0$ , c'est-à-dire quand les  $\rho_k$  sont supposés nuls pour tous les horizons autres que l'horizon 0, et donc la série chronologique est complètement aléatoire, i.e. un bruit blanc. Alors, les



écart-types de (2.5) pour les autocorrélations estimées  $r_k$  prennent la forme simple suivante

$$SE(r_k) \simeq \frac{1}{\sqrt{N}} \quad k > 0$$

De plus, dans ce cas, le résultat de (2.6) indique que les autocorrélations estimées  $r_k$  et  $r_{k+s}$  à deux horizons différents ne sont pas corrélées. Et puisque les  $r_k$  sont généralement supposés être distribués approximativement normalement pour  $N$  grand, une suite d'autocorrélations estimées pour différents horizons aura tendance à être distribuée indépendamment et normalement avec une moyenne nulle et une variance  $\frac{1}{\sqrt{N}}$ .

### Périodogramme d'une série chronologique

Une autre façon d'analyser une série chronologique est basée sur l'hypothèse qu'elle est constituée d'ondes sinusoïdales de différentes fréquences. Un dispositif qui utilise cette idée, introduit par Schuster (1898), est le *périodogramme*. Ce dernier était à l'origine utilisé pour détecter et estimer l'amplitude d'une composante sinusoïdale, de fréquence connue, masquée par un bruit. Nous l'utiliserons plus tard pour vérifier le caractère aléatoire d'une série (généralement une série de résidus après ajustement d'un modèle particulier).

Pour illustrer le calcul du périodogramme, supposons que le nombre d'observations  $N = 2q + 1$  est impair. Considérons le développement en série de Fourier de la série chronologique

$$X_t = \alpha_0 + \sum_{i=0}^q (\alpha_i c_{it} + \beta_i s_{it}) + e_t \quad (2.7)$$

où  $c_{it} = \cos(2\pi f_i t)$ ,  $s_{it} = \sin(2\pi f_i t)$ .

La fréquence  $f_i = \frac{i}{N}$  est la  $i^{\text{ième}}$  harmonique de la fréquence fondamentale  $\frac{1}{N}$  associée à la  $i^{\text{ième}}$  composante sinusoïdale dans (2.7) de fréquence  $f_i$  et de période  $\frac{1}{f_i} = \frac{N}{i}$ .

Les estimateurs des moindres carrés des coefficients  $\alpha_0$  et  $(\alpha_i, \beta_i)$  sont

$$a_0 = \overline{X} \quad (2.8)$$

et pour  $i = 1, 2, \dots, q$

$$a_i = \frac{2}{N} \sum_{t=1}^N X_t c_{it} \quad (2.9)$$

$$b_i = \frac{2}{N} \sum_{t=1}^N X_t s_{it} \quad (2.10)$$

Puisque  $\sum_{t=1}^N c_{it}^2 = \sum_{t=1}^N s_{it}^2 = \frac{N}{2}$ , tous les termes dans (2.7) sont mutuellement orthogonaux pour  $t = 1, 2, \dots, N$ .

Le périodogramme consiste alors en  $q = \frac{N-1}{2}$  valeurs

$$I(f_i) = \frac{N}{2} (a_i^2 + b_i^2) \quad (2.11)$$

où  $I(f_i)$  est appelé *intensité* à la fréquence  $f_i$ .

Quand  $N$  est pair, on pose  $N = 2q$  et les relations (2.8)-(2.11) sont appliquées pour  $i = 1, 2, \dots, q-1$ , mais

$$a_q = \frac{1}{N} \sum_{t=1}^N (-1)^t X_t$$

$$b_q = 0$$

et

$$I(f_q) = I(0, 5) = N a_q^2$$

Notons que la fréquence la plus élevée est de 0,5 cycle par intervalle de temps parce que la plus petite période est de deux intervalles.

### 2.1.2 Spectre et densité spectrale

Pour compléter notre étude, nous ajoutons ici une brève discussion du spectre et de la fonction de densité spectrale. L'utilisation de ces outils importants est décrite plus en détail par Jenkins et Watts (1968), Bloomfield (2000) et Shumway et Stoffer (2011), entre autres. Nous ne les appliquons pas à l'analyse des séries chronologiques dans ce cours.

#### Spectre empirique

La définition du périodogramme dans (2.11) suppose que les fréquences  $f_i = \frac{i}{N}$  sont des harmoniques de la fréquence fondamentale  $\frac{1}{N}$ . En guise d'introduction au spectre, nous assouplissons cette hypothèse et permettons

à la fréquence  $f_i$  de varier continûment dans l'intervalle  $[0-0,5]$ . La définition (2.11) du périodogramme peut être modifiée telle que

$$I(f) = \frac{N}{2} (a_f^2 + b_f^2) \quad (2.12)$$

et  $I(f)$  est alors appelé *spectre empirique* (Jenkins et Watts, 1968). Comme le périodogramme, il peut être utilisé pour détecter et estimer l'amplitude d'une composante sinusoïdale de fréquence inconnue  $f$  cachée dans le bruit. En effet, il apparaît comme un outil plus approprié à cette fin si l'on sait que la fréquence  $f$  n'est pas en relation harmonique avec la longueur de la série. De plus, il fournit un point de départ pour la théorie de l'analyse spectrale. On peut montrer que le spectre empirique  $I(f)$  et l'estimateur  $c_k$  de la fonction d'autocovariance sont liés par l'importante relation

$$I(f) = 2 \left[ c_0 + 2 \sum_{k=1}^{N-1} c_k \cos(2\pi f k) \right] \quad 0 \leq f \leq \frac{1}{2} \quad (2.13)$$

*Autrement dit, le spectre empirique est la transformée de Fourier discrète de l'estimateur de la fonction d'autocovariance.*

## Spectre

Le périodogramme et le spectre empirique sont des outils appropriés pour analyser des séries chronologiques constituées de mélanges d'ondes sinusoïdales, à des fréquences fixes bruitées. Cependant, les séries chronologiques stationnaires du type décrit plus haut sont caractérisées par des changements aléatoires de fréquence, d'amplitude et de phase. Pour ce type de série, le spectre empirique  $I(f)$  fluctue énormément et ne peut pas faire l'objet d'interprétations utiles.

Cependant, supposons que le spectre empirique soit déterminé pour une série chronologique de taille  $N$ , qui est une réalisation d'un processus gaussien stationnaire. Comme déjà mentionné, un tel processus n'aurait pas de composantes déterministes en cosinus ou sinus. Mais, on peut mener une analyse de Fourier et obtenir des valeurs de  $(a_f, b_f)$  pour une fréquence donnée  $f$ . Si on répète la réalisations de  $N$  observations du processus stochastique, nous pourrions construire une population de valeurs pour  $a_f$ ,  $b_f$ , et  $I(f)$ . Ainsi, nous pourrions calculer la valeur moyenne de  $I(f)$  à l'aide des réalisations

répétées de taille  $N$ , à savoir,

$$E[I(f)] = 2 \left[ E(c_0) + 2 \sum_{k=1}^{N-1} E(c_k) \cos(2\pi f k) \right] \quad (2.14)$$

Pour  $N$  grand, on peut montrer (Jenkins et Watts, 1968) que la moyenne de l'estimateur  $c_k$  du coefficient d'autocovariance tend vers l'autocovariance théorique  $\gamma_k$ , c'est-à-dire

$$\lim_{N \rightarrow \infty} E(c_k) = \gamma_k$$

On définit la *puissance du spectre*  $p(f)$  telle que

$$p(f) = \lim_{N \rightarrow \infty} E[I(f)] = 2 \left[ \gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos(2\pi f k) \right] \quad (2.15)$$

Il est clair que

$$|p(f)| \leq 2 \left( |\gamma_0| + 2 \sum_{k=1}^{\infty} |\gamma_k| \right) \quad (2.16)$$

Une condition suffisante pour que le spectre converge est que  $\gamma_k$  tende vers zéro rapidement pour que la série (2.16) converge.

Puisque la puissance du spectre est la transformée de Fourier de la fonction d'autocovariance, la connaissance de la fonction d'autocovariance est mathématiquement équivalente à la connaissance du spectre, et vice versa. A partir de maintenant, nous nous référons à "puissance du spectre" simplement par "spectre".

En intégrant (2.15) entre les limites 0 et  $\frac{1}{2}$ , la variance du processus  $X_t$  est alors

$$\gamma_0 = \sigma_{X_t}^2 = \int_0^{\frac{1}{2}} p(f) df \quad (2.17)$$

Ainsi, de la même manière que le périodogramme  $I(f)$  montre comment la variance (2.3) d'une série, composée d'un mélange de sinus et de cosinus, est répartie entre les différentes fréquences harmoniques, le spectre  $p(f)$  montre comment la variance d'un processus stochastique est répartie entre une gamme continue de fréquences. On peut interpréter  $p(f) df$  comme mesurant approximativement la variance du processus dans la gamme de fréquences allant de  $f$  à  $f + df$ . De plus, à partir de la définition de (2.15),

la représentation spectrale de la fonction d'autocovariance  $\{\gamma_k\}$  peut être obtenue comme

$$\gamma_k = \int_0^{\frac{1}{2}} \cos(2\pi f k) p(f) df$$

qui, avec (2.15), montre directement la correspondance biunivoque entre la puissance du spectre et la fonction d'autocovariance d'un processus.

Inversement, puisque les  $\gamma_k$  forment une suite définie positive, à condition que la série (2.16) converge, il résulte du théorème de Herglotz (Loève, 1977) qu'une fonction unique  $p(f)$  existe telle que  $\gamma_k$  admet la représentation spectrale  $\gamma_k = \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi f k} p(f) df$ . Par conséquent, la puissance du spectre  $p(f)$  d'un processus stationnaire, pour lequel (2.16) converge, peut être définie comme une fonction unique, dont l'existence est garantie et doit avoir la forme du membre de droite de (2.15) par représentation spectrale.

### Fonction de densité spectrale

Il est parfois plus commode de baser la définition (2.15) du spectre sur les autocorrélations  $\rho_k$  plutôt que sur les autocovariances  $\gamma_k$ . La fonction résultante

$$g(f) = \frac{p(f)}{\sigma_X^2} = 2 \left[ 1 + 2 \sum_{k=1}^{\infty} \rho_k \cos(2\pi f k) \right] \quad 0 \leq f \leq \frac{1}{2} \quad (2.18)$$

est appelée "*fonction de densité spectrale*". En utilisant (2.17), on voit que la fonction de densité spectrale a la propriété suivante

$$\int_0^{\frac{1}{2}} g(f) df = 1$$

Puisque la fonction  $g(f)$  est aussi positive, elle a les mêmes propriétés qu'une fonction de densité de probabilité ordinaire. Cette analogie s'étend aux propriétés d'estimation de ces deux fonctions, comme nous le verrons plus loin.

### Estimation du spectre

On pourrait s'attendre à ce qu'une estimation du spectre puisse être obtenue à partir de (2.15), en remplaçant les autocovariances théoriques  $\gamma_k$  par leurs estimations  $c_k$ . En raison de (2.5), cela revient à prendre le spectre empirique comme estimateur de  $p(f)$ . Cependant, on peut montrer (Jenkins et

Watts, 1968) que le spectre empirique d'une série chronologique stationnaire fluctue fortement autour du spectre théorique. Une explication intuitive de ce fait est que le spectre de l'échantillon correspond à l'utilisation d'un intervalle, dans le domaine fréquentiel, dont la largeur est trop petite. Ceci est analogue à l'utilisation d'un intervalle de groupe trop petit pour l'histogramme lors de l'estimation d'une distribution de probabilité ordinaire. En utilisant une estimation modifiée ou *lissée*

$$\hat{p}(f) = 2 \left[ c_0 + 2 \sum_{k=1}^{N-1} \lambda_k c_k \cos(2\pi f k) \right] \quad (2.19)$$

où les  $\lambda_k$  sont des poids convenablement choisis, appelés *fenêtre de retard*, il est possible d'augmenter le *bande passante* de l'estimation et d'obtenir une estimation plus douce du spectre. Les poids  $\lambda_k$  dans (2.19) sont généralement choisis de sorte qu'ils chutent à zéro pour les horizons  $k > M$ , où  $M$  est appelé point de troncature et  $M < N$  est modérément petit par rapport à la longueur  $N$  de la série. Comme une forme alternative de calcul, on peut aussi obtenir une estimation d'un spectre plus lisse en considérant les estimateurs des  $p(f_i)$  tels que

$$\hat{p}(f_i) = \sum_{j=-m}^m W(f_i) I\left(f_i + \frac{j}{N}\right)$$

où  $\sum_{j=-m}^m W(f_i) = 1$ .

La fonction de pondération symétrique  $W(f_i)$  est appelée *fenêtre spectrale*, et  $m$  est choisi pour être beaucoup plus petit que  $\frac{N}{2}$ .

**Remarque 2** La commande "*spectrum()*" peut être utilisée pour estimer le spectre dans *R*. Pour utiliser cette commande, une *fenêtre de lissage* doit être spécifiée ; voir l'aide (*spectrum*).

## Avantages et inconvénients des fonctions d'autocorrélation et de densité spectrale

Parce que la fonction d'autocorrélation et le spectre sont des transformées l'une de l'autre, elles sont mathématiquement équivalentes, et donc toute discussion de leurs avantages et inconvénients ne tourne pas sur des questions mathématiques mais sur leur représentativité du processus. Parce que,

comme nous l'avons vu, chacun fait la lumière sur un aspect différent des données, ils doivent être considérés comme complémentaires et non contradictoires. Chacun contribue à la compréhension du processus stochastique en question.

L'obtention des estimations de la fonction d'autocorrélation et du spectre sont des approches non structurelles, analogues à la représentation d'une fonction de distribution empirique par un histogramme. Ce sont deux des moyens qui font parler les données des séries stationnaires et fournir une première étape dans l'analyse des séries chronologiques, tout comme un histogramme peut fournir une première étape dans l'analyse de la distribution des données.

Le choix entre le spectre et la fonction d'autocorrélation en tant qu'outil de modélisation dépend de la nature des modèles qui s'avèrent utiles dans la pratique. Les modèles que nous avons trouvés intéressants, et que nous considérons dans la suite de ce cours, sont simplement décrits en termes de fonction d'autocorrélation, et c'est cet outil que nous utiliserons pour la spécification ou l'identification d'un modèle.

### 2.1.3 Lien entre le spectre empirique et l'estimateur de la fonction d'autocovariance

Dans ce paragraphe, nous allons montrer le résultat (2.5) :

$$I(f) = 2 \left[ c_0 + 2 \sum_{k=1}^{N-1} c_k \cos(2\pi f k) \right] \quad 0 \leq f \leq \frac{1}{2}$$

qui relie le spectre empirique  $I(f)$  et l'estimateur  $c_k$  de la fonction d'autocovariance.

Supposons que les estimateurs des moindres carrés  $a_f$  et  $b_f$  des composantes sinusoïdales, à la fréquence  $f$ , d'une série soient combinés tel que  $d_f = a_f - ib_f$ , où  $i = \sqrt{-1}$ ; alors

$$I(f) = \frac{N}{2} (a_f - ib_f)(a_f + ib_f) = \frac{N}{2} d_f d_f^* \quad (2.20)$$

où  $d_f^*$  est le conjugué complexe de  $d_f$ . Ensuite, en utilisant (2.1) et (2.2), nous obtenons

$$\begin{aligned} d_f &= \frac{2}{N} \sum_{t=1}^N X_t [\cos(2\pi ft) - i \sin(2\pi ft)] \\ &= \frac{2}{N} \sum_{t=1}^N X_t e^{-i2\pi ft} = \frac{2}{N} \sum_{t=1}^N (X_t - \bar{X}) e^{-i2\pi ft} \end{aligned} \quad (2.21)$$

Le remplacement de (2.21) dans (2.20) donne

$$I(f) = \frac{N}{2} \sum_{t=1}^N \sum_{t'=1}^N (X_t - \bar{X}) (X_{t'} - \bar{X}) e^{-i2\pi f(t-t')} \quad (2.22)$$

Puisque

$$c_k = \frac{1}{N} \sum_{t=1}^{N-k} (X_t - \bar{X}) (X_{t+k} - \bar{X})$$

Le changement de variable  $k = t - t'$  dans (2.22) donne le résultat requis :

$$\begin{aligned} I(f) &= 2 \sum_{k=-N+1}^{N-1} c_k e^{-i2\pi f k} \\ &= 2 \left[ c_0 + 2 \sum_{k=1}^{N-1} c_k \cos(2\pi f k) \right] \quad 0 \leq f \leq \frac{1}{2} \end{aligned}$$



# Chapitre 3

## Modèles stationnaires linéaires

Dans ce chapitre, nous allons décrire un modèle stochastique linéaire général qui suppose qu'un processus aléatoire est généré par une combinaison linéaire de bruits blancs. Pour une représentation pratique, il est souhaitable de construire des modèles qui utilisent des paramètres avec parcimonie. Cette parcimonie peut souvent être obtenue par une représentation du processus linéaire en termes d'un petit nombre de termes autorégressifs-moyenne mobile (ARMA). Les propriétés des modèles ARMA qui en résultent sont discutées en préparation de leur utilisation dans la construction de modèles pour les séries chronologiques.

### 3.1 Processus linéaire général

#### 3.1.1 Deux formes équivalentes pour un processus linéaire

Nous avons déjà discuté de la représentation d'un processus stochastique centré  $X_t$  comme la sortie d'un filtre linéaire, dont l'entrée est le bruit blanc  $\varepsilon_t$ , c'est-à-dire

$$\begin{aligned} X_t &= \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots \\ &= \varepsilon_t + \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j} \end{aligned} \tag{3.1}$$

Le *processus linéaire général* (3.1) nous permet de représenter  $X_t$  comme une somme pondérée des valeurs passées et présente du processus "bruit blanc"

$\varepsilon_t$ . La forme de représentation (3.1) est dite représentation en  $\psi$ -poids du processus linéaire.

Les premières références importantes sur le développement des modèles stochastiques linéaires incluent Yule (1927), Walker (1931), Slutsky (1937), Wold (1938), Kendall (1945), Bartlett (1946), Quenouille (1952,1957), Doob (1953), Grenander et Rosenblatt (1957), Hannan (1960), Robinson (1967), entre autres. L'utilité de ces modèles est bien mise en évidence dans la littérature ultérieure. Le processus de bruit blanc  $\varepsilon_t$  consiste en une suite de variables aléatoires non corrélées de moyenne nulle et variance constante, c'est-à-dire

$$E(\varepsilon_t) = 0 \quad Var(\varepsilon_t) = \sigma_\varepsilon^2$$

Puisque les variables aléatoires  $\varepsilon_t$  sont supposées non corrélées, il s'ensuit que leur fonction d'autocovariance est

$$\gamma_k = E(\varepsilon_t \varepsilon_{t+k}) = \begin{cases} \sigma_\varepsilon^2, & k = 0 \\ 0, & k \neq 0 \end{cases} \quad (3.2)$$

Ainsi, la fonction d'autocorrélation du bruit blanc a une forme particulièrement simple

$$\rho_k = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases} \quad (3.3)$$

Un résultat fondamental dans le développement des processus stationnaires est celui de Wold (1938), qui a établi que tout processus stationnaire centré purement non-déterministe  $X_t$  possède une représentation linéaire du type (3.1) avec  $\sum_{j=1}^{\infty} \psi_j^2 < \infty$ . Les  $\varepsilon_t$  ne sont pas corrélés et possèdent une variance commune  $\sigma_\varepsilon^2$  mais ne sont pas nécessairement indépendants. Nous réserverons le terme processus linéaire pour un processus  $X_t$  de la forme de (3.1) dans lequel les variables  $\varepsilon_t$  sont des variables aléatoires indépendantes.

Pour qu'un processus  $X_t$  défini par (3.1) représente un processus stationnaire valable, il est nécessaire que les coefficients  $\psi_j$  soient absolument sommables, c'est-à-dire  $\sum_{j=1}^{\infty} |\psi_j| < \infty$ . Sous certaines conditions (Koopmans, 1974),  $X_t$  peut être exprimé aussi comme une somme pondérée des valeurs de  $X_t$  passées et une innovation  $\varepsilon_t$ , c'est-à-dire

$$\begin{aligned} X_t &= \pi_1 X_{t-1} + \pi_2 X_{t-2} + \dots + \varepsilon_t \\ &= \sum_{j=1}^{\infty} \pi_j X_{t-j} + \varepsilon_t \end{aligned} \quad (3.4)$$

Dans cette forme alternative, la valeur présente  $X_t$  peut être considérée comme étant "régressée" sur les valeurs passées  $X_{t-1}, X_{t-2}, \dots$  du processus. La forme de représentation alternative (3.4) est dite représentation en  $\pi$ -poids du processus linéaire.

### Relations entre les poids $\psi$ et les poids $\pi$

Les relations entre les poids  $\psi$  et les poids  $\pi$  peuvent être obtenues en utilisant l'opérateur de retard "backward shift"  $B$  précédemment défini, tel que

$$BX_t = X_{t-1} \text{ et donc } B^j X_t = X_{t-j}$$

Plus tard, nous aurons aussi besoin d'utiliser l'opérateur d'avance "forward shift"  $F = B^{-1}$ , défini tel que

$$FX_t = X_{t+1} \text{ et donc } F^j X_t = X_{t+j}$$

A titre d'exemple de l'utilisation de l'opérateur  $B$ , considérons le modèle suivant

$$X_t = \varepsilon_t - \theta \varepsilon_{t-1} = (1 - B) \varepsilon_t$$

dans lequel  $\psi_1 = -\theta$ ,  $\psi_j = 0$  pour  $j > 1$ .

En exprimant  $\varepsilon_t$  en termes de  $X_t$ , on obtient

$$(1 - B)^{-1} X_t = \varepsilon_t$$

De plus, pour  $|\theta| < 1$ ,

$$(1 - B)^{-1} = 1 + \theta B + \theta^2 B^2 + \theta^3 B^3 + \dots$$

Et par conséquent,  $X_t$  s'exprime en termes de valeurs antérieures, comme dans (3.4), tel que

$$X_t = -\theta X_{t-1} - \theta^2 X_{t-2} + \theta^3 X_{t-3} + \dots + \varepsilon_t$$

de sorte que pour ce modèle,  $\pi_j = -\theta^j$ .

En utilisant l'opérateur "backshift"  $B$ , le modèle (3.1) peut être écrit tel que

$$X_t = \left( 1 + \sum_{j=1}^{\infty} \psi_j B^j \right) \varepsilon_t$$

ou bien

$$X_t = \psi(B) \varepsilon_t \quad (3.5)$$

où

$$\psi(B) = \left(1 + \sum_{j=1}^{\infty} \psi_j B^j\right) = \sum_{j=0}^{\infty} \psi_j B^j$$

avec  $\psi_0 = 1$ .

La fonction  $\psi(B)$  est appelée *fonction de transfert* du filtre linéaire qui relie  $X_t$  à  $\varepsilon_t$ . Elle peut être considérée comme la fonction génératrice des poids  $\psi$ , avec  $B$  maintenant considéré simplement comme une variable dont la  $k^{\text{ième}}$  puissance est le coefficient de  $\psi_j$ .

De même, (3.4) peut être écrite telle que

$$\left(1 - \sum_{j=1}^{\infty} \pi_j B^j\right) X_t = \varepsilon_t$$

ou bien

$$\pi(B) X_t = \varepsilon_t \quad (3.6)$$

Ainsi,

$$\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j$$

est la fonction génératrice des poids  $\pi$ .

Après avoir multiplié des deux côtés de l'expression (3.6) par  $\psi(B)$ , on obtient

$$\psi(B) \pi(B) X_t = \psi(B) \varepsilon_t = X_t$$

D'où,  $\psi(B) \pi(B) = 1$ , de sorte que

$$\pi(B) = \psi^{-1}(B) \quad (3.7)$$

Cette relation peut être utilisée pour obtenir les poids  $\pi$  connaissant les poids  $\psi$ , et vice versa.

### 3.1.2 Fonction génératrice des autocovariances d'un processus linéaire

Un outil de base pour l'analyse des séries chronologiques afin d'identifier les modèles adéquats, est la fonction d'autocorrélation. Ainsi, il est très utile de bien connaître la fonction d'autocorrélation d'un processus linéaire.

## Autocovariances

L'autocovariance à horizon  $k$  d'un processus linéaire  $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$  avec  $\psi_0 = 1$  est évidemment telle que

$$\begin{aligned} \gamma_k &= E(X_t X_{t-k}) = E\left(\sum_{j=0}^{\infty} \sum_{h=0}^{\infty} \psi_j \psi_h \varepsilon_{t-j} \varepsilon_{t+k-h}\right) \\ &= \sum_{j=0}^{\infty} \sum_{h=0}^{\infty} \psi_j \psi_h E(\varepsilon_{t-j} \varepsilon_{t+k-h}) = \sigma_{\varepsilon}^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k} \end{aligned} \quad (3.8)$$

en utilisant la propriété (3.2) pour la fonction d'autocovariance du bruit blanc. En particulier, en posant  $k = 0$ , on déduit la variance du processus  $X_t$

$$\gamma_0 = \sigma_X^2 = \sigma_{\varepsilon}^2 \sum_{j=0}^{\infty} \psi_j^2 \quad (3.9)$$

Il s'ensuit que la condition de l'absolue sommabilité de la suite des coefficients  $\psi_j$  d'un processus linéaire stationnaire,  $\sum_{j=0}^{\infty} |\psi_j| < \infty$ , implique que la série du côté droit de équation (3.9) converge, et garantit ainsi au processus une variance finie.

## Fonction génératrice des autocovariances

Une autre façon d'obtenir les autocovariances d'un processus linéaire est via la *fonction génératrice des autocovariances* définie telle que

$$\gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_k B^k \quad (3.10)$$

où  $\gamma_0$ , la variance du processus, est le coefficient de  $B^0 = 1$ , alors que  $\gamma_k$ , l'autocovariance à horizon  $k$ , est le coefficient des deux termes  $B^k$  et  $B^{-k} = F^k$ . En substituant le résultat de (3.8) dans la fonction génératrice des autocovariances (3.10) on obtient

$$\begin{aligned} \gamma(B) &= \sigma_{\varepsilon}^2 \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} \psi_j \psi_{j+k} B^k \\ &= \sigma_{\varepsilon}^2 \sum_{j=0}^{\infty} \sum_{k=-j}^{\infty} \psi_j \psi_{j+k} B^k \end{aligned}$$

puisque  $\psi_h = 0$  pour  $h < 0$ .

Posons  $j + k = h$ , si bien que  $k = h - j$ , on aura

$$\begin{aligned}\gamma(B) &= \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \sum_{h=0}^{\infty} \psi_j \psi_h B^{h-j} \\ &= \sigma_\varepsilon^2 \sum_{h=0}^{\infty} \psi_h B^h \sum_{j=0}^{\infty} \psi_j B^{-j}\end{aligned}$$

Finalement, la fonction génératrice des autocovariances s'écrit telle que

$$\gamma(B) = \sigma_\varepsilon^2 \psi(B) \psi(B^{-1}) = \sigma_\varepsilon^2 \psi(B) \psi(F) \quad (3.11)$$

Par exemple, supposons que  $X_t = \varepsilon_t - \theta \varepsilon_{t-1} = (1 - \theta B) \varepsilon_t$  de sorte que  $\psi(B) = 1 - \theta B$ . Alors,

$$\begin{aligned}\gamma(B) &= \sigma_\varepsilon^2 (1 - \theta B) (1 - \theta B^{-1}) \\ &= \sigma_\varepsilon^2 [-\theta B^{-1} + (1 + \theta^2) B^0 - \theta B]\end{aligned}$$

En comparant avec (3.10), on déduit les autocovariances par identification

$$\begin{aligned}\gamma_0 &= \sigma_\varepsilon^2 (1 + \theta^2) \\ \gamma_1 &= -\theta \sigma_\varepsilon^2 \\ \gamma_k &= 0 \quad k \geq 2\end{aligned}$$

Dans la suite,  $B$  sera considéré comme une variable dans la fonction génératrice, et ainsi il pourra prendre des valeurs complexes. En particulier, il sera souvent nécessaire de considérer les différents cas où  $|B| < 1$ ,  $|B| = 0$ , ou  $|B| > 1$ , c'est-à-dire lorsque le nombre complexe  $B$  se trouve l'intérieur, sur ou en dehors du cercle unité.

### 3.1.3 Conditions de stationnarité et d'inversibilité pour un processus linéaire

#### Stationnarité

La convergence de la série (3.9) garantit que le processus possède une variance finie. De plus, nous avons vu plus haut que les autocovariances et les autocorrélations doivent satisfaire à certaines conditions pour assurer la

stationnarité du processus. Pour le processus linéaire (3.1), ces conditions sont résumées par la seule condition que  $\sum_{j=0}^{\infty} |\psi_j| < \infty$ . Cette condition peut également être incorporée dans la condition que la série  $\psi(B)$ , qui est la fonction génératrice des poids  $\psi$ , doit converger pour  $|B| < 1$ , c'est-à-dire sur ou dans le cercle de l'unité.

#### Conditions de stationnarité.

Si nous remplaçons  $B = e^{-i2\pi f}$  et  $F = B^{-1} = e^{i2\pi f}$  dans la fonction génératrice des autocovariances (3.11), nous obtenons la moitié de la puissance du spectre. Par conséquent, la puissance du spectre d'un processus linéaire est alors telle que

$$\begin{aligned} p(f) &= 2\sigma_\varepsilon^2 \psi(e^{-i2\pi f}) \psi(e^{i2\pi f}) \\ &= 2\sigma_\varepsilon^2 |\psi(e^{-i2\pi f})|^2 \quad 0 \leq f \leq \frac{1}{2} \end{aligned} \quad (3.12)$$

Il s'ensuit que la variance du processus est

$$\sigma_X^2 = \int_0^{\frac{1}{2}} p(f) df = 2\sigma_\varepsilon^2 \int_0^{\frac{1}{2}} |\psi(e^{-i2\pi f})|^2 df \quad (3.13)$$

Maintenant, si l'intégrale (3.13) doit converger, il a été établi (Grenander et Rosenblatt, 1957) que la série infinie  $\psi(B)$  doit converger pour  $B$  sur ou dans le cercle unité. De façon plus directe, pour un processus linéaire  $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ , la condition  $\sum_{j=0}^{\infty} |\psi_j| < \infty$  de l'absolue sommabilité des coefficients  $\psi_j$  implique (Brockwell et Davis, 1991 ; Fuller, 1996) que la somme  $\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$  converge avec la probabilité 1 et représente donc un processus stationnaire valable.

#### Spectre d'un processus linéaire stationnaire

Il a été montré que si nous substituons  $B = e^{-i2\pi f}$ , où  $i = \sqrt{-1}$ , dans la fonction génératrice des autocovariances (3.11), nous obtenons la moitié de la puissance du spectre. Ainsi, le spectre d'un processus linéaire est

$$\begin{aligned} p(f) &= 2\sigma_\varepsilon^2 \psi(e^{-i2\pi f}) \psi(e^{i2\pi f}) \\ &= 2\sigma_\varepsilon^2 |\psi(e^{-i2\pi f})|^2 \quad 0 \leq f \leq \frac{1}{2} \end{aligned} \quad (3.14)$$

En fait, c'est l'expression bien connue (Jenkins et Watts, 1968) qui relie le spectre  $p(f)$  de la sortie d'un système linéaire au spectre uniforme  $2\sigma_\varepsilon^2$  d'un bruit blanc en entrée multiplié par le carré du gain du système  $G^2(f) = |\psi(e^{-i2\pi f})|^2$ .

## Inversibilité

On a déjà vu que les poids  $\psi$  d'un processus linéaire doivent satisfaire à la condition que  $\psi(B)$  converge sur ou dans le cercle d'unité pour que le processus soit stationnaire. Nous considérons maintenant une restriction similaire appliquée aux poids  $\pi$  pour assurer ce qu'on appelle l'inversibilité de la représentation du processus. Cette condition d'inversibilité est indépendante de la condition de stationnarité et s'applique également aux modèles linéaires non stationnaires, que nous introduirons plus tard. Pour illustrer l'idée de base de l'inversibilité, considérons à nouveau le cas particulier

$$X_t = (1 - \theta B) \varepsilon_t \quad (3.15)$$

Exprimant les  $\varepsilon_t$  en fonction des  $X_t$  présent et passés, le modèle devient

$$\varepsilon_t = (1 - \theta B)^{-1} X_t = (1 + \theta B + \theta^2 B^2 + \dots + \theta^k B^k) (1 - \theta^{k+1} B^{k+1})^{-1} X_t$$

Si  $|\theta| < 1$ , en faisant tendre  $k$  vers l'infini, on obtient la série

$$X_t = - \sum_{j=1}^{\infty} \theta^j X_{t-j} + \varepsilon_t \quad (3.16)$$

et les poids  $\pi$  du modèle sous la forme de (3.4) sont  $\pi_j = -\theta^j$ .

Quelle que soit la valeur de  $\theta$ ,  $X_t = (1 - \theta B) \varepsilon_t$  définit un processus stationnaire parfaitement correct. Pour donner un sens à la relation (3.16) nous supposons que  $|\theta| < 1$ . On dit alors que le processus (3.15) est inversible. Nous remarquons que cette condition est équivalente à  $\sum_{j=0}^{\infty} |\theta|^j \equiv \sum_{j=0}^{\infty} |\pi_j| < \infty$ , de sorte que la série

$$\pi(B) = (1 - \theta B)^{-1} = \sum_{j=0}^{\infty} \theta^j B^j$$

converge pour tout  $B < 1$ , c'est-à-dire sur ou dans le cercle de l'unité.

L'exigence de l'inversibilité est nécessaire pour associer un événement présent à ses valeurs passées d'une manière sensée. C'est ce qu'on appelle la propriété de "*causalité*" d'un processus aléatoire.

Le processus linéaire général (3.1) est inversible et peut être écrit sous la forme

$$\pi(B) X_t = \varepsilon_t$$



si les poids  $\pi_j$  sont absolument sommables, c'est-à-dire, si  $\sum_{j=0}^{\infty} |\pi_j| < \infty$ , qui implique que la série  $\pi(B)$  converge sur ou dans le cercle de l'unité. Ainsi, pour résumer, un processus linéaire (3.1) est stationnaire si  $\sum_{j=0}^{\infty} |\psi_j| < \infty$  et est inversible si  $\sum_{j=0}^{\infty} |\pi_j| < \infty$ , où  $\pi(B) = \psi^{-1}(B) = 1 - \sum_{j=0}^{\infty} \pi_j B^j$ .

### 3.1.4 Processus autorégressif et moyenne mobile

Les représentations (3.1) et (3.4) d'un processus linéaire général ne seraient pas très utiles en pratique si elles contenaient un nombre infini de paramètres  $\psi_j$  et  $\pi_j$ . Nous allons décrire maintenant un moyen d'introduire la parcimonie et d'arriver à des modèles qui soient des représentations utiles pour des applications pratiques.

#### Processus autorégressif

Considérons d'abord le cas particulier du modèle (3.4) dans lequel seulement les  $p$  premiers poids sont différents de zéro. Le modèle peut alors être écrit tel que

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t \quad (3.17)$$

où nous utilisons maintenant les symboles  $\phi_1, \phi_2, \dots, \phi_p$  pour l'ensemble fini des poids ou paramètres. Le processus obtenu est appelé *processus autorégressif* d'ordre  $p$ , ou plus succinctement, processus  $AR(p)$ .

Le modèle  $AR(p)$  peut être écrit sous la forme équivalente suivante

$$(1 - \phi_1 B - \phi_2^2 B^2 - \dots - \phi_p B^p) X_t = \varepsilon_t$$

ou bien

$$\phi(B) X_t = \varepsilon_t \quad (3.18)$$

Cela implique que

$$X_t = \frac{1}{\phi(B)} \varepsilon_t = \phi^{-1}(B) \varepsilon_t = \psi(B) \varepsilon_t$$

Par conséquent, le processus autorégressif peut être considéré comme la sortie  $X_t$  d'un filtre linéaire ayant pour fonction de transfert  $\phi^{-1}(B) = \psi(B)$  lorsque l'entrée est un bruit blanc  $\varepsilon_t$ .

### Processus moyenne mobile

Considérons maintenant le cas particulier du modèle (3.1), quand seulement les  $q$  premiers poids  $\psi$  sont différent de zéro. Le processus peut alors être écrit tel que

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (3.19)$$

où nous utilisons maintenant les symboles  $-\theta_1, -\theta_2, \dots, -\theta_q$  pour l'ensemble fini des poids ou paramètres.

Ce processus est appelé *processus moyenne mobile* d'ordre  $q$ , que nous écrivons en abrégé  $MA(q)$ .

En utilisant l'opérateur 'backshift'  $B$ , le modèle  $MA(q)$  peut être écrit sous la forme équivalente

$$X_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t$$

ou plus succinctement tel que

$$X_t = \theta(B) \varepsilon_t \quad (3.20)$$

Par conséquent, le processus moyenne mobile peut être considéré comme la sortie  $X_t$  d'un filtre linéaire de fonction de transfert  $\theta(B)$  lorsque l'entrée est un bruit blanc  $\varepsilon_t$ .

### Processus mixte autorégressif-moyenne mobile

On a déjà montré plus haut qu'un processus en moyenne mobile *fini*  $MA(1)$

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} = (1 - \theta B) \varepsilon_t \quad |\theta| < 1$$

peut être exprimé comme un processus autorégressif *infini*

$$X_t = -\theta_1 X_{t-1} - \theta_1^2 X_{t-2} - \dots + \varepsilon_t$$

Cependant, si le processus était réellement un  $MA(1)$ , nous n'obtiendrions pas de représentation parcimonieuse en utilisant un modèle autorégressif. Inversement, un processus  $AR(1)$  ne peut pas être représenté parcimonieusement en utilisant un modèle en moyenne mobile. En pratique, pour obtenir

un paramétrage parcimonieux, il est souvent utile d'inclure des termes autorégressifs et moyenne mobile dans le modèle. Le modèle ainsi obtenu s'écrit tel que

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

ou bien

$$\phi(B) X_t = \theta(B) \varepsilon_t \quad (3.21)$$

et est appelé processus *mixte autorégressif-moyenne mobile* d'ordre  $(p, q)$ , que nous noterons en abrégé  $ARMA(p, q)$ .

Ecrivons maintenant

$$X_t = \phi^{-1}(B) \theta(B) \varepsilon_t = \frac{1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q}{1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p} \varepsilon_t$$

nous remarquons que le processus mixte ARMA peut être considéré comme la sortie  $X_t$  d'un filtre linéaire, dont la fonction de transfert est le rapport de deux opérateurs polynomiaux  $\theta(B)$  et  $\phi(B)$ , quand l'entrée est un bruit blanc  $\varepsilon_t$ .

Dans la suite, nous discuterons de certaines caractéristiques importantes des modèles autorégressifs, moyennes mobiles et mixtes. En particulier, nous étudierons leurs variances, leurs fonctions d'autocorrélation, leurs spectres et les conditions de stationnarité et d'inversibilité.

## 3.2 Processus autoregressifs

### 3.2.1 Conditions de stationnarité pour un processus autorégressif

Les paramètres  $\phi_1, \phi_2, \dots, \phi_p$  d'un processus  $AR(p)$

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t$$

ou bien

$$(1 - \phi_1 B - \dots - \phi_p B^p) X_t = \varepsilon_t$$

doivent satisfaire certaines conditions pour que le processus soit stationnaire. A titre d'illustration, le processus  $AR(1)$

$$(1 - \phi_1 B) X_{t-1} = \varepsilon_t$$

peut être écrit tel que

$$X_t = (1 - \phi_1 B)^{-1} \varepsilon_t = \sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j} \quad (3.22)$$

à condition que la série dans le membre de droite de (3.22) converge. Par conséquent,

$$\psi(B) = (1 - \phi_1 B)^{-1} = \sum_{j=0}^{\infty} \phi_1^j B^j \quad (3.23)$$

Nous avons vu plus haut que pour la stationnarité du processus,  $\psi(B)$  doit converger pour  $|B| < 1$ , ou de façon équivalente que  $\sum_{j=0}^{\infty} |\phi_1|^j < \infty$ . Ceci implique que le paramètre  $\phi_1$  d'un processus  $AR(1)$  doit satisfaire la condition  $|\phi_1| < 1$  pour assurer la stationnarité. Du fait que la racine de  $1 - \phi_1 B = 0$  est  $B = \phi_1^{-1}$ , cette condition équivaut à dire que la racine de  $1 - \phi_1 B = 0$  doit être à l'extérieur du cercle unité.

Le processus général  $AR(p)$ ,  $\phi(B) X_t = \varepsilon_t$ , peut être exprimé tel que

$$X_t = \phi^{-1}(B) \varepsilon_t = \psi(B) \varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

à condition que l'expression dans le membre de droite converge.

En utilisant la factorisation de  $\phi(B)$

$$\phi(B) = (1 - G_1 B)(1 - G_2 B) \dots (1 - G_p B)$$

où  $G_1^{-1}, \dots, G_p^{-1}$  sont les racines de  $\phi(B) = 0$ , et en développant la fraction rationnelle  $\phi^{-1}(B)$  en éléments simples, on obtient

$$X_t = \phi^{-1}(B) \varepsilon_t = \sum_{i=1}^p \frac{K_i}{1 - G_i B} \varepsilon_t$$

Donc, pour que  $\psi(B) = \phi^{-1}(B)$  soit une série convergente pour  $|B| < 1$ , i.e., pour que les poids  $\psi_j = \sum_{i=1}^p K_i G_i^j$  soient absolument sommables, on doit avoir  $|G_i| < 1$  pour  $i = 1, 2, \dots, p$ . De manière équivalente, les racines de  $\phi(B) = 0$  doivent se situer en dehors du cercle unité. Ces racines sont généralement appelées *zéros du polynôme  $\phi(B)$* . Ainsi, pour la stationnarité, les zéros de  $\phi(B)$  doivent se situer en dehors du cercle unité. Un argument similaire peut être appliqué lorsque les zéros de  $\phi(B)$  ne sont pas tous distincts. L'équation  $\phi(B) = 0$  est appelée *équation caractéristique du processus*. Notons aussi que les racines de  $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  sont les inverses des racines de l'équation polynomiale en  $x$ ,

$$x^p - \phi_1 x^{p-1} - \dots - \phi_p = 0$$

Puisque la série  $\pi(B) = \phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  est finie, aucune restriction n'est requise sur paramètres d'un processus autorégressif pour assurer l'inversibilité.

### Les poids $\psi$

Du fait que  $\psi(B) = \frac{1}{\phi(B)}$  alors  $\phi(B) \psi(B) = 1$ , on déduit aisément que les poids  $\psi_j$  pour un processus  $AR(p)$  satisfont l'équation de différence

$$\psi_j = \phi_1 \psi_{j-1} + \phi_2 \psi_{j-2} + \dots + \phi_p \psi_{j-p} \quad j > 0 \quad (3.24)$$

avec  $\psi_0 = 1$  et  $\psi_j = 0$  pour  $j < 0$ .

A partir de l'équation (3.24) les poids  $\psi_j$  peuvent facilement être calculés récursivement en fonction des  $\phi_i$ . En effet, d'après les propriétés des équations aux différences linéaires, le fait que les poids  $\psi_j$  satisfont l'équation de différence (3.24) implique qu'ils ont une représentation explicite sous forme de  $\psi_j = \sum_{i=1}^p K_i G_i^j$  pour le cas des racines distinctes.

### 3.2.2 Fonction d'autocorrélation et spectre d'un processus autorégressif

#### Fonction d'autocorrélation

Une relation de récurrence utile pour le calcul des autocorrélations d'un processus autorégressif stationnaire est obtenue comme suit ; on multiplie l'équation

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t$$

par  $X_{t-k}$ , pour  $k \geq 0$ , pour obtenir

$$X_{t-k} X_t = \phi_1 X_{t-k} X_{t-1} + \dots + \phi_p X_{t-k} X_{t-p} + X_{t-k} \varepsilon_t \quad (3.25)$$

Et en calculant l'espérance mathématique de (3.25), on obtient l'équation de différence

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p} \quad k > 0 \quad (3.26)$$

Notons que l'espérance  $E(X_{t-k} \varepsilon_t)$  est nulle pour  $k > 0$ , puisque  $X_{t-k}$  ne peut impliquer que les innovations  $\varepsilon_j$  jusqu'à l'instant  $t-k$ , qui ne sont évidemment pas corrélées avec  $\varepsilon_t$ . En divisant l'équation (3.26) par  $\gamma_0$ , on constate que la fonction d'autocorrélation satisfait à la même forme d'équation de différence

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad k > 0 \quad (3.27)$$

Remarquons aussi que l'équation (3.27) est analogue à l'équation de différence satisfaite par le processus  $X_t$  lui-même, mais sans  $\varepsilon_t$ .

Supposons maintenant que cette équation s'écrit

$$\phi(B) \rho_k = 0$$

où  $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  et  $B$  opère maintenant sur  $k$  et non  $t$ .

Le polynôme caractéristique  $\phi(B)$  peut être écrit tel que

$$\phi(B) = \prod_{i=1}^p (1 - G_i B)$$

Alors, la solution générale pour  $\rho_k$  dans (3.27) est

$$\rho_k = A_1 G_1^k + A_2 G_2^k + \dots + A_p G_p^k \quad (3.28)$$

où  $G_1^{-1}, G_2^{-1}, \dots, G_p^{-1}$  sont les racines de l'équation caractéristique

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p = 0$$

Pour la stationnarité, nous avons besoin que  $|G_i| < 1$  pour tout  $i$ . Ainsi, deux situations peuvent survenir en pratique si l'on suppose que les racines  $G_i$  sont distinctes.

1. La racine  $G_i$  est réelle, auquel cas le terme  $A_i G_i^k$  dans (3.28) décroît géométriquement quand  $k$  augmente.
2. Deux racines  $G_i$  et  $G_j$  sont complexes et conjuguées, auquel cas elles contribuent avec un terme

$$D^k \sin(2\pi f k + F)$$

à la fonction d'autocorrélation (3.28), qui aura une forme sinusoïdale amortie, avec un facteur d'amortissement  $D = |G_i| = |G_j|$ .

En général, la fonction d'autocorrélation d'un processus autorégressif stationnaire consistera en un mélange d'exponentielles amorties et d'ondes sinusoïdales amorties.

### Equations de Yule-Walker

Si on développe l'équation (3.27) pour  $k = 1, 2, \dots, p$ , nous obtenons un ensemble d'équations linéaires pour  $\phi_1, \phi_2, \dots, \phi_p$  en fonction de  $\rho_1, \rho_2, \dots, \rho_p$ , c'est-à-dire

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1} \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 + \dots + \phi_p \rho_{p-2} \\ &\dots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \dots + \phi_p \end{aligned} \tag{3.29}$$

Ce sont les équations bien connues de *Yule-Walker* (Yule, 1927; Walker, 1931). Nous obtenons des estimations de Yule-Walker des paramètres en remplaçant dans (3.29) les autocorrélations théoriques  $\rho_k$  par les autocorrélations estimées  $r_k$ . À noter que si on écrit

$$\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \dots \\ \phi_p \end{pmatrix} \quad \rho_p = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \dots \\ \rho_p \end{pmatrix} \quad \mathbf{P}_p = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-2} \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & 1 \end{bmatrix}$$

Alors, la solution de (3.29) pour les paramètres  $\phi$  en fonction des auto-corrélations peut être exprimée telle que

$$\phi = \mathbf{P}_p^{-1} \rho_p \quad (3.30)$$

### Variance

Lorsque  $k = 0$ , la contribution au terme  $E(X_{t-k}\varepsilon_t)$ , en prenant l'espérance dans (3.25), est  $E(\varepsilon_t^2) = \sigma_\varepsilon^2$  puisque la partie de  $X_t$  qui sera corrélée avec  $\varepsilon_t$  est l'innovation la plus récente  $\varepsilon_t$ . Par conséquent, lorsque  $k = 0$ ,

$$\gamma_0 = \phi_1\gamma_{-1} + \phi_2\gamma_{-2} + \dots + \phi_p\gamma_{-p} + \sigma_\varepsilon^2 \quad k > 0 \quad (3.31)$$

En substituant  $\gamma_k = \gamma_{-k}$  dans (3.31) et en écrivant  $\gamma_k = \gamma_0\rho_k$ , la variance  $\gamma_0 = \sigma_X^2$  peut être écrite telle que

$$\sigma_X^2 = \frac{\sigma_\varepsilon^2}{1 - \phi_1\rho_1 - \phi_2\rho_2 - \dots - \phi_p\rho_p} \quad (3.32)$$

### Spectre

Pour un processus  $AR(p)$ ,  $\psi(B) = \phi^{-1}(B)$  et

$$\phi(B) = 1 - \phi_1B - \dots - \phi_pB^p$$

Donc, en utilisant (3.14), le spectre d'un processus autorégressif s'écrit tel que

$$p(f) = \frac{2\sigma_\varepsilon^2}{|1 - \phi_1e^{-i2\pi f} - \phi_2e^{-i4\pi f} - \dots - \phi_pe^{-i2p\pi f}|^2} \quad (3.33)$$

Dans la suite, on va étudier deux processus autorégressifs particulièrement importants, ceux du premier et du second ordre.

### 3.2.3 Processus autorégressif du premier ordre

Le processus autorégressif de premier ordre est

$$\begin{aligned} X_t &= \phi_1X_{t-1} + \varepsilon_t \\ &= \varepsilon_t + \phi_1\varepsilon_{t-1} + \phi_1^2\varepsilon_{t-2} + \dots \end{aligned} \quad (3.34)$$

où  $\phi_1$  doit satisfaire la condition  $-1 < \phi_1 < 1$  pour que le processus soit stationnaire.



### Fonction d'autocorrélation

En utilisant (3.27), la fonction d'autocorrélation satisfait l'équation de différence de premier ordre

$$\rho_k = \phi_1 \rho_{k-1} \quad k > 0 \quad (3.35)$$

qui, avec  $\rho_0 = 1$ , admet pour la solution

$$\rho_k = \phi_1^k \quad k \geq 0$$

Puisque  $-1 < \phi_1 < 1$ , alors la fonction d'autocorrélation d'un  $AR(1)$  décroît exponentiellement vers zéro lorsque  $\phi_1$  est positif mais décroît exponentiellement vers zéro oscillant en signe quand  $\phi_1$  est négatif. En particulier, on peut remarquer que

$$\rho_1 = \phi_1 \quad (3.36)$$

### Variance

En utilisant (3.32), la variance du processus est

$$\sigma_X^2 = \frac{\sigma_\varepsilon^2}{1 - \rho_1 \phi_1} = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2} \quad (3.37)$$

en substituant  $\rho_1 = \phi_1$ .

### Spectre

Finalement, en utilisant (3.33), le spectre est obtenu tel que

$$\begin{aligned} p(f) &= \frac{2\sigma_\varepsilon^2}{|1 - \phi_1 e^{-i2\pi f}|^2} \\ &= \frac{2\sigma_\varepsilon^2}{1 + \phi_1^2 - 2\phi_1 \cos(2\pi f)} \quad 0 \leq f \leq \frac{1}{2} \end{aligned} \quad (3.38)$$

### Exemple

La figure 3.1 montre les réalisations de deux processus  $AR(1)$  avec  $\phi_1 = 0,8$  et  $\phi_1 = -0,8$ , et les fonctions d'autocorrélation théoriques et les spectres correspondants. Ainsi, lorsque le paramètre  $\phi_1 = 0,8$ , i.e. positif et voisin de 1, les valeurs adjacentes dans la série sont similaires et la série présente

une tendance prononcée. Cela se reflète dans la fonction d'autocorrélation qui décroît lentement vers zéro, et dans le spectre qui est dominé par les basses fréquences. Lorsque le paramètre  $\phi_1 = -0,8$ , i.e. négatif et voisin de  $-1$ , la série a une tendance oscillante, ce qui se reflète dans la fonction d'autocorrélation, qui alterne en signe tout en décroissant vers zéro, et dans le spectre qui est dominé par les hautes fréquences.

### 3.2.4 Processus autorégressif du second ordre

#### Condition de stationnarité

Un processus autorégressif du second ordre peut être exprimé tel que

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t \quad (3.39)$$

Pour la stationnarité, les racines de

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 = 0 \quad (3.40)$$

doivent se situer en dehors du cercle unité, ce qui implique que les paramètres  $\phi_1$  et  $\phi_2$  doivent se situer dans la région triangulaire suivante

$$\begin{aligned} \phi_1 + \phi_2 &< 1 \\ \phi_2 - \phi_1 &< 1 \\ -1 &< \phi_2 < 1 \end{aligned} \quad (3.41)$$

#### Fonction d'autocorrélation

En utilisant (3.27), la fonction d'autocorrélation vérifie l'équation de différence de second ordre suivante

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \quad k > 0 \quad (3.42)$$

avec les valeurs initiales  $\rho_0 = 1$  et  $\rho_1 = \frac{\phi_1}{1-\phi_2}$ .

A partir de (3.28), on obtient la solution générale de l'équation de différence (3.42) telle que

$$\begin{aligned} \rho_k &= A_1 G_1^k + A_2 G_2^k \\ &= \frac{G_1 (1 - G_2) G_1^k - G_2 (1 - G_1) G_2^k}{(G_1 - G_2) (1 + G_1 G_2)} \end{aligned} \quad (3.43)$$

où  $G_1^{-1}$  et  $G_2^{-1}$  sont les racines de l'équation caractéristique  $\phi(B) = 0$ . Lorsque les racines sont réelles, la fonction d'autocorrélation consiste en un mélange d'exponentielles amorties. Cela se produit lorsque  $\phi_1^2 + 4\phi_2 \geq 0$ . La fonction d'autocorrélation reste positive tout en tendant vers zéro, ce qui correspond à une racine dominante positive dans (3.43). La fonction d'autocorrélation alterne en signe tout en tendant vers zéro, ce qui correspond à une racine dominante négative. Si les racines  $G_1$  et  $G_2$  sont complexes ( $\phi_1^2 + 4\phi_2 < 0$ ), un processus autorégressif de second ordre affiche un comportement pseudopériodique.

### Equations de Yule-Walker

Pour un modèle  $AR(2)$ , les équations de Yule-Walker s'expriment telles que

$$\begin{aligned}\rho_1 &= \phi_1 + \phi_2 \rho_1 \\ \rho_2 &= \phi_1 \rho_1 + \phi_2\end{aligned}\tag{3.44}$$

qui, lorsqu'elles sont résolues par rapport à  $\phi_1$  et  $\phi_2$ , donnent

$$\begin{aligned}\phi_1 &= \frac{\rho_1(1-\rho_2)}{1-\rho_1^2} \\ \phi_2 &= \frac{\rho_2-\rho_1^2}{1-\rho_1^2}\end{aligned}\tag{3.45}$$

Ces équations peuvent également être résolues pour exprimer  $\rho_1$  et  $\rho_2$  en fonction de  $\phi_1$  and  $\phi_2$  pour donner

$$\begin{aligned}\rho_1 &= \frac{\phi_1}{1-\phi_2} \\ \rho_2 &= \phi_2 + \frac{\phi_1^2}{1-\phi_2}\end{aligned}\tag{3.46}$$

Les valeurs  $\rho_1$  et  $\rho_2$  dans (3.46) fournissent les valeurs initiales pour une solution récursive dans l'équation (3.42).

Cependant, pour calculer les autocorrélations d'un processus  $AR(2)$ , il est plus simple d'utiliser directement les récursions dans (3.42).

En utilisant la condition de stationnarité (3.41) et les expressions de  $\rho_1$  et  $\rho_2$  dans (3.46), on peut déduire que les valeurs admissibles pour  $\rho_1$  et  $\rho_2$ , pour un processus stationnaire  $AR(2)$ , doivent se situer dans la région

$$\begin{aligned}-1 &< \rho_1 < 1 \\ -1 &< \rho_2 < 1 \\ \rho_1^2 &< \frac{1}{2}(\rho_2 + 1)\end{aligned}$$

## Variance

A partir de la relation (3.32), on peut d  duire la variance du processus  $AR(2)$  telle que

$$\begin{aligned}\sigma_X^2 &= \frac{\sigma_\varepsilon^2}{1 - \rho_1\phi_1 - \rho_2\phi_2} \\ &= \frac{1 - \phi_2}{1 + \phi_2} \frac{\sigma_\varepsilon^2}{(1 - \phi_2)^2 - \phi_1^2}\end{aligned}\quad (3.47)$$

## Spectre

A partir de la relation (3.33), on d  duit le spectre d'un processus  $AR(2)$  tel que

$$\begin{aligned}p(f) &= \frac{2\sigma_\varepsilon^2}{|1 - \phi_1 e^{-i2\pi f} - \phi_2 e^{-i4\pi f}|^2} \\ &= \frac{2\sigma_\varepsilon^2}{1 + \phi_1^2 + \phi_2^2 - 2\phi_1(1 - \phi_2)\cos(2\pi f) - 2\phi_2\cos(2\pi f)} \quad 0 \leq f \leq \frac{1}{2}\end{aligned}\quad (3.48)$$

### 3.2.5 Fonction d'autocorr  lation partielle

En pratique, on ne connaît pas l'ordre du processus autor  gressif. Alors il doit   tre sp  cifi      partir des donn  es.

La fonction d'autocorr  lation partielle est un outil qui exploite le fait que, alors qu'un processus  $AR(p)$  poss  de une fonction d'autocorr  lation infinie, les autocorr  lations partielles sont nulles au-del   d'un horizon  $p$ .

Les autocorr  lations partielles peuvent   tre d  crites en termes de  $p$  fonctions non nulles des autocorr  lations. Notons  $\phi_{kj}$  le  $j^{\text{i  me}}$  coefficient dans une repr  sentation autor  gressive d'ordre  $k$ , de sorte que  $\phi_{kk}$  est le dernier coefficient. De (3.27), les  $\phi_{kj}$  satisfont au syst  me d'  quations

$$\rho_j = \phi_{k1}\rho_{j-1} + \dots + \phi_{k(k-1)}\rho_{j-k+1} + \phi_{kk}\rho_{j-k} \quad j = 1, 2, \dots, k \quad (3.49)$$

conduisant aux   quations de Yule-Walker (3.29), qui peuvent s'  crire

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-2} \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \dots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \dots \\ \rho_k \end{bmatrix} \quad (3.50)$$

ou bien

$$\mathbf{P}_k \boldsymbol{\phi}_k = \boldsymbol{\rho}_k \quad (3.51)$$

En résolvant ces équations successivement pour  $k = 1, 2, 3, \dots$ , on obtient

$$\begin{aligned} \phi_{11} &= \rho_1 \\ \phi_{22} &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \\ \phi_{33} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} \end{aligned} \quad (3.52)$$

En général, pour  $\phi_{kk}$ , le déterminant du numérateur a les mêmes éléments que celui du dénominateur, mais sa dernière colonne est remplacée par le vecteur  $\boldsymbol{\rho}_k$ . La quantité  $\phi_{kk}$ , considérée comme une fonction de l'horizon  $k$ , est appelée *fonction d'autocorrélation partielle*.

Pour un processus  $AR(p)$ , les autocorrélations partielles  $\phi_{kk}$  sont non nulles pour  $k \leq p$  et nulles pour  $k > p$ . En d'autres termes, la fonction d'autocorrélation partielle d'un processus  $AR(p)$  chute à zéro après l'horizon  $p$ . A titre d'exemple, les autocorrélations partielles du processus  $AR(2)$   $X_t = 0,75X_{t-1} - 0,50X_{t-2} + \varepsilon_t$  sont  $\phi_{11} = \rho_1 = 0.5$ ,  $\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = -0.5 \equiv \phi_2$ , et  $\phi_{kk} = 0$ , pour tout  $k > 2$ .

La quantité  $\phi_{kk}$  est appelée autocorrélation partielle du processus  $\{X_t\}$  à horizon  $k$ , puisqu'elle est égale à la corrélation entre  $X_t$  et  $X_{t-k}$  sans prendre en compte  $X_{t-1}, X_{t-2}, \dots, X_{t-k+1}$ . Maintenant, il est facile d'établir à partir de la théorie des moindres carrés que les valeurs  $\phi_{k1}, \phi_{k2}, \dots, \phi_{kk}$ , qui sont les solutions de (3.50), sont les coefficients dans la régression linéaire de  $X_t$  sur  $X_{t-1}, \dots, X_{t-k}$ , c'est-à-dire, ce sont les valeurs des coefficients  $b_1, \dots, b_k$ , qui minimise  $E \left[ \left( X_t - b_0 - \sum_{i=1}^k b_i X_{t-i} \right)^2 \right]$ . Donc, en supposant par commodité que le processus  $\{X_t\}$  est centré, le meilleur prédicteur linéaire, dans le sens de l'erreur quadratique moyenne, de  $X_t$  basé sur  $X_{t-1}, X_{t-2}, \dots, X_{t-k+1}$  est

$$\hat{X}_t = \phi_{k-1,1} X_{t-1} + \phi_{k-1,2} X_{t-2} + \dots + \phi_{k-1,k-1} X_{t-k+1}$$

si le processus est un  $AR$  ou pas.

De même, le meilleur prédicteur linéaire de  $X_{t-k}$  basé sur les valeurs futures  $X_{t-1}, X_{t-2}, \dots, X_{t-k+1}$  est

$$\hat{X}_{t-k} = \phi_{k-1,1}X_{t-k+1} + \phi_{k-1,2}X_{t-k+2} + \dots + \phi_{k-1,k-1}X_{t-1}$$

Ainsi, l'autocorrélation partielle à horizon  $k$  d'un processus  $\{X_t\}$ ,  $\phi_{kk}$  peut être définie comme la corrélation entre les résidus de ces deux régressions, c'est-à-dire

$$\phi_{kk} = \text{Corr}(X_t - \hat{X}_t, X_{t-k} - \hat{X}_{t-k}) \quad (3.53)$$

Par exemple, on trouve que  $\phi_{11} = \text{Corr}(X_t, X_{t-1}) = \rho_1$ , alors que

$$\begin{aligned} \phi_{22} &= \text{Corr}(X_t - \rho_1 X_{t-1}, X_{t-2} - \rho_1 X_{t-1}) \\ &= \frac{\gamma_2 - 2\rho_1\gamma_1 + \rho_1^2\gamma_0}{\left[(\gamma_0 + \rho_1^2\gamma_0 - 2\rho_1\gamma_1)^2\right]^{\frac{1}{2}}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \end{aligned}$$

qui est conforme aux résultats énoncés dans (3.52) et dérivés des équations de Yule-Walker. Les autocorrélations partielles d'ordre supérieur  $\phi_{kk}$  définies par (3.53) peuvent également être considérées comme étant la solution d'un système d'équations de Yule-Walker approprié.

### 3.2.6 Estimation de la fonction d'autocorrélation partielle

Les autocorrélations partielles peuvent être estimées en ajustant successivement les modèles autorégressifs d'ordres  $1, 2, 3, \dots$  par la méthode des moindres carrés et en choisissant les estimations  $\hat{\phi}_{11}, \hat{\phi}_{22}, \hat{\phi}_{33}, \dots$  du dernier coefficient ajusté à chaque étape. Sinon, si les valeurs des paramètres ne sont pas trop proches des limites non stationnaires, des estimations approximatives de Yule-Walker des modèles autorégressifs successifs peuvent être utilisées. Les estimations des autocorrélations partielles peuvent alors être obtenues en substituant les estimations  $r_j$  aux autocorrélations théoriques dans (3.49), pour obtenir

$$r_j = \hat{\phi}_{k1}r_{j-1} + \dots + \hat{\phi}_{k(k-1)}r_{j-k+1} + \hat{\phi}_{kk}r_{j-k} \quad j = 1, 2, \dots, k \quad (3.54)$$

et résoudre les équations résultantes pour  $k = 1, 2, \dots$

Cela peut être fait en utilisant une méthode récursive simple due à Levinson (1947) et Durbin (1960), que nous décrirons plus tard. Cependant, ces

estimations obtenues à partir de (3.54) deviennent très sensibles aux erreurs d'arrondi et ne devraient pas être utilisées si les valeurs des paramètres sont proches des limites non stationnaires.

### 3.2.7 Erreurs standards des estimateurs des autocorrélations partielles

Il a été montré par Quenouille (1949) que dans l'hypothèse que le processus est autorégressif d'ordre  $p$ , les autocorrélations partielles estimées d'ordre  $p + 1$ , et plus, sont centrées, approximativement indépendantes et normalement distribuées. En outre, si  $n$  est le nombre d'observations utilisées dans l'ajustement,

$$\text{Var}(\hat{\phi}_{kk}) \simeq \frac{1}{n} \quad k \geq p + 1$$

Ainsi, l'erreur standard (SE) de l'autocorrélation partielle estimée  $\hat{\phi}_{kk}$  est

$$SE(\hat{\phi}_{kk}) = \hat{\sigma}(\hat{\phi}_{kk}) = \frac{1}{\sqrt{n}} \quad k \geq p + 1 \quad (3.55)$$

### 3.2.8 Calculs par R

L'estimation de la fonction d'autocorrélation partielle est commodément effectuée à l'aide du logiciel R. Par exemple, la commande "*pacf()*" dans le package "*stats*" donne les autocorrélations partielles estimées. Une alternative est d'utiliser la commande "*acf2()*" dans le package R "*astsa*". Cette commande a l'avantage de produire des tracés des fonctions d'autocorrélation et d'autocorrélation partielle dans un seul graphe. Ceci permet une comparaison facile des deux fonctions, ce qui sera utile pour spécifier un modèle pour la série chronologique étudiée.

## 3.3 Processus moyenne mobile

### 3.3.1 Conditions d'inversibilité pour les processus moyenne mobile

Nous allons établir les conditions que doivent vérifier les paramètres  $\theta_1, \theta_2, \dots, \theta_q$  pour assurer l'inversibilité d'un processus  $MA(q)$  :

$$\begin{aligned} X_t &= \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \\ &= (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t \\ &= \theta(B) \varepsilon_t \end{aligned} \quad (3.56)$$

Nous avons déjà vu plus haut que le processus moyenne mobile ordre 1

$$X_t = (1 - \theta_1 B) \varepsilon_t$$

est inversible si  $|\theta_1| < 1$  ; i.e. la série

$$\pi(B) = (1 - \theta_1 B)^{-1} = \sum_{j=0}^{\infty} \theta_1^j B^j$$

converge sur ou dans le cercle unité. Ce qui est équivalent à dire que la racine,  $B = \theta_1^{-1}$  de l'équation  $1 - \theta_1 B = 0$ , se trouve en dehors du cercle unité.

La condition d'inversibilité pour un processus  $MA$  d'ordre  $q$  quelconque peut être obtenue en écrivant  $X_t = \theta(B) \varepsilon_t$  tel que

$$\varepsilon_t = \theta^{-1}(B) X_t$$

où  $\theta(B) = \prod_{i=1}^q (1 - H_i B)$  et  $H_1^{-1}, \dots, H_q^{-1}$  sont les racines de  $\theta(B) = 0$ .

Alors, en développant la fraction rationnelle  $\pi(B) = \theta^{-1}(B)$  en éléments simples on obtient

$$\pi(B) = \theta^{-1}(B) = \sum_{i=1}^q \frac{M_i}{1 - H_i B}$$

Ou bien de façon plus explicite

$$\pi(B) = \sum_{i=1}^q M_i \left( \sum_{j=0}^{\infty} H_i^j B^j \right) = \sum_{j=0}^{\infty} \left( \sum_{i=1}^q M_i H_i^j \right) B^j \quad (3.57)$$



La série dans l'équation (3.57) converge, ou de manière équivalente les poids  $\pi_j = -\sum_{i=1}^q M_i H_i^j$  sont absolument sommables, si  $|H_i| < 1$ , pour  $i = 1, 2, \dots, q$ . Il s'ensuit que la condition d'inversibilité pour un processus  $MA(q)$  est que les racines  $H_i^{-1}$ ,  $i = 1, 2, \dots, q$  de l'équation caractéristique

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q = 0 \quad (3.58)$$

doivent se trouver à l'extérieur du cercle unité.

De la relation  $\theta(B)\pi(B) = 1$ , il s'ensuit que les poids  $\pi_j$  satisfont à l'équation de différence

$$\pi_j = \theta_1 \pi_{j-1} + \theta_2 \pi_{j-2} + \dots + \theta_q \pi_{j-q} \quad j > 0$$

Avec la convention  $\pi_0 = -1$  et  $\pi_j = 0$  pour  $j < 0$ , les poids  $\pi_j$  peuvent facilement être calculés récursivement en fonction des  $\theta_i$ .

Il est clair que, puisque la série

$$\psi(B) = \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

est finie, aucune condition sur les paramètres du processus moyenne mobile n'est nécessaire pour assurer la stationnarité.

### 3.3.2 Fonction d'autocorrélation et spectre d'un processus moyenne mobile

#### Fonction d'autocorrélation

La fonction d'autocovariance d'un processus  $MA(q)$  s'écrit

$$\begin{aligned} \gamma_k &= E[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q})(\varepsilon_{t-k} - \theta_1 \varepsilon_{t-k-1} - \dots - \theta_q \varepsilon_{t-k-q})] \\ &= -\theta_k E(\varepsilon_{t-k}^2) + \theta_1 \theta_{k-1} E(\varepsilon_{t-k-1}^2) + \dots + \theta_{q-k} \theta_q E(\varepsilon_{t-q}^2) \end{aligned}$$

Puisque les innovations  $\varepsilon_t$  ne sont pas corrélées, et  $\gamma_k = 0$  pour  $k > q$ , alors la variance du processus est telle que

$$\gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma_\varepsilon^2 \quad (3.59)$$

et l'autocovariance est

$$\gamma_k = \begin{cases} (-\theta_k + \theta_1 \theta_{k-1} + \dots + \theta_{q-k} \theta_q) \sigma_\varepsilon^2 & k = 1, 2, \dots, q \\ 0 & k > q \end{cases} \quad (3.60)$$

D'où la fonction d'autocorrélation s'exprime telle que

$$\rho_k = \begin{cases} \frac{-\theta_k + \theta_1\theta_{k-1} + \dots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} & k = 1, 2, \dots, q \\ 0 & k > q \end{cases} \quad (3.61)$$

Nous constatons que la fonction d'autocorrélation d'un processus  $MA(q)$  chute à zéro à partir de l'ordre  $q + 1$  du processus aléatoire. En d'autres termes, la fonction d'autocorrélation d'un processus moyenne mobile chute à zéro après l'horizon  $q$ .

### Paramètres 'moyenne mobile' en fonction des autocorrélations

Si les autocorrélations  $\rho_1, \rho_2, \dots, \rho_q$  sont connues, les  $q$  équations (3.61) peuvent être résolues pour les paramètres  $\theta_1, \theta_2, \dots, \theta_q$ . Cependant, contrairement aux équations de Yule-Walker (3.29) pour un processus autorégressif, les équations (3.61) sont non linéaires. Par conséquent, sauf dans le cas simple où  $q = 1$ , ces équations doivent être résolues itérativement à l'aide de méthodes numériques.

Les estimations des paramètres de la moyenne mobile peuvent être obtenues en substituant les estimations  $r_k$  à  $\rho_k$  et en résolvant les équations résultantes. Cependant, contrairement aux estimations autorégressives obtenues à partir des équations de Yule-Walker, les estimations des paramètres de la moyenne mobile peuvent être moins efficaces. Néanmoins, elles peuvent fournir des estimations approximatives utiles à l'étape d'identification de modèle discutée plus loin. En outre, elles fournissent des valeurs initiales utiles pour une procédure itérative d'estimation des paramètres qui convergent vers les estimations efficaces du maximum de vraisemblance.

### Spectre

Pour un processus  $MA(q)$ ,

$$\psi(B) = \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

Donc, en utilisant la relation (3.14), le spectre d'un processus  $MA(q)$  est tel que

$$p(f) = 2\sigma_\varepsilon^2 \left| 1 - \theta_1 e^{-i2\pi f} - \theta_2 e^{-i4\pi f} - \dots - \theta_q e^{-i2q\pi f} \right|^2 \quad 0 \leq f \leq \frac{1}{2} \quad (3.62)$$

Nous allons maintenant étudier plus en détail les processus de moyenne mobile du premier et du second ordre, qui ont une importance pratique considérable.

### 3.3.3 Processus moyenne mobile du premier ordre

Nous avons déjà introduit le processus  $MA(1)$

$$\begin{aligned} X_t &= \varepsilon_t - \theta_1 \varepsilon_{t-1} \\ &= (1 - \theta_1 B) \varepsilon_t \end{aligned}$$

et nous avons vu que  $\theta_1$  doit se situer dans l'intervalle  $-1 < \theta_1 < 1$  pour que le processus soit inversible. Le processus est, bien sûr, stationnaire pour toutes les valeurs de  $\theta_1$ .

#### Fonction d'autocorrélation

Il est facile de voir que la variance de ce processus est égale à

$$\gamma_0 = (1 + \theta_1^2) \sigma_\varepsilon^2$$

La fonction d'autocorrélation est telle que

$$\rho_k = \begin{cases} \frac{-\theta_1}{1+\theta_1^2} & k = 1 \\ 0 & k > 1 \end{cases} \quad (3.63)$$

De là on peut remarquer que  $\rho_1$  doit satisfaire  $|\rho_1| = \frac{|\theta_1|}{1+\theta_1^2} \leq \frac{1}{2}$ . Aussi, pour  $k = 1$ , on trouve que l'équation

$$\rho_1 \theta_1^2 + \theta_1 + \rho_1 = 0 \quad (3.64)$$

admet des racines pour  $\theta_1$  égales à  $\theta_1 = -1 \pm \frac{\sqrt{1-4\rho_1^2}}{2\rho_1}$ .

Puisque le produit des racines est l'unité, nous remarquons que si  $\theta_1$  est une solution alors  $\theta_1^{-1}$  l'est aussi. De plus, si  $\theta_1$  satisfait la condition d'inversion  $|\theta_1| < 1$ , l'autre racine  $\theta_1^{-1}$  sera plus grande que l'unité et ne pourra pas satisfaire la condition. Par exemple, si  $\rho_1 = -0,4$ , les deux solutions sont  $\theta_1 = 0,5$  et  $\theta_1 = 2,0$ . Cependant, seule la solution  $\theta_1 = 0,5$  correspond à un modèle inversible.

## Spectre

En utilisant la relation (3.62), le spectre du processus  $MA(1)$  s'écrit tel que

$$\begin{aligned} p(f) &= 2\sigma_\varepsilon^2 |1 - \theta_1 e^{-i2\pi f}|^2 \\ &= 2\sigma_\varepsilon^2 [1 + \theta_1^2 - 2\theta_1 \cos(2\pi f)] \quad 0 \leq f \leq \frac{1}{2} \end{aligned} \quad (3.65)$$

En général, lorsque  $\theta_1$  est négatif,  $\rho_1$  est positif, et le spectre est dominé par les basses fréquences. Inversement, lorsque  $\theta_1$  est positif,  $\rho_1$  est négatif, et le spectre est dominé par les hautes fréquences.

## Fonction d'autocorrélation partielle

En utilisant (3.50) avec  $\rho_1 = \frac{-\theta_1}{1+\theta_1^2}$  et  $\rho_k = 0$ , pour  $k > 1$ , on obtient après quelques manipulations algébriques

$$\phi_{kk} = \frac{-\theta_1^k (1 - \theta_1^2)}{1 - \theta_1^{2(k+1)}}$$

Ainsi,  $|\phi_{kk}| < |\theta_1|^k$ , et la fonction d'autocorrélation partielle est dominée par une exponentielle amortie. Si  $\rho_1$  est positif, de sorte que  $\theta_1$  soit négatif, les autocorrélations partielles alternent en signe. Si, par contre,  $\rho_1$  est négatif, de sorte que  $\theta_1$  soit positif, les autocorrélations partielles sont négatives. On a vu que les poids  $\pi_j$  pour le processus  $MA(1)$  sont tels que  $\pi_j = -\theta_1^j$ , et donc que ce sont des coefficients dans la forme autorégressive infinie du processus, il est logique que la fonction d'autocorrélation partielle  $\phi_{kk}$  pour le  $MA(1)$  mime essentiellement la caractéristique de décroissance exponentielle des poids  $\pi_j$ .

**Remarque 3** *Il est clair qu'il y a dualité entre les processus  $AR(1)$  et  $MA(1)$ . En effet,*

1. *Alors que la fonction d'autocorrélation d'un processus  $MA(1)$  chute à zéro après l'horizon 1, la fonction d'autocorrélation d'un processus  $AR(1)$  tend vers zéro exponentiellement.*
2. *Réciproquement, alors que la fonction d'autocorrélation partielle d'un processus  $MA(1)$  tend vers zéro tout en étant dominé par une exponentielle amortie, la fonction d'autocorrélation partielle d'un processus  $AR(1)$  chute à zéro après l'horizon 1.*

3. Il s'avère qu'une dualité approximative correspondante de ce genre se produit en général dans les fonctions d'autocorrélation et d'autocorrélation partielle entre les processus AR et MA.

### 3.3.4 Processus moyen mobile du second ordre

#### Conditions d'inversibilité

Le processus moyenne mobile du second ordre est défini tel que

$$\begin{aligned} X_t &= \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \\ &= (1 - \theta_1 B - \theta_2 B^2) \varepsilon_t \end{aligned}$$

et est stationnaire pour toutes les valeurs de  $\theta_1$  et de  $\theta_2$ . Cependant, il n'est inversible que si les racines de l'équation caractéristique

$$1 - \theta_1 B - \theta_2 B^2 = 0 \quad (3.66)$$

se situe à l'extérieur du cercle unité, c'est-à-dire

$$\begin{aligned} \theta_1 + \theta_2 &< 1 \\ \theta_2 - \theta_1 &< 1 \\ -1 &< \theta_2 < 1 \end{aligned} \quad (3.67)$$

Ces conditions sont semblables aux conditions (3.41) requises pour la stationnarité d'un processus  $AR(2)$ .

#### Fonction d'autocorrélation

En utilisant (3.59), on obtient la variance du processus telle que

$$\gamma_0 = (1 + \theta_1^2 + \theta_2^2) \sigma_\varepsilon^2$$

et en utilisant (3.61), la fonction d'autocorrélation s'écrit telle que

$$\begin{aligned} \rho_1 &= \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2} \\ \rho_2 &= \frac{-\theta_2}{1+\theta_1^2+\theta_2^2} \\ \rho_k &= 0 \quad k > 2 \end{aligned} \quad (3.68)$$

Ainsi, la fonction d'autocorrélation s'annule après l'horizon 2.

Il résulte de (3.67) et (3.68) que les deux premières autocorrélations d'un

processus  $MA(2)$  inversible doivent se situer dans la zone délimitée par les segments de courbe suivants

$$\begin{aligned}\rho_1 + \rho_2 &= -0,5 \\ \rho_2 - \rho_1 &= -0,5 \\ \rho_1^2 &= 4\rho_2(1 - 2\rho_2)\end{aligned}\tag{3.69}$$

## Spectre

En utilisant (3.62), le spectre, pour  $0 \leq f \leq \frac{1}{2}$ , d'un processus  $MA(2)$  est tel que

$$\begin{aligned}p(f) &= 2\sigma_\varepsilon^2 |1 - \theta_1 e^{-i2\pi f} - \theta_2 e^{-i4\pi f}|^2 \\ &= 2\sigma_\varepsilon^2 [1 + \theta_1^2 + \theta_2^2 - 2\theta_1(1 - \theta_2)\cos(2\pi f) - 2\theta_2\cos(4\pi f)]\end{aligned}\tag{3.70}$$

et est l'inverse du spectre (3.48) d'un processus autorégressif du second ordre, à l'exception de la constante  $2\sigma_\varepsilon^2$ .

## Fonction d'autocorrélation partielle

L'expression exacte de la fonction d'autocorrélation partielle d'un processus  $MA(2)$  est compliquée. Cependant on peut dire qu'elle est dominée par la somme de deux exponentielles si les racines de l'équation caractéristique  $1 - \theta_1 B - \theta_2 B^2 = 0$  sont réelles, et par une onde sinusoïdale amortie si les racines sont complexes. Ainsi, elle se comporte comme la fonction d'autocorrélation d'un processus  $AR(2)$ .

**Exemple.** A titre d'illustration, considérons le modèle moyenne mobile du second ordre

$$X_t = \varepsilon_t - 0,8\varepsilon_{t-1} + 0,5\varepsilon_{t-2}$$

La variance du processus est  $\gamma_0 = \sigma_\varepsilon^2 (1 + (0,8)^2 + (-0,5)^2) = 1,89\sigma_\varepsilon^2$  et à partir de (3.68) les autocorrélations théoriques sont telles que

$$\begin{aligned}\rho_1 &= \frac{-0,8(1+0,5)}{1+(0,8)^2+(-0,5)^2} = -0,635 \\ \rho_2 &= \frac{0,5}{1,89} = 0,265 \\ \rho_k &= 0 \quad k > 2\end{aligned}$$

Les autocorrélations partielles théoriques sont obtenues en résolvant successivement les équations (3.49) ou (3.50).

Les fonctions d'autocorrélation et d'autocorrélation partielle peuvent être générées à l'aide de la fonction '*ARMAacf()*' du package de R '*stats*'.

### 3.3.5 Dualité entre les processus autorégressifs et les processus moyenne mobile

Nous avons étudié les propriétés des processus autorégressifs et des processus moyenne mobile et discuté de la dualité entre ces deux types de processus. Cette dualité a les conséquences suivantes :

1. Dans un processus autorégressif stationnaire d'ordre  $p$ ,  $\varepsilon_t$  peut être représenté comme une somme pondérée finie des  $X$  précédents, ou bien  $X_t$  comme une somme pondérée infinie

$$X_t = \phi^{-1}(B) \varepsilon_t$$

des  $\varepsilon$  antérieurs.

Réciproquement, un processus moyenne mobile inversible d'ordre  $q$ ,  $X_t$ , peut être représenté comme une somme pondérée finie des  $\varepsilon$  précédents, ou bien  $\varepsilon_t$  comme une somme pondérée infinie

$$\theta^{-1}(B) X_t = \varepsilon_t$$

des  $X$  antérieurs.

2. Un processus  $MA$  fini a une fonction d'autocorrélation nulle au-delà d'un certain point, mais étant équivalent à un processus  $AR$  infini, sa fonction d'autocorrélation partielle est infinie et est dominée par des exponentielles amorties et/ou d'ondes sinusoïdales amorties. Inversement, le processus  $AR$  a une fonction d'autocorrélation partielle qui est nulle au-delà d'un certain point, mais sa fonction d'autocorrélation est infinie et consiste en un mélange d'exponentielles amorties et/ou d'ondes sinusoïdales amorties.
3. Pour un processus autorégressif d'ordre fini  $p$ , les paramètres ne sont pas astreints aux conditions permettant d'assurer l'inversibilité. Cependant, pour la stationnarité, les racines de  $\phi(B) = 0$  doivent se situer en dehors du cercle unité. Inversement, les paramètres d'un processus  $MA$  ne sont pas quant à eux astreints aux conditions permettant d'assurer la stationnarité. Cependant, pour l'inversion, les racines de  $\theta(B) = 0$  doivent se situer en dehors du cercle de l'unité.
4. Le spectre d'un processus moyenne mobile est l'inverse du spectre du processus autorégressif correspondant.

## 3.4 Processus mixte autorégressif-moyenne mobile

### 3.4.1 Propriétés de stationnarité et d'inversibilité

Nous avons remarqué plus haut que pour obtenir la parcimonie, il peut être nécessaire d'inclure des termes autorégressifs et des termes moyenne mobile dans la construction d'un modèle. Ainsi, nous pouvons avoir besoin d'utiliser le modèle mixte ARMA défini tel que

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (3.71)$$

ou bien,

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) X_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t$$

ou plus succinctement

$$\phi(B) X_t = \theta(B) \varepsilon_t$$

où  $\phi(B)$  et  $\theta(B)$  sont des opérateurs polynomiaux en  $B$  de degrés  $p$  et  $q$  respectivement.

Nous nous référons par la suite à ce type de processus comme un processus  $ARMA(p, q)$ . Cela peut être interprété de deux façons : il est évident que les termes 'moyenne mobile' à droite de l'équation (3.71) n'affecteront pas les arguments précédents qui établissent les conditions de stationnarité d'un processus autorégressif. Ainsi,  $\phi(B) X_t = \theta(B) \varepsilon_t$  définira un processus stationnaire à condition que l'équation caractéristique  $\phi(B) = 0$  admet toutes ses racines en dehors du cercle unité. De même, les racines de  $\theta(B) = 0$  doivent se trouver à l'extérieur du cercle unité si le processus doit être inversible. Ainsi, le processus  $ARMA(p, q)$  stationnaire et inversible (3.71) possède à la fois la représentation moyenne mobile infinie

$$X_t = \psi(B) \varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

où  $\psi(B) = \phi^{-1}(B) \theta(B)$ , et la représentation autorégressive infinie

$$\pi(B) X_t = X_t - \sum_{j=1}^{\infty} \pi_j X_{t-j} = \varepsilon_t$$



où  $\pi(B) = \theta^{-1}(B)\phi(B)$ , avec simultanément les poids  $\psi_j$  et les poids  $\pi_j$  absolument sommables. Les poids  $\psi_j$  sont déterminés à partir de la relation  $\phi(B)\psi(B) = \theta(B)$  et vérifient

$$\psi_j = \phi_1\psi_{j-1} + \phi_2\psi_{j-2} + \dots + \phi_p\psi_{j-p} - \theta_j \quad j > 0$$

avec  $\psi_0 = 1$ ,  $\psi_j = 0$  pour  $j < 0$ , et  $\theta_j = 0$  pour  $j > q$ .

Alors qu'à partir de la relation  $\theta(B)\pi(B) = \phi(B)$  on détermine les poids  $\pi_j$  qui satisfont à la relation suivante

$$\pi_j = \theta_1\pi_{j-1} + \theta_2\pi_{j-2} + \dots + \theta_q\pi_{j-q} - \phi_j \quad j > 0$$

avec  $\pi_0 = -1$ ,  $\pi_j = 0$  pour  $j < 0$ , et  $\phi_j = 0$  pour  $j > p$ . De ces relations, les poids  $\psi_j$  et  $\pi_j$  peuvent facilement être calculés récursivement en fonction des coefficients  $\phi_i$  et  $\theta_i$ .

### 3.4.2 Fonction d'autocorrélation et spectre des processus mixtes

#### Fonction d'autocorrélation

La fonction d'autocorrélation du processus mixte peut être obtenue par une méthode similaire à celle utilisée pour les processus autorégressifs. En multipliant l'équation (3.71) par  $X_{t-k}$  et en passant aux espérances, nous remarquons que la fonction d'autocovariance vérifie l'équation de différence suivante

$$\gamma_k = \phi_1\gamma_{k-1} + \dots + \phi_p\gamma_{k-p} + \gamma_{X\varepsilon}(k) - \theta_1\gamma_{X\varepsilon}(k-1) - \theta_q\gamma_{X\varepsilon}(k-q)$$

où  $\gamma_{X\varepsilon}(k)$  est la fonction de covariance croisée entre  $X$  et  $\varepsilon$  et est définie par  $\gamma_{X\varepsilon}(k) = E[X_{t-k}\varepsilon_t]$ .

Puisque  $X_{t-k}$  ne dépend que des innovations qui se sont produites jusqu'à l'instant  $t-k$  à travers la représentation moyenne mobile infinie  $X_{t-k} = \psi(B)\varepsilon_{t-k} = \sum_{j=0}^{\infty} \psi_j\varepsilon_{t-k-j}$ , il s'ensuit que

$$\gamma_{X\varepsilon}(k) = \begin{cases} 0 & k > 0 \\ \psi_{-k}\sigma_{\varepsilon}^2 & k \leq 0 \end{cases}$$

Par conséquent, l'équation précédente pour  $\gamma_k$  peut être exprimée telle que

$$\gamma_k = \phi_1\gamma_{k-1} + \dots + \phi_p\gamma_{k-p} - \sigma_{\varepsilon}^2(\theta_k\psi_0 + \theta_{k+1}\psi_1 + \dots + \theta_q\psi_{q-k}) \quad (3.72)$$

avec la convention  $\psi_0 = -1$ .

On constate que cela implique

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p} \quad k \geq q+1$$

et donc

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad k \geq q+1 \quad (3.73)$$

ou bien

$$\phi(B) \rho_k = 0 \quad k \geq q+1$$

Ainsi, pour un processus  $ARMA(p, q)$ , il y aura  $q$  autocorrélations  $\rho_1, \rho_2, \dots, \rho_q$  dont les valeurs dépendent directement du choix des  $q$  paramètres moyenne mobile  $\theta_i$ , ainsi que des  $p$  paramètres autorégressifs  $\phi_j$ . De même, les  $p$  valeurs  $\rho_{q-p+1}, \dots, \rho_q$  fournissent les valeurs initiales nécessaires pour l'équation de différence  $\phi(B) \rho_k = 0$ , où  $k \geq q+1$ , qui détermine alors entièrement les autocorrélations à des horizons plus élevés. Si  $q - p < 0$ , toutes les autocorrélations  $\rho_j$ ,  $j = 0, 1, 2, \dots$ , seront constituées d'un mélange d'exponentielles amorties et/ou d'ondes sinusoïdales amorties, dont la nature est déterminée par les racines du polynôme  $\phi(B)$  et les valeurs initiales. Si, cependant,  $q - p \geq 0$ , il y aura  $q - p + 1$  valeurs initiales  $\rho_0, \rho_1, \dots, \rho_{q-p}$ , qui ne suivent pas cette allure générale. Ces caractéristiques sont utiles pour identifier les modèles mixtes.

## Variance

Quand  $k = 0$ , on a

$$\gamma_0 = \phi_1 \gamma_1 + \dots + \phi_p \gamma_p + \sigma_\varepsilon^2 (1 - \theta_1 \psi_1 - \dots - \theta_q \psi_q) \quad (3.74)$$

que l'on peut résoudre à l'aide des  $p$  équations (3.72) pour  $k = 1, 2, \dots, p$  pour obtenir  $\gamma_0, \gamma_1, \dots, \gamma_p$ .

## Spectre

En utilisant (3.14), le spectre du processus mixte  $ARMA(p, q)$  est tel que

$$\begin{aligned} p(f) &= 2\sigma_\varepsilon^2 \frac{|\theta(e^{-i2\pi f})|^2}{|\phi(e^{-i2\pi f})|^2} \\ &= 2\sigma_\varepsilon^2 \frac{|1 - \theta_1 e^{-i2\pi f} - \dots - \theta_q e^{-i2q\pi f}|^2}{|1 - \phi_1 e^{-i2\pi f} - \dots - \phi_q e^{-i2q\pi f}|^2} \quad 0 \leq f \leq \frac{1}{2} \end{aligned} \quad (3.75)$$

### Fonction d'autocorrélation partielle

Le processus mixte  $\phi(B)X_t = \theta(B)\varepsilon_t$  peut être écrit tel que

$$\varepsilon_t = \theta^{-1}(B)\phi(B)X_t$$

où  $\theta^{-1}(B)$  est une série en  $B$ .

Par conséquent, la fonction d'autocorrélation partielle d'un processus mixte est infinie. Elle se comporte finalement comme la fonction d'autocorrélation partielle d'un processus moyenne mobile pure, dominée par un mélange d'exponentielles amorties et/ou d'ondes sinusoïdales amorties, en fonction de l'ordre la partie moyenne mobile et des valeurs des paramètres qu'elle contient.

### 3.4.3 Processus autorégressif du premier ordre-moyenne mobile du premier ordre

Un des processus ARMA mixtes d'une importance pratique considérable est le processus  $ARMA(1, 1)$

$$X_t - \phi_1 X_{t-1} = \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad (3.76)$$

ou bien

$$(1 - \phi_1 B)X_t = (1 - \theta_1 B)\varepsilon_t$$

Nous allons étudier maintenant certaines de ses propriétés les plus importantes.

#### Conditions de stationnarité et d'inversibilité

Tout d'abord, notons que le processus est stationnaire si  $-1 < \phi_1 < 1$ , et est inversible si  $-1 < \theta_1 < 1$ . De plus, à partir des relations  $\psi_1 = \phi_1 - \theta_1$  et  $\psi_j = \phi_1 \psi_{j-1}$  pour  $j \geq 2$ , on remarque que les poids  $\psi_j$  sont donnés par  $\psi_j = (\phi_1 - \theta_1) \phi_1^{j-1}$ ,  $j \geq 1$ . De même il est facile de voir que  $\pi_j = (\phi_1 - \theta_1) \theta_1^{j-1}$ ,  $j \geq 1$ , pour le processus stationnaire et inversible  $ARMA(1, 1)$ .

#### Fonction d'autocorrélation

A partir des équation (3.72) et (3.74) on obtient

$$\begin{aligned} \gamma_0 &= \phi_1 \gamma_1 + \sigma_\varepsilon^2 (1 - \theta_1 \psi_1) \\ \gamma_1 &= \phi_1 \gamma_0 - \theta_1 \sigma_\varepsilon^2 \\ \gamma_k &= \phi_1 \gamma_{k-1} \quad k \geq 2 \end{aligned}$$

avec  $\psi_1 = \phi_1 - \theta_1$ .

Par conséquent, en résolvant les deux premières équations pour  $\gamma_0$  et  $\gamma_1$ , la fonction d'autocovariance du processus est alors

$$\begin{aligned}\gamma_0 &= \frac{1+\theta_1^2-2\phi_1\theta_1}{1-\phi_1^2}\sigma_\varepsilon^2 \\ \gamma_1 &= \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1-\phi_1^2}\sigma_\varepsilon^2 \\ \gamma_k &= \phi_1\gamma_{k-1} \quad k \geq 2\end{aligned}\tag{3.77}$$

La dernière équation donne  $\gamma_k = \phi_1\gamma_{k-1}$ ,  $k \geq 2$ , de sorte que  $\rho_k = \rho_1\phi_1^{k-1}$ ,  $k \geq 2$ . Ainsi, la fonction d'autocorrélation décroît exponentiellement à partir de la valeur initiale  $\rho_1$ , qui dépend de  $\theta_1$  et de  $\phi_1$ . Cette décroissance exponentielle est lisse si  $\phi_1$  est positif et alterne en signe si  $\phi_1$  est négatif. De plus, le signe de  $\rho_1$  est déterminé par le signe de  $(\phi_1 - \theta_1)$  et indique de quel côté du zéro la décroissance exponentielle a lieu.

Les deux premières autocorrélations peuvent être exprimées en fonction des paramètres du processus  $ARMA(1,1)$ , tels que :

$$\begin{aligned}\rho_1 &= \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1+\theta_1^2-2\phi_1\theta_1} \\ \rho_2 &= \phi_1\rho_1\end{aligned}\tag{3.78}$$

En utilisant ces expressions et les conditions de stationnarité et d'inversibilité, on peut montrer que  $\rho_1$  et  $\rho_2$  doivent se trouver dans la région

$$\begin{aligned}|\rho_2| &< |\rho_1| \\ \rho_2 &> \rho_1(2\rho_1 + 1) \quad \rho_1 < 0 \\ \rho_2 &> \rho_1(2\rho_1 - 1) \quad \rho_1 > 0\end{aligned}\tag{3.79}$$

### Fonction d'autocorrélation partielle

La fonction d'autocorrélation partielle du processus mixte  $ARMA(1,1)$  est constituée d'une seule valeur initiale  $\phi_{11} = \rho_1$ . Ensuite, elle se comporte comme la fonction d'autocorrélation partielle d'un processus  $MA(1)$  pur et est dominée par une exponentielle amortie. Ainsi, comme on peut le montrer graphiquement, lorsque  $\theta_1$  est positif, elle est dominée par une exponentielle uniformément amortie qui décroît à partir de la valeur  $\rho_1$ , avec un signe déterminé par le signe de  $(\phi_1 - \theta_1)$ . De même, quand  $\theta_1$  est négatif, elle est dominée par une exponentielle qui oscille en tendant vers zéro à partir de la valeur  $\rho_1$ , avec un signe déterminé par le signe de  $(\phi_1 - \theta_1)$ .

### Exemple numérique.

Pour une illustration numérique, considérons le processus  $ARMA(1, 1)$  suivant

$$(1 - 0,8B)X_t = (1 + 0,6B)\varepsilon_t$$

de sorte que  $\phi = 0,8$  et  $\theta = -0,6$ .

En supposant  $\sigma_\varepsilon^2 = 1$ , on peut déduire à partir des équations (3.77) et (3.78) que la variance de  $X_t$  est  $\gamma_0 = 6,444$ , et  $\rho_1 = 0,893$ . En outre, la fonction d'autocorrélation satisfait  $\rho_j = 0,8\rho_{j-1}$ ,  $j \geq 2$ , de sorte que  $\rho_j = 0,893(0,8)^{j-1}$ , pour  $j \geq 2$ .

La décroissance exponentielle dans la fonction d'autocorrélation est clairement évidente. La fonction d'autocorrélation partielle exhibera également une allure à décroissance exponentielle qui oscille dans le signe en raison de la valeur négative de  $\theta$ .

# Chapitre 4

## Modèles linéaires non stationnaires

De nombreuses séries chronologiques (par exemple, les séries des cours de la bourse) se comportent comme si elles n'avaient pas de moyenne fixe. Même ainsi, elles présentent une homogénéité dans le sens où, en dehors du niveau local, ou peut-être du niveau et de la tendance locale, une partie de la série se comporte comme n'importe quelle autre partie. Les modèles qui décrivent un tel comportement non stationnaire homogène peuvent être obtenus en supposant qu'une *différenciation* appropriée du processus est stationnaire. Dans ce chapitre, nous aborderont les propriétés de la classe importante des modèles pour lesquels la  $d^{\text{ème}}$  différence de la série procure un processus stationnaire mixte autorégressif-moyenne mobile. Ces modèles sont appelés processus autorégressif-moyenne mobile intégré (Autoregressive-Integrated-Moving Average) ou bien *ARIMA*.

### 4.1 Processus autorégressif-moyenne mobile intégré

#### 4.1.1 Processus autorégressif de premier ordre non stationnaire

Il existe un nombre illimité de façons dont un processus peut être non stationnaire. Cependant, les types de séries économiques et industrielles que nous souhaitons analyser présentent fréquemment un type particulier de

comportement non stationnaire homogène qui peut être représenté par un modèle stochastique, qui est une forme modifiée du modèle autorégressif-moyenne mobile (ARMA). Prédemment, nous avons considéré le modèle mixte ARMA

$$\phi(B) X_t = \theta(B) \varepsilon_t \quad (4.1)$$

avec  $\phi(B)$  et  $\theta(B)$  des opérateurs polynomiaux en  $B$ , de degré  $p$  et  $q$ , respectivement.

Pour assurer la stationnarité, les racines de  $\phi(B) = 0$  doivent se situer en dehors du cercle unité. Un moyen naturel d'obtenir des processus non stationnaires consiste à assouplir cette restriction.

Pour avoir un aperçu des possibilités, considérons le modèle autorégressif de premier ordre,

$$(1 - \phi B) X_t = \varepsilon_t \quad (4.2)$$

qui est stationnaire pour  $|\phi| < 1$ .

Si on étudie le comportement de ce type de processus pour  $\phi = 2$ , une valeur en dehors de la gamme stationnaire, on voit qu'après une courte période d'induction, la série 'se déchaîne' et suit essentiellement une courbe exponentielle, les générateurs  $\varepsilon_t$  ne jouent presque aucun rôle. Le comportement des séries générées par des processus d'ordre supérieur, qui violent la condition de stationnarité, est similaire. De plus, ce comportement est essentiellement le même, que les termes 'moyenne mobile' soient ou non introduits à droite de l'équation du modèle.

#### 4.1.2 Modèle général pour un processus non stationnaire présentant une homogénéité

##### Modèle autorégressif-moyen mobile intégré

Bien que les modèles non stationnaires du type décrit ci-dessus soient utiles pour représenter un comportement explosif ou évolutif (comme la croissance bactérienne), les applications que nous décrivons dans ce cours ne sont pas de ce type. Jusqu'ici, nous avons vu qu'un processus ARMA est stationnaire si les racines de  $\phi(B) = 0$  se trouvent à l'extérieur du cercle unité, et montre un comportement non stationnaire explosif si les racines se trouvent à l'intérieur du cercle unité. Le seul cas restant est que les racines de  $\phi(B) = 0$  se trouvent sur le cercle de l'unité. Il s'avère que les modèles résultants sont d'une grande utilité pour représenter des séries chronologiques homogènes

non stationnaires. En particulier, les séries non saisonnières sont souvent bien représentées par des modèles dans lesquels une ou plusieurs de ces racines sont l'unité.

Considérons le modèle

$$\varphi(B) X_t = \theta(B) \varepsilon_t \quad (4.3)$$

où  $\varphi(B)$  est un opérateur autorégressif non stationnaire tel que  $d$  des racines de  $\varphi(B) = 0$  sont égales à l'unité et le reste se trouve en dehors du cercle unité.

Par suite, le modèle peut être écrit tel que

$$\varphi(B) X_t = \phi(B) (1 - \phi B)^d X_t = \theta(B) \varepsilon_t \quad (4.4)$$

où  $\phi(B)$  est un opérateur autorégressif stationnaire.

En utilisant l'opérateur de différenciation  $\nabla = 1 - B$ , nous pouvons écrire le modèle tel que

$$\phi(B) \nabla^d X_t = \theta(B) \varepsilon_t \quad (4.5)$$

De manière équivalente, le processus est défini par les deux équations

$$\phi(B) W_t = \theta(B) \varepsilon_t \quad (4.6)$$

et

$$W_t = \nabla^d X_t \quad (4.7)$$

Ainsi, nous constatons que le modèle correspond au fait que la  $d^{\text{ème}}$  différence de la série peut être représentée par un processus *ARMA* stationnaire et inversible. Une autre façon de voir le processus pour  $d \geq 1$  résulte de l'inversion de l'équation (4.7)

$$X_t = S^d W_t \quad (4.8)$$

où  $S$  est l'opérateur de sommation infinie défini tel que

$$\begin{aligned} Sx_t &= \sum_{h=-\infty}^t x_h = (1 + B + B^2 + \dots) x_t \\ &= (1 - B)^{-1} x_t = \nabla^{-1} x_t \end{aligned}$$

Ainsi,

$$S = (1 - B)^{-1} = \nabla^{-1}$$



L'opérateur  $S^2$  est similairement défini tel que

$$\begin{aligned} S^2 x_t &= Sx_t + Sx_{t-1} + Sx_{t-2} + \dots \\ &= \sum_{i=-\infty}^t \sum_{h=-\infty}^i x_h = (1 + 2B + 3B^2 + \dots) x_t \end{aligned}$$

et ainsi de suite pour l'ordre supérieur  $d$ .

L'équation (4.8) implique que le processus (4.5) peut être obtenu en additionnant (ou en 'intégrant') le processus stationnaire (4.6)  $d$  fois. Par conséquent, nous appelons le processus (4.5) un processus de autorégressif-moyenne mobile intégré (*ARIMA*).

Les modèles *ARIMA* pour les séries chronologiques non stationnaires, qui ont également été considérés plus tôt par Yaglom (1955), sont d'une importance fondamentale pour la prévision et le contrôle comme discuté par Box et Jenkins (1962, 1963, 1965, 1968a, 1968b, 1969) et Box et al. . (1967a). Zadeh et Ragazzini (1950), Kalman (1960) et Kalman et Bucy (1961) discutent également des processus non stationnaires. Une procédure plus ancienne pour l'analyse des séries chronologiques qui utilisaient la différenciation fut la méthode des '*differences de variables*' (voir Tintner (1940) et Rao et Tintner (1963)). Cependant, la motivation, les méthodes et les objectifs de cette procédure étaient très différents de ceux discutés ici. Techniquement, l'opérateur de sommation *infinie*  $S = (1 - B)^{-1}$  dans (4.8) ne peut pas vraiment être utilisé pour définir les processus *ARIMA* non stationnaires, du fait que les sommes infinies impliquées ne sont pas convergentes. Au lieu de cela, nous pouvons considérer l'opérateur de sommation fini  $S_m$  pour tout entier positif  $m$ , donné par

$$S_m = (1 + B + B^2 + \dots + B^{m-1}) \equiv \frac{1 - B^m}{1 - B}$$

De même, l'opérateur de sommation double fini peut être défini tel que

$$\begin{aligned} S_m^{(2)} &= \sum_{j=0}^{m-1} \sum_{i=j}^{m-1} B^i = (1 + 2B + 3B^2 + \dots + mB^{m-1}) \\ &\equiv \frac{1 - B^m - mB^m(1 - B)}{(1 - B)^2} \end{aligned}$$

car  $(1 - B) S_m^{(2)} = S_m - mB^m$ , et ainsi de suite.

Alors la relation entre un processus *ARMA* intégré  $X_t$  avec  $d = 1$ , par

exemple, et le processus *ARMA* correspondant  $W_t = (1 - B) X_t$ , en termes de valeurs retournées à une origine plus ancienne  $k < t$ , peut être exprimée telle que

$$X_t = \frac{S_{t-k}}{1 - B^{t-k}} W_t = \frac{1}{1 - B^{t-k}} (W_t + W_{t-1} + \dots + W_{k+1})$$

de sorte que  $X_t = W_t + W_{t-1} + \dots + W_{k+1} + X_k$  peut être considéré comme la somme d'un nombre fini de termes du processus stationnaire  $W$  plus une valeur d'initialisation du processus  $X$  à l'instant  $k$ .

Par conséquent, dans la définition formelle des propriétés stochastiques d'un processus *ARIMA* non-stationnaire tel que généré dans l'équation (4.3), il serait généralement nécessaire de spécifier des conditions d'initialisation pour le processus à un certain point  $k$  dans un passé fini (mais éventuellement lointain). Cependant, ces spécifications de la condition initiale auront peu d'effet sur la plupart des caractéristiques importantes du processus, et de telles spécifications ne seront pour la plupart pas en considération dans ce cours. Comme mentionné dans le chapitre 1, le modèle (4.5) équivaut à représenter le processus  $X_t$  comme la sortie d'un filtre linéaire (sauf si  $d = 0$ , c'est un filtre linéaire instable), dont l'entrée est le bruit blanc  $\varepsilon_t$ . Alternativement, on peut le considérer comme un dispositif *pour transformer le processus hautement dépendant, et éventuellement non stationnaire  $X_t$ , en une suite de variables aléatoires non corrélées  $\varepsilon_t$* , c'est-à-dire, transformer le processus en un bruit blanc.

Si dans (4.5), l'opérateur autorégressif  $\phi(B)$  est d'ordre  $p$ , la  $d^{\text{ème}}$  différence est prise, et l'opérateur de moyenne mobile  $\phi(B)$  est d'ordre  $q$ , on dit qu'on a un modèle *ARIMA* d'ordre  $(p, d, q)$ , ou simplement un processus *ARIMA*( $p, d, q$ ).

## Deux interprétations du modèle ARIMA

On montre maintenant que le modèle *ARIMA* est un modèle intuitivement raisonnable pour de nombreuses séries chronologiques qui se produisent dans la pratique. Premièrement, on notera que le comportement local d'une série chronologique stationnaire dépend fortement du niveau de  $X_t$ .

Si nous devons utiliser des modèles pour lesquels le comportement du processus est indépendant de son niveau, nous devons choisir l'opérateur autorégressif  $\varphi(B)$  tel que

$$\varphi(B)(X_t + c) = \varphi(B)X_t$$

où  $c$  est une constante quelconque.

Ainsi  $\varphi(B)$  doit être de la forme

$$\varphi(B) = \phi_1(B)(1 - B) = \phi_1(B)\nabla$$

Par conséquent, une classe de processus ayant la propriété désirée sera de la forme

$$\phi_1(B)W_t = \theta(B)\varepsilon_t$$

où  $W_t = \nabla X_t$ .

L'homogénéité requise exclut la possibilité que  $W_t$  augmente de façon explosive. Cela signifie que  $\phi_1(B)$  est un opérateur autorégressif stationnaire ou  $\phi_1(B) = \phi_2(B)(1 - B)$ , de telle sorte que  $\phi_2(B)W_t = \theta(B)\varepsilon_t$ , où maintenant  $W_t = \nabla^2 X_t$ . Dans ce dernier cas, le même argument peut être appliqué à la seconde différence, et ainsi de suite.

Finalement, nous arrivons à la conclusion que pour la représentation des séries chronologiques qui sont non stationnaires mais qui présentent néanmoins une homogénéité, l'opérateur à gauche de (4.3) devrait être de la forme  $\phi(B)\nabla^d$ , où  $\phi(B)$  est un opérateur autorégressif stationnaire. Ainsi, nous sommes ramenés au modèle (4.5).

Pour approcher le modèle d'un point de vue quelque peu différent, considérons la situation où  $d = 0$  dans (4.4), de telle sorte que  $\phi(B)X_t = \theta(B)\varepsilon_t$ . L'exigence que les zéros de  $\phi(B)$  soient en dehors du cercle unité assurerait non seulement que le processus  $X_t$  soit stationnaire et de moyenne nulle, mais aussi que  $\nabla X_t$ ,  $\nabla^2 X_t$ ,  $\nabla^3 X_t$ , ... soient chacun stationnaire et de moyenne nulle.

Une série chronologique homogène sauf en son niveau, en ce sens qu'à l'exception d'une translation verticale, une partie ressemble beaucoup à une autre, peut être représentée en conservant l'exigence que chacune des différences soit stationnaire et de moyenne nulle, mais en laissant le niveau 'libre'. Nous faisons cela en utilisant le modèle

$$\phi(B)\nabla X_t = \theta(B)\varepsilon_t$$

Une série chronologique qui n'a ni un niveau fixe ni une pente fixe, mais dont comportement est homogène si l'on tient compte des différences de ces caractéristiques, peut-être représentée par le modèle

$$\phi(B)\nabla^2 X_t = \theta(B)\varepsilon_t$$

ce qui assure la stationnarité et une moyenne nulle pour toutes les différences après la première et la seconde mais permet au niveau et à la pente d'être 'libre'.

### 4.1.3 Forme générale d'un modèle ARIMA

Pour des raisons que l'on développera plus loin, il est parfois utile de considérer une légère extension du modèle *ARIMA* dans (4.5), en ajoutant un terme constant  $\theta_0$ , donnant la forme plus générale suivante

$$\varphi(B) X_t = \phi(B) \nabla^d X_t = \theta_0 + \theta(B) \varepsilon_t \quad (4.9)$$

où

$$\begin{aligned} \phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \\ \theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \end{aligned}$$

1.  $\phi(B)$  est appelé *opérateur autorégressif* ; il est supposé être stationnaire, c'est-à-dire que les racines de l'équation  $\phi(B) = 0$  se trouvent à l'extérieur du cercle unité.
2.  $\varphi(B) = \phi(B) \nabla^d$  est appelé *opérateur autorégressif généralisé* ; c'est un opérateur non stationnaire avec  $d$  racines de  $\varphi(B) = 0$  égales à l'unité.
3.  $\theta(B)$  est appelé *opérateur moyenne mobile* ; il est supposé être inversible, c'est-à-dire que les racines de  $\theta(B) = 0$  se trouvent à l'extérieur du cercle unité.

Lorsque  $d = 0$ , ce modèle représente un processus stationnaire. Les exigences de stationnarité et d'inversion s'appliquent indépendamment, et, en général, les opérateurs  $\phi(B)$  et  $\theta(B)$  ne sont pas du même ordre. Des exemples de régions de stationnarité pour les cas simples de  $p = 1, 2$  et les régions d'inversion identiques pour  $q = 1, 2$  ont été donnés au chapitre 3.

### Tendances stochastiques et déterministes

Lorsque le terme constant  $\theta_0$  est omis, le modèle (4.9) est capable de représenter des séries qui ont des tendances stochastiques, caractérisées, par exemple, par des changements aléatoires du niveau et de la pente de la série. Cependant, nous pouvons souhaiter inclure une fonction déterministe du temps  $f(t)$  dans le modèle. En particulier, une acceptation automatique

pour une tendance polynomiale déterministe, de degré  $d$ , peut être tolérée en permettant à  $\theta_0$  d'être différent de zéro. Par exemple, lorsque  $d = 1$ , on peut utiliser le modèle avec  $\theta_0 \neq 0$  pour représenter une possible tendance linéaire déterministe en présence de bruit non stationnaire. Puisque autoriser  $\theta_0$  à être différent de zéro équivaut à permettre à

$$E(W_t) = E(\nabla^d X_t) = \mu_W = \frac{\theta_0}{1 - \phi_1 - \phi_2 - \dots - \phi_p} \quad (4.10)$$

d'être non nulle.

Une autre façon d'exprimer le modèle (4.9) plus généralement, consiste à l'exprimer sous la forme d'un processus *ARMA* inversible stationnaire en terme de  $\widetilde{W}_t = W_t - \mu_W$ , c'est-à-dire

$$\varphi(B) \widetilde{W}_t = \theta(B) \varepsilon_t \quad (4.11)$$

Remarquons que, quand  $d = 1$  par exemple,  $\nabla X_t = W_t = \widetilde{W}_t + \mu_W$  implique que  $X_t = \widetilde{X}_t + \mu_W t + \alpha$ , où  $\alpha$  est une constante et le processus  $\widetilde{X}_t$  est tel que  $\nabla \widetilde{X}_t = \widetilde{W}_t$ , qui est de moyenne nulle. Ainsi,  $\theta_0 \neq 0$  autorise une composante de tendance linéaire déterministe dans  $X_t$  avec une pente  $\mu_W = \frac{\theta_0}{1 - \phi_1 - \phi_2 - \dots - \phi_p}$ .

Dans de nombreuses applications, où il n'y a pas de raison physique pour une composante déterministe, on peut supposer que la moyenne de  $W$  est nulle à moins qu'une telle hypothèse ne soit incompatible avec les données. Dans de nombreux cas, l'hypothèse d'une tendance stochastique est plus réaliste que l'hypothèse d'une tendance déterministe. Ceci est d'une importance particulière dans la prévision, car une tendance stochastique n'exige pas que la série suive la tendance observée dans le passé. Dans ce qui suit, lorsque  $d > 0$ , nous supposerons souvent que  $\mu_W = 0$ , ou de façon équivalente, que  $\theta_0 = 0$ , à moins que cela ne soit clair à partir des données ou de la nature du problème qu'une moyenne non nulle, ou plus généralement une composante déterministe de forme connue, est nécessaire.

### Quelques cas particuliers importants du modèle ARIMA

Un peu plus haut, nous avons examiné quelques cas particuliers importants du modèle (4.9), correspondant à la situation stationnaire,  $d = 0$ . Les modèles suivants représentent quelques cas particuliers du modèle non stationnaire ( $d \geq 1$ ), qui semblent communs dans la pratique.

1. Le processus  $(0, 1, 1)$  :

$$\begin{aligned}\nabla X_t &= \varepsilon_t - \theta_1 \varepsilon_{t-1} \\ &= (1 - \theta_1 B) \varepsilon_t\end{aligned}$$

correspondant à  $p = 0, d = 1, q = 1, \phi(B) = 1, \theta(B) = 1 - \theta_1 B$ .

2. Le processus  $(0, 2, 2)$  :

$$\begin{aligned}\nabla^2 X_t &= \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \\ &= (1 - \theta_1 B - \theta_2 B^2) \varepsilon_t\end{aligned}$$

correspondant à  $p = 0, d = 2, q = 2, \phi(B) = 1, \theta(B) = 1 - \theta_1 B - \theta_2 B^2$ .

3. Le processus  $(1, 1, 1)$  :

$$\nabla X_t - \phi_1 \nabla X_{t-1} = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

ou bien

$$(1 - \phi_1 B) \nabla X_t = (1 - \theta_1 B) \varepsilon_t$$

correspondant à  $p = 1, d = 1, q = 1, \phi(B) = 1 - \phi_1 B, \theta(B) = 1 - \theta_1 B$ .

Pour la représentation des séries chronologiques non saisonnières nous avons rarement rencontré des situations pour lesquelles  $p, d$  ou  $q$  ont été supérieurs à 2. Souvent, les valeurs 0 ou 1 sont les plus souvent rencontrées dans la pratique.

### Transformation non linéaire de du processus $X_t$

La gamme d'applications utiles du modèle (4.9) s'élargit considérablement si l'on admet la possibilité de transformation. Ainsi, nous pouvons substituer  $X_t^{(\lambda)}$  pour  $X_t$ , dans (4.9), où  $X_t^{(\lambda)}$  est une transformation nonlinéaire de  $X_t$ , impliquant un ou plusieurs paramètres  $\lambda$ . Une transformation appropriée peut être suggérée par l'application ou, dans certains cas, elle peut être estimée à partir des données. Par exemple, si nous nous intéressons aux ventes d'un produit récemment introduit, nous pourrions constater que le volume des ventes augmentait rapidement et que c'était la fluctuation en pourcentage qui présentait une stabilité non stationnaire (homogénéité) plutôt qu'une fluctuation absolue. Cela permettrait d'analyser le logarithme des ventes puisque

$$\nabla \log(X_t) = \log\left(\frac{X_t}{X_{t-1}}\right) = \log\left(1 + \frac{\nabla X_t}{X_{t-1}}\right) \simeq \frac{\nabla X_t}{X_{t-1}}$$

où  $\frac{\nabla X_t}{X_{t-1}}$  représente le changement relatif ou en pourcentage.

L'approximation étant maintenue si les changements relatifs ne sont pas excessivement grands. Lorsque les données couvrent une large gamme et surtout les données saisonnières, l'estimation de la transformation en utilisant l'approche de Box et Cox (1964) peut être utile. Cette approche considère la famille des transformations de puissance de la forme  $X_t^{(\lambda)} = \frac{X_t^\lambda - 1}{\lambda}$  pour  $\lambda \neq 0$  et  $X_t^{(0)} = \log(X_t)$  pour  $\lambda = 0$ .

Le logiciel permettant d'estimer le paramètre  $\lambda$  dans la transformation de puissance de Box-Cox est disponible dans les bibliothèques *TSA* et *MASS* de R. Par exemple, la fonction '*BoxCox.ar()*' dans le paquet *TSA* donne une transformation de puissance de sorte que la série transformée soit approximativement un processus AR gaussien.

## 4.2 Deux formes explicites pour le modèle ARIMA

Considérons maintenant trois formes 'explicites' différentes pour le modèle général (4.9). Chacune d'entre elles permet de valoriser un aspect particulier. Ainsi, la valeur actuelle  $X_t$  du processus peut être exprimée

1. En termes de valeurs passées de  $X$  et des valeurs actuelle et passées des innovations  $\varepsilon$ , en utilisant directement l'équation de différence,
2. En termes des innovations actuelle et passées  $\varepsilon_{t-j}$  seulement, et
3. En termes de somme pondérée des valeurs passées  $X_{t-j}$  du processus et de l'innovation actuelle  $\varepsilon_t$ .

Dans ce qui suit, nous nous intéressons principalement aux modèles *non stationnaires* dans lesquels  $\nabla^d X_t$  est un processus stationnaire et  $d$  est supérieur à zéro. Pour de tels modèles, nous pouvons, sans perte de généralité, omettre  $\mu$  dans la spécification.

### 4.2.1 Modèle sous forme d'équation aux différences

L'utilisation directe de l'équation de différence nous permet d'exprimer la valeur actuelle  $X_t$  du processus en fonction des valeurs précédentes des  $X$  et des valeurs actuelle et antérieures des innovations  $\varepsilon$ . Ainsi, si

$$\varphi(B) = \phi(B)(1 - B)^d = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_{p+d} B^{p+d}$$

le modèle général (4.9), avec  $\theta_0 = 0$ , peut être écrit tel que

$$X_t = \varphi_1 X_{t-1} + \dots + \varphi_{p+d} X_{t-p-d} - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (4.12)$$

Par exemple, considérons le processus représenté par le modèle d'ordre  $(1, 1, 1)$

$$(1 - \phi B)(1 - B)X_t = (1 - \theta B)\varepsilon_t$$

Alors ce processus peut être exprimé tel que

$$[1 - (1 + \phi)B + \phi B^2]X_t = (1 - \theta B)\varepsilon_t$$

c'est-à-dire,

$$X_t = (1 + \phi)X_{t-1} - \phi X_{t-2} + \varepsilon_t - \theta \varepsilon_{t-1} \quad (4.13)$$

avec  $\varphi_1 = 1 + \phi$  et  $\varphi_2 = -\phi$  dans la notation introduite ci-dessus.

Pour de nombreuses raisons, et en particulier pour le calcul des prévisions, l'équation de différence (4.12) est la forme la plus pratique à utiliser.

## 4.2.2 Modèle sous forme d'innovations aléatoires

### Modèle sous forme d'innovations présente et passées

Comme il a été établi précédemment, un modèle linéaire peut être exprimé comme la sortie  $X_t$  du filtre linéaire

$$\begin{aligned} X_t &= \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots \\ &= \varepsilon_t + \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j} = \psi(B) \varepsilon_t \end{aligned} \quad (4.14)$$

dont l'entrée est un bruit blanc, ou une suite de variables aléatoires non corrélées  $\varepsilon_t$  avec une moyenne 0 et une variance commune  $\sigma_\varepsilon^2$ .

Il est parfois utile d'exprimer le modèle *ARIMA* sous cette forme et, en particulier, les poids  $\psi$  seront nécessaires au chapitre 5 pour calculer la variance des erreurs de prévision. Cependant, comme les processus *ARIMA* non stationnaires ne sont pas en équilibre statistique au cours du temps, ils ne peuvent pas être supposés se propager indéfiniment dans le passé, et donc une représentation infinie comme dans (4.14) ne sera pas possible. Mais une forme tronquée finie approximative, dont il sera question plus loin, existe toujours. Nous montrons maintenant que les poids  $\psi$  pour un processus *ARIMA* peuvent être obtenus directement à partir de la forme 'équation de différence' du modèle.



### Expression générale pour les poids $\psi$

Si on multiplie des deux côtés de (??) avec l'opérateur autorégressif généralisé  $\varphi(B)$ , on obtient

$$\varphi(B) X_t = \varphi(B) \psi(B) \varepsilon_t$$

Cependant, puisque  $\varphi(B) X_t = \theta(B) \varepsilon_t$ , il s'ensuit que

$$\varphi(B) \psi(B) = \theta(B) \quad (4.15)$$

Par conséquent, les poids  $\psi$  peuvent être obtenus en égalisant les coefficients de  $B$  dans le développement suivant

$$(1 - \varphi_1 B - \dots - \varphi_{p+d} B^{p+d}) (1 + \psi_1 B + \psi_2 B^2 + \dots) = 1 - \theta_1 B - \dots - \theta_q B^q \quad (4.16)$$

Par ailleurs, on remarque que les poids  $\psi_j$  du processus *ARIMA* peuvent être déterminés récursivement à l'aide des équations

$$\psi_j = \varphi_1 \psi_{j-1} + \varphi_2 \psi_{j-2} + \dots + \varphi_{p+d} \psi_{j-p-d} - \theta_j \quad j > 0$$

avec  $\psi_0 = 1$ ,  $\psi_j = 0$  pour  $j < 0$ , et  $\theta_j = 0$  pour  $j > q$ .

On peut noter aussi que pour  $j > \sup(p + d - 1, q)$ , les poids  $\psi$  satisfont l'équation de différence homogène définie par l'opérateur autorégressif généralisé, c'est-à-dire

$$\varphi(B) \psi_j = \phi(B) (1 - B)^d \psi_j = 0 \quad (4.17)$$

où  $B$  opère maintenant sur l'indice  $j$ .

Ainsi, pour  $j$  suffisamment grand, les poids  $\psi_j$  sont représentés par un mélange de polynômes en  $j$ , d'exponentielles amorties et de sinusoides amorties.

### Exemple

A titre d'exemple, considérons le processus  $(1, 1, 1)$  donné par l'équation (4.13), pour lequel

$$\begin{aligned} \varphi(B) &= (1 - \phi B) (1 - B) \\ &= 1 - (1 + \phi) B + \phi B^2 \end{aligned}$$

et

$$\theta(B) = 1 - \theta B$$

En substituant dans (4.16) on obtient

$$(1 - (1 + \phi)B + \phi B^2) (1 + \psi_1 B + \psi_2 B^2 + \dots) = 1 - \theta B$$

et donc les  $\psi_j$  satisfont à l'équation de récurrence  $\psi_j = (1 + \phi) \psi_{j-1} - \phi \psi_{j-2}$ ,  $j \geq 2$  avec  $\psi_0 = 1$  et  $\psi_1 = (1 + \phi) - \theta$ . Ainsi, puisque les racines de l'équation  $\varphi(B) = (1 - \phi B)(1 - B) = 0$  sont  $G_1^{-1} = 1$  et  $G_2^{-1} = \phi^{-1}$ , on a en général,

$$\psi_j = A_0 + A_1 \phi^j \quad (4.18)$$

où les constantes  $A_0$  et  $A_1$  sont déterminées à partir des valeurs initiales  $\psi_0 = A_0 + A_1$  et  $\psi_1 = A_0 + A_1 \phi = 1 + \phi - \theta$  tel que

$$A_0 = \frac{1-\theta}{1-\phi} \quad A_1 = \frac{\theta-\phi}{1-\phi}$$

Ainsi, de manière informelle, on peut vouloir exprimer le modèle (4.13) sous une forme équivalente

$$X_t = \sum_{j=0}^{\infty} (A_0 + A_1 \phi^j) \varepsilon_{t-j} \quad (4.19)$$

Puisque  $|\phi| < 1$ , les poids  $\psi_j$  tendent vers  $A_0$  pour  $j$  assez grand, de sorte que les innovations  $\varepsilon_{t-j}$ , qui sont assez éloignés dans le passé, reçoivent un poids constant  $A_0$ . Cependant, la représentation en (4.19) n'est strictement pas valable parce que la somme infinie sur la droite ne converge en aucun sens ; c'est-à-dire que les poids  $\psi_j$  ne sont pas absolument sommables comme dans le cas d'un processus stationnaire. Une version tronquée connexe de la forme en somme d'innovations du modèle est toujours valable, comme on le verra plus loin. Néanmoins, pour des commodités de notation, on se référera souvent à la forme en somme infinie d'innovations (4.14) d'un processus *ARIMA* même si cette forme n'est strictement pas convergente, comme un simple dispositif de notation pour représenter la forme tronquée dans (4.25), dans des situations où la distinction entre les deux formes n'est pas importante.

### Forme tronquée du modèle sous forme d'innovations aléatoires

Pour des raisons techniques, il est nécessaire, et dans certains cas pratique, de considérer le modèle sous une forme légèrement différente de (4.14). Supposons que nous souhaitons exprimer la valeur actuelle  $X_t$  du processus

en fonction des  $t - k$  innovations  $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{k+1}$ , assimilées par le système depuis l'instant origine  $k < t$ . Cette instant d'origine  $k$  peut, par exemple, être l'instant dans lequel le processus a été observé pour la première fois. Le modèle général

$$\varphi(B) X_t = \theta(B) \varepsilon_t \quad (4.20)$$

est une équation aux différences ayant pour solution

$$X_t = C_k(t - k) + I_k(t - k) \quad (4.21)$$

Un bref résumé sur les équations aux différences linéaires est donné à l'annexe A4.1.

Nous rappelons au lecteur que la solution de telles équations est très proche de celle des équations différentielles linéaires. La *fonction complémentaire*  $C_k(t - k)$  est la solution générale de l'équation aux différences homogène

$$\varphi(B) C_k(t - k) = 0 \quad (4.22)$$

En général, cette solution consiste en une combinaison linéaire de certaines fonctions du temps. Ces fonctions sont des puissances  $t^j$ , des termes géométriques réels (exponentiels)  $G^t$ , et des termes géométriques complexes (exponentiels)  $D^t \sin(2\pi f_0 t - F)$ , où les constantes  $G$ ,  $f_0$ , et  $F$  sont des fonctions des paramètres  $(\phi, \theta)$  du modèle. Les coefficients qui forment les combinaisons linéaires de ces termes peuvent être déterminés de manière à satisfaire un ensemble de conditions initiales définies par les valeurs du processus avant l'instant  $k + 1$ . L'*intégrale particulière*  $I_k(t - k)$  est une fonction qui satisfait

$$\varphi(B) I_k(t - k) = \theta(B) \varepsilon_t \quad (4.23)$$

Il faut noter que dans cette expression  $B$  opère sur  $t$  et non sur  $k$ . Il est montré dans l'annexe A4.1 que cette équation est satisfaite pour  $t - k > q$  par

$$I_k(t - k) = \sum_{j=0}^{t-k-1} \psi_j \varepsilon_{t-j} = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \dots + \psi_{t-k-1} \varepsilon_{k+1} \quad t > k \quad (4.24)$$

avec  $I_k(t - k) = 0, t \leq k$ .

L'intégrale particulière  $I_k(t - k)$ , représente ainsi la forme tronquée finie de la forme infinie en innovations (4.14), tandis que la fonction complémentaire  $C_k(t - k)$  résume les fonctions 'initialisantes' du processus  $X$  dans le sens où

$C_k(t-k)$  est déjà déterminé ou spécifié par l'instant  $k+1$ . Par conséquent, la forme tronquée du modèle en innovations pour le processus *ARIMA* (4.3) est donnée par

$$X_t = \sum_{j=0}^{t-k-1} \psi_j \varepsilon_{t-j} + C_k(t-k) \quad (4.25)$$

A titre d'exemple, considérons la figure 4.4. La discussion ci-dessus implique que toute observation  $X_t$  peut être considérée par rapport à tout instant antérieur  $k$  et peut être divisée en deux parties additives. La première partie  $C_k(t-k)$  est la composante de  $X_t$ , *déjà déterminée à l'instant  $k$* , et indique ce que les observations antérieures à l'instant  $k+1$  devaient nous parler de la valeur de la série à l'instant  $t$ . Elle représente la trajectoire que prendrait le processus si, à l'instant  $k$ , la source des innovations  $\varepsilon_t$  avait été 'éteinte'. La deuxième partie,  $I_k(t-k)$ , représente un composant supplémentaire, *imprévisible à l'instant  $k$* , qui représente l'effet total des innovations entrant dans le système à l'instant  $k$ . Par conséquent, pour spécifier un processus *ARIMA*, il faut spécifier le composant d'initialisation  $C_k(t-k)$  dans (4.25) pour une origine temporelle  $k$  dans le passé fini mais peut-être lointain, la trajectoire restante du processus étant déterminée par les termes du processus des innovations tronqué dans (4.25).

La fonction complémentaire est la solution de l'équation aux différences

$$(1 - \phi B)(1 - B)C_k(t-k) = 0$$

c'est-à-dire,

$$C_k(t-k) = b_0^{(k)} + b_1^{(k)} \phi^{t-k}$$

où  $b_0^{(k)}$ ,  $b_1^{(k)}$  sont des coefficients qui dépendent de l'historique du processus et, on le remarquera, changent avec l'origine  $k$ .

En utilisant les poids  $\psi$  (4.18), l'intégrale particulière (4.24) est telle que

$$I_k(t-k) = \sum_{j=0}^{t-k-j} (A_0 + A_1 \phi^j) \varepsilon_{t-j}$$

de sorte que, finalement, nous pouvons écrire le modèle (4.19) sous la forme équivalente suivante

$$X_t = b_0^{(k)} + b_1^{(k)} \phi^{t-k} + \sum_{j=0}^{t-k-j} (A_0 + A_1 \phi^j) \varepsilon_{t-j} \quad (4.26)$$

Notons que puisque  $|\phi| < 1$ , si  $t - k$  est choisi suffisamment grand, le terme impliquant  $\phi^{t-k}$  dans cette expression est négligeable et peut être ignoré.

### **Lien entre les formes tronquées et non-tronquées du modèle sous forme d'équation aux différences**

En revenant au cas général, on peut toujours considérer le processus en se référant à une origine finie  $k$  (éventuellement éloignée), le processus étant sous forme d'innovations aléatoires tronqué comme dans (4.25). Par comparaison avec la forme non tronquée de (4.14), on peut constater qu'on peut, de façon informelle, faire la correspondance en représentant la fonction complémentaire  $C_k(t - k)$  en fonction des poids  $\psi$  tel que

$$C_k(t - k) = \sum_{j=t-k}^{\infty} \psi_j \varepsilon_{t-j} \quad (4.27)$$

même si, formellement, la somme infinie à droite de l'équation (4.27) ne converge pas. Comme mentionné précédemment, pour la simplicité de notation, nous utiliserons souvent cette correspondance.

En résumé, pour le modèle général (4.20),

1. Nous pouvons exprimer la valeur  $X_t$  du processus, de manière informelle, comme une somme infinie pondérée des innovations présente et passées  $\varepsilon_{t-j}$ , en fonction de

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \psi(B) \varepsilon_t$$

2. La valeur de  $X_t$  peut être exprimée, plus formellement, comme une somme finie pondérée des  $t - k$  innovations présente et passées survenant après une origine  $k$ , plus une fonction complémentaire  $C_k(t - k)$ . Cette somme finie se compose des  $t - k$  premiers termes de la somme infinie, de sorte que

$$X_t = C_k(t - k) + \sum_{j=0}^{t-k-1} \psi_j \varepsilon_{t-j} \quad (4.28)$$

Enfin, la fonction complémentaire  $C_k(t - k)$  peut être considérée, par commodité de notation, comme étant la somme infinie tronquée, de

sorte que

$$C_k(t-k) = \sum_{j=t-k}^{\infty} \psi_j \varepsilon_{t-j}$$

Par exemple, considérons une fois de plus le modèle

$$(1 - \phi B)(1 - B)X_t = (1 - \theta B)\varepsilon_t$$

Nous pouvons écrire  $X_t$  soit, de manière informelle, comme une somme infinie de  $\varepsilon_{t-j}$

$$X_t = \sum_{j=0}^{\infty} (A_0 + A_1 \phi^j) \varepsilon_{t-j}$$

soit, plus formellement, sous forme d'une somme finie pondérée telle que

$$X_t = C_k(t-k) + \sum_{j=0}^{t-k-1} (A_0 + A_1 \phi^j) \varepsilon_{t-j}$$

De plus, la fonction complémentaire peut être écrite telle que

$$C_k(t-k) = b_0^{(k)} + b_1^{(k)} \phi^{t-k}$$

où  $b_0^{(k)}$  et  $b_1^{(k)}$ , qui satisfont aux conditions initiales à travers l'instant  $k$ , sont tels que

$$\begin{aligned} b_0^{(k)} &= \frac{X_k - \phi X_{k-1} - \theta \varepsilon_k}{1 - \phi} \\ b_1^{(k)} &= \frac{-\phi(X_k - X_{k-1}) + \theta \varepsilon_k}{1 - \phi} \end{aligned}$$

La fonction complémentaire peut également être représentée, de manière informelle, comme la somme infinie tronquée

$$C_k(t-k) = \sum_{j=t-k}^{\infty} (A_0 + A_1 \phi^j) \varepsilon_{t-j}$$

à partir de laquelle on peut remarquer que  $b_0^{(k)}$  et  $b_1^{(k)}$  peuvent être représentés tels que

$$\begin{aligned} b_0^{(k)} &= A_0 \sum_{j=t-k}^{\infty} \varepsilon_{t-j} = \frac{1 - \theta}{1 - \phi} \sum_{j=t-k}^{\infty} \varepsilon_{t-j} \\ b_1^{(k)} &= A_1 \sum_{j=t-k}^{\infty} \phi^{j-(t-k)} \varepsilon_{t-j} = \frac{\theta - \phi}{1 - \phi} \sum_{j=t-k}^{\infty} \phi^{j-(t-k)} \varepsilon_{t-j} \end{aligned}$$

## Fonction complémentaire comme une espérance conditionnelle

Une conséquence de la forme tronquée (4.25) est que pour  $m > 0$

$$C_k(t-k) = C_{k-m}(t-k+m) + \psi_{t-k}\varepsilon_k + \psi_{t-k+1}\varepsilon_{k-1} + \dots + \psi_{t-k+m-1}\varepsilon_{k-m+1} \quad (4.29)$$

qui montre comment la fonction complémentaire change lorsque l'origine  $k$  est modifiée.

Désignons maintenant par  $C_k(t-k)$  l'*espérance conditionnelle* de  $X_t$ , à l'*instant*  $k$ . C'est l'espérance connaissant l'historique complète du processus jusqu'à, mais pas au-delà de l'instant  $k$ . Pour calculer cette espérance, notons que

$$E_k[\varepsilon_j] = \begin{cases} 0 & j > k \\ \varepsilon_j & j \leq k \end{cases}$$

C'est-à-dire que, à l'instant  $k$ , les valeurs attendues des futures innovations  $\varepsilon$  sont nulles et de celles qui ont déjà eu lieu sont les valeurs effectivement réalisées.

En prenant des espérances conditionnelles à l'instant  $k$  des deux côtés de (4.28), on obtient  $E_k[X_t] = C_k(t-k)$ . Ainsi, pour  $(t-k) > q$ , la fonction complémentaire fournit la valeur attendue de la future valeur  $X_t$  du processus, *observé à partir de l'instant*  $k$  et basée sur la connaissance du passé. L'intégrale particulière montre comment cette espérance est modifiée par les événements suivants représentés par les innovations  $\varepsilon_{k+1}, \varepsilon_{k+2}, \dots, \varepsilon_t$ . Dans le problème de la prévision, que nous discuterons plus loin, il apparaîtra que  $C_k(t-k)$  est la prévision d'erreur quadratique moyenne minimale de  $X_t$  faite à l'instant  $k$ . L'équation (4.29) peut être utilisée pour "mettre à jour" cette prévision.

# Chapitre 5

## Modèles conditionnels hétéroscédastiques

Dans ce chapitre on présente un aperçu de certains modèles qui ont été élaborés pour décrire la variabilité dans le temps de la variance ou la volatilité d'une série chronologique. Pour introduire quelques notations, notons que le processus  $ARMA(p, p)$ ,  $\phi(B) X_t = \theta_0 + \theta(B) \varepsilon_t$ , peut être écrit comme la somme de sa prévision par rapport à son passé et d'une erreur de prédiction

$$X_t = E(X_t | \mathcal{F}_{t-1}) + \varepsilon_t$$

où  $\mathcal{F}_{t-1}$  représente l'information passée disponible à l'instant  $t - 1$  et  $\varepsilon_t$  représente l'erreur de prédiction. Pour le modèle  $ARMA$ ,  $\mathcal{F}_{t-1}$  est une fonction des observations passées et des termes d'erreurs passés, mais pourrait plus généralement inclure des variables de régression externes  $Y_t$ . L'hypothèse faite jusqu'ici est que les erreurs de prédiction  $\varepsilon_t$  sont des variables aléatoires *indépendantes* avec une variance constante  $Var(\varepsilon_t) = \sigma_\varepsilon^2$  qui est indépendante du passé. Cependant, cette hypothèse semble incompatible avec l'hétéroscédasticité souvent observée pour les séries chronologiques en économie et en finance, en particulier. Par exemple, les séries chronologiques financières telles que les rendements boursiers présentent souvent des périodes où la volatilité est élevée et des périodes où elle est plus faible. Cette caractéristique de l'évolution d'une série chronologique, ou *fait stylisé*, est communément appelée *phénomène de grappe* ou 'clustering' de la volatilité. Une autre caractéristique commune des séries chronologiques financières est que les distributions marginales sont leptokurtiques et tendent à avoir des queues plus lourdes que celles d'une distribution normale. Un certain nombre



d'autres faits stylisés ont été étudiés pour des données financières (pour des discussions et des références, voir, par exemple, Teräsvirta et al., 2010, chapitre 8).

Le modèle autorégressif hétéroscédastique conditionnel (*ARCH*) a été introduit par Engle (1982) pour décrire la variabilité dans le temps de la variance d'une série de taux d'inflation. Une extension de ce modèle appelé modèle conditionnel hétéroscédastique généralisé (*GARCH*) a été proposée par Bollerslev (1986). Ces modèles sont capables de décrire non seulement le regroupement de la volatilité, mais aussi des caractéristiques telles que le comportement des queues lourdes qui est commun dans de nombreuses séries chronologiques en économie et en finance. Cependant, il existe d'autres caractéristiques liées à la volatilité qui ne sont pas capturées par les modèles de base *ARCH* et *GARCH*. Cela a conduit à un certain nombre d'extensions et de formulations alternatives visant à résoudre ces problèmes. Cette section présente une brève description des modèles *ARCH* et *GARCH* ainsi que certaines extensions proposées dans la littérature. La littérature dans ce domaine est vaste et seul un certain nombre de développements seront discutés. Une couverture plus complète peut être trouvée dans les articles de Bollerslev et al. (1992, 1994), Bera et Higgins (1993), Li et al. (2003) et Teräsvirta (2009), entre autres. La modélisation de la volatilité est également discutée dans plusieurs publications sur les séries chronologiques, y compris Franses et van Dijk (2000), Mills et Markellos (2008), Teräsvirta et al. (2010) et Tsay (2010). Les manuels consacrés à la modélisation de la volatilité incluent Francq et Zakoïan (2010) et Xekalaki et Degiannakis (2010).

## 5.1 Le modèle *ARCH*

Pour un processus *ARMA* stationnaire, la moyenne inconditionnelle est constante dans le temps tandis que la moyenne conditionnelle  $E(X_t|\mathcal{F}_{t-1})$  varie en fonction des observations passées. Parallèlement à cela, le modèle *ARCH* suppose que la variance inconditionnelle du processus d'erreur est constante dans le temps mais permet à la variance conditionnelle de  $\varepsilon_t$  de varier en fonction des erreurs quadratiques passées.

Notons  $\sigma_t^2 = \text{Var}(\varepsilon_t|\mathcal{F}_{t-1})$  la variance conditionnelle de  $\varepsilon_t$ , étant donné le passé  $\mathcal{F}_{t-1}$ , le modèle *ARCH*( $q$ ) de base peut être formulé tel que

$$\varepsilon_t = \sigma_t e_t \tag{5.1}$$

où  $\{e_t\}$  est une suite de variables aléatoires *iid* de moyenne nulle et de variance 1, et

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 \quad (5.2)$$

avec  $\alpha_0 > 0$ ,  $\alpha_i \geq 0$ , pour  $i = 1, \dots, q-1$ , et  $\alpha_q > 0$ .

Les conditions sur les paramètres sont imposées pour s'assurer que la variance conditionnelle  $\sigma_t^2$  est positive. La contrainte additionnelle  $\sum_{i=1}^q \alpha_i < 1$

assure que les  $\varepsilon_t$  sont de covariance stationnaire avec une variance finie inconditionnelle  $\sigma_\varepsilon^2$ . Pour certaines séries chronologiques, telles que les rendements boursiers, les observations originales sont généralement non corrélées et les  $\varepsilon_t$  sont directement observés. Alternativement, le  $\varepsilon_t$  peut être la suite des innovations associée à un modèle de type *ARMA* ou de régression. A des fins de modélisation, on suppose généralement que  $e_t$  dans (5.1) suit une distribution normale ou une distribution de Student.

Le modèle *ARCH* a été utilisé par Engle (1982) pour étudier la variance des taux d'inflation au Royaume-Uni et par Engle (1983) pour décrire la variance des taux d'inflation aux États-Unis. Le modèle *ARCH* et ses extensions ultérieures par Bollerslev (1986) et d'autres ont rapidement trouvé d'autres applications. Par exemple, Diebold et Nerlove (1989) ont montré que le modèle *ARCH* peut être utilisé pour générer des mesures statistiquement et économiquement significatives de la volatilité des taux de change.

Bollerslev (1987) a utilisé l'extension *GARCH* du modèle *ARCH* pour analyser la volatilité conditionnelle des rendements financiers observés à une fréquence mensuelle ou supérieure. Dans Weiss (1984), des modèles *ARMA* avec des erreurs *ARCH* ont été utilisés pour modéliser le comportement en séries chronologiques de 13 séries macroéconomiques américaines différentes. Bollerslev et al. (1992) décrivent un grand nombre d'autres applications dans leur étude des modèles de volatilité. Alors que la majorité des applications ont été réalisées en finance et en économie, les modèles ont également été utilisés dans d'autres domaines. Par exemple, Campbell et Diebold (2005) ont utilisé des modèles de volatilité dans leur analyse des températures moyennes quotidiennes de quatre villes américaines. Les modèles ont également été utilisés pour des variables telles que les vitesses du vent, les mesures de la qualité de l'air, les amplitudes de séismes et dans l'analyse des signaux vocaux. Pour des références choisies, voir Francq et Zekoïan (2010, p.12).

### 5.1.1 Quelques propriétés du modèle *ARCH*

Pour établir certaines propriétés du modèle *ARCH*, nous examinons d'abord le modèle *ARCH*(1) où

$$\sigma_t^2 = \text{Var}(\varepsilon_t | \mathcal{F}_{t-1}) = E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \quad (5.3)$$

avec  $\alpha_0 > 0$  et  $\alpha_1 > 0$ .

La forme du modèle montre que la variance conditionnelle  $\sigma_t^2$  sera grande si  $\varepsilon_{t-1}$  est grand en valeur absolue et vice versa. Une grande (petite) valeur de  $\sigma_t^2$  aura à son tour tendance à générer une grande (petite) valeur de  $\varepsilon_t$ , donnant ainsi lieu à un regroupement de volatilité (phénomène de grappe).

Il résulte de (5.1) que  $E(\varepsilon_t | \mathcal{F}_{t-1}) = 0$ . La moyenne *inconditionnelle* de  $\varepsilon_t$  est aussi zéro puisque

$$E(\varepsilon_t) = E[E(\varepsilon_t | \mathcal{F}_{t-1})]$$

De plus, les  $\varepsilon_t$  sont décorrélés puisque pour  $j > 0$ ,

$$E(\varepsilon_t \varepsilon_{t-j}) = E[E(\varepsilon_t \varepsilon_{t-j} | \mathcal{F}_{t-1})] = E[\varepsilon_{t-j} E(\varepsilon_t | \mathcal{F}_{t-1})] = 0$$

Mais les  $\varepsilon_t$  ne sont pas mutuellement indépendants puisqu'ils sont liés entre eux par leurs variances conditionnelles. Le manque d'autocorrélation est une propriété importante qui rend le modèle *ARCH* approprié pour la modélisation des rendements d'actifs qui ne sont pas corrélés par l'hypothèse d'un marché efficace.

On suppose également que  $\varepsilon_t$  ont des variances *inconditionnelles* égales,  $\text{Var}(\varepsilon_t) = E(\varepsilon_t^2) = \sigma_\varepsilon^2$ , pour tout  $t$ , de sorte que le processus soit faiblement stationnaire. Si  $\alpha_1 < 1$ , la variance inconditionnelle existe et est égale à

$$\sigma_\varepsilon^2 = \text{Var}(\varepsilon_t) = \frac{\alpha_0}{1 - \alpha_1} \quad (5.4)$$

Ceci vient du fait que

$$\sigma_\varepsilon^2 = E(\varepsilon_t^2) = E[E(\varepsilon_t^2 | \mathcal{F}_{t-1})] = E(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2) = \alpha_0 + \alpha_1 \sigma_\varepsilon^2$$

En substituant  $\alpha_0 = \sigma_\varepsilon^2 (1 - \alpha_1)$  de (5.4) dans (5.3), on constate que

$$\sigma_t^2 = \sigma_\varepsilon^2 + \alpha_1 (\varepsilon_{t-1}^2 - \sigma_\varepsilon^2) \quad (5.5)$$

ou de manière équivalente,  $\sigma_t^2 - \sigma_\varepsilon^2 = \alpha_1 (\varepsilon_{t-1}^2 - \sigma_\varepsilon^2)$ .

Par conséquent, la variance conditionnelle de  $\varepsilon_t$  sera supérieure à la variance

inconditionnelle lorsque  $\varepsilon_{t-1}^2$  est plus grande que la variance inconditionnelle  $\sigma_\varepsilon^2$ .

Pour étudier le comportement en queue de  $\varepsilon_t$ , nous examinons le quatrième moment  $\mu_4 = E(\varepsilon_t^4)$ . Si  $\varepsilon_t$  est normalement distribué, alors, conditionnellement au passé, on a

$$E(\varepsilon_t^4 | \mathcal{F}_{t-1}) = 3\sigma_t^4 = 3(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2)^2$$

Par conséquent, le quatrième moment inconditionnel de  $\varepsilon_t$  satisfait à

$$E(\varepsilon_t^4) = E[E(\varepsilon_t^4 | \mathcal{F}_{t-1})] = 3[\alpha_0^2 + 2\alpha_0\alpha_1 E(\varepsilon_{t-1}^2) + \alpha_1^2 E(\varepsilon_{t-1}^4)]$$

Ainsi, si  $\{\varepsilon_t\}$  est stationnaire au quatrième ordre, de sorte que  $\mu_4 = E(\varepsilon_t^4) = E(\varepsilon_{t-1}^4)$ , alors

$$\mu_4 = \frac{3(\alpha_0^2 + 2\alpha_0\alpha_1\sigma_\varepsilon^2)}{1 - 3\alpha_1^2} = \frac{3\alpha_0^2(1 - \alpha_1^2)}{(1 - \alpha_1)^2(1 - 3\alpha_1^2)} \quad (5.6)$$

Puisque  $\mu_4 = E(\varepsilon_t^4) > 0$ , cette expression montre que  $\alpha_1$  doit satisfaire  $0 < \alpha_1 < \frac{1}{\sqrt{3}}$  pour que  $\varepsilon_t$  ait un quatrième moment fini. De plus, si on note  $\kappa$  le kurtosis inconditionnel de  $\varepsilon_t$ , alors

$$\kappa = \frac{E(\varepsilon_t^4)}{[E(\varepsilon_t^2)]^2} = \frac{3(1 - \alpha_1^2)}{1 - 3\alpha_1^2}$$

La valeur du kurtosis dépasse 3, le kurtosis de la distribution normale. Par conséquent, la distribution marginale de  $\varepsilon_t$  a une queue plus lourde que celle de la distribution normale. C'est une caractéristique supplémentaire du modèle *ARCH* qui le rend utile pour la modélisation des retours d'actifs financiers où le comportement à queue lourde est la norme.

Pour définir une autre forme du processus *ARCH*, supposons que  $v_t = \varepsilon_t^2 - \sigma_t^2$ , de sorte que  $\varepsilon_t^2 = v_t + \sigma_t^2$ . Les variables aléatoires  $v_t$  ont alors une moyenne nulle et sont décorrélées puisque

$$\begin{aligned} E[(\varepsilon_t^2 - \sigma_t^2)(\varepsilon_{t-j}^2 - \sigma_{t-j}^2)] &= E[E\{(\varepsilon_t^2 - \sigma_t^2)(\varepsilon_{t-j}^2 - \sigma_{t-j}^2) | \mathcal{F}_{t-1}\}] \\ &= E[(\varepsilon_{t-j}^2 - \sigma_{t-j}^2) E\{(\varepsilon_t^2 - \sigma_t^2) | \mathcal{F}_{t-1}\}] = 0 \end{aligned}$$

De plus, puisque  $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$ , on constate que le modèle *ARCH*(1) peut être écrit tel que

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + v_t \quad (5.7)$$

Cette forme montre que le processus des carrés des erreurs  $\{\varepsilon_t^2\}$  peut être considéré comme un modèle  $AR(1)$  avec des innovations non corrélées  $v_t$ . Les innovations sont hétéroscédastiques et non-gaussiennes dans ce cas.

Pour le modèle  $ARCH(q)$  (5.2), nous avons également

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 + v_t$$

de sorte que  $\varepsilon_t^2$  possède la forme d'un processus  $AR(q)$ .

D'autres résultats liés aux moments et au kurtosis du modèle  $ARCH(1)$  s'étendent également aux modèles  $ARCH$  d'ordre supérieur. En particulier,

si  $\sum_{i=1}^q \alpha_i < 1$ , alors la variance inconditionnelle est

$$\sigma_\varepsilon^2 = \frac{\alpha_0}{1 - \sum_{i=1}^q \alpha_i}$$

comme montré par Engle (1982).

Les conditions nécessaires et suffisantes pour l'existence de moments pairs d'ordre supérieur du processus  $ARCH(q)$  ont été données par Milhoi (1985).

### 5.1.2 Erreurs de prévision pour le modèle ARCH

Les prévisions d'une valeur future  $X_{t+\ell}$  générée à partir d'un modèle  $ARMA$  avec erreurs *iid*  $\varepsilon_t$  ont des erreurs de prévision qui dépendent de l'horizon  $\ell$  mais sont indépendantes de l'origine de prévision  $t$  à partir de laquelle les prévisions sont faites. Baillie et Bollerslev (1992) ont montré que les prévisions d'erreur quadratique moyenne minimale de  $X_{t+\ell}$  sont les mêmes, que les innovations  $\varepsilon_t$  soient ou non hétéroscédastiques. Pour un processus  $ARMA$  avec erreurs  $ARCH$ , ceci implique, en particulier, que l'erreur de prévision à horizon 1 est égale à  $\varepsilon_{t+1}$  alors que l'erreur de prévision à horizon

$\ell$  peut être définie telle que  $e_t(\ell) = \sum_{j=0}^{\ell-1} \psi_j \varepsilon_{t+\ell-j}$  avec  $\psi_0 = 1$ . La présence de

l'hétéroscédasticité conditionnelle aura cependant un impact sur la variance des erreurs de prévision.

Pour un processus  $ARCH(1)$ , la variance conditionnelle de l'erreur de prévision à horizon 1,  $\varepsilon_{t+1}$ , est donnée par (5.5) telle que

$$E[e_t^2(1) | \mathcal{F}_t] = \sigma_{t+1}^2 = \sigma_\varepsilon^2 + \alpha_1 (\varepsilon_t^2 - \sigma_\varepsilon^2) \quad (5.8)$$

La variance conditionnelle de l'erreur de prévision à horizon 1 peut donc être plus petite ou plus grande que la variance inconditionnelle en fonction de la différence entre la dernière erreur au carré  $\varepsilon_t^2$  et  $\sigma_\varepsilon^2$ .

On peut également montrer que les variances conditionnelles des erreurs de prévision à horizon  $\ell$ ,  $e_t(\ell)$  dépendent des erreurs passées au carré en se basant sur la relation

$$E[e_t(\ell) | \mathcal{F}_t] = \sum_{j=0}^{\ell-1} \psi_j^2 E(\varepsilon_{t+\ell-j}^2 | \mathcal{F}_t)$$

où pour le modèle  $ARCH(1)$  on a

$$\begin{aligned} E(\varepsilon_{t+h}^2 | \mathcal{F}_t) &= E[E(\varepsilon_{t+h}^2 | \mathcal{F}_t)] \\ &= \alpha_0 + \alpha_1 E(\varepsilon_{t+h}^2 | \mathcal{F}_t) \\ &= \alpha_0 + \alpha_1 + \dots + \alpha_1^{h-1} + \alpha_1^h \varepsilon_t^2 \quad \text{pour } h > 0 \end{aligned}$$

A partir de ceci et en utilisant (5.4), on peut vérifier que

$$E[e_t(\ell) | \mathcal{F}_t] = \sigma_\varepsilon^2 \sum_{j=0}^{\ell-1} \psi_j^2 + \sum_{j=0}^{\ell-1} \psi_j^2 \alpha_1^{\ell-j} (\varepsilon_t^2 - \sigma_\varepsilon^2) \quad (5.9)$$

ce qui se réduit à (5.8), pour  $\ell = 1$ .

Le premier terme du côté droit de cette expression est la variance d'erreur de prédiction conventionnelle en supposant que les erreurs  $\varepsilon_t$  sont homoscedastiques tandis que le second terme reflète l'impact des effets  $ARCH$ . Ce dernier varie avec le temps et peut à nouveau être positif ou négatif selon la différence  $\varepsilon_t^2 - \sigma_\varepsilon^2$ . La variance des valeurs prédites varie donc avec le temps et peut être plus grande ou plus petite que celle de l'homoscedasticité. Pour le modèle  $ARCH(q)$  général, le deuxième terme du côté droit sera une fonction de  $q$  valeurs passées  $\varepsilon_t^2, \dots, \varepsilon_{t-q+1}^2$ .

Si la série chronologique  $X_t$  suit un modèle  $AR(1)$ , les poids  $\psi$  sont donnés par  $\psi_j = \phi^{j-1}$ . Si  $\phi$  est égal à zéro, de sorte que la moyenne de la série soit une constante indépendante du passé, l'expression (5.9) se réduit à  $\sigma_\varepsilon^2 + \alpha_1^\ell (\varepsilon_t^2 - \sigma_\varepsilon^2)$ . Nous notons qu'il s'agit de la prévision conditionnelle à horizon  $\ell$  de la variance conditionnelle  $\sigma_{t+\ell}^2$  pour le modèle  $ARCH(1)$ . Cette prévision pourrait être calculée plus directement telle que  $E[\sigma_{t+\ell}^2 | \mathcal{F}_t] = \alpha_0 + \alpha_1 E(\varepsilon_{t+\ell-1}^2 | \mathcal{F}_t)$ , où  $E(\varepsilon_{t+\ell-1}^2 | \mathcal{F}_t)$  peut être généré récursivement à partir du modèle  $AR$  pour  $\varepsilon_t^2$ . Le résultat suit en posant  $\alpha_0 = \sigma_\varepsilon^2 (1 - \alpha_1)$ .

## 5.2 Le modèle *GARCH*

Le modèle *ARCH* a un inconvénient en ce sens qu'il nécessite souvent un ordre  $q$  élevé pour décrire de façon adéquate l'évolution de la volatilité au fil du temps. Une extension du modèle *ARCH* appelée modèle *ARCH généralisé*, ou *GARCH* (General-ARCH), a été introduite par Bollerslev (1986) pour résoudre ce problème. Le modèle *GARCH*( $q, p$ ) suppose que  $\varepsilon_t = \sigma_t e_t$ , où les  $\{e_t\}$  sont à nouveau des variables aléatoires *iid* de moyenne nulle et de variance 1, et où  $\sigma_t$  est donné par

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad (5.10)$$

avec  $\alpha_0 > 0$ ,  $\alpha_i \geq 0$ ,  $i = 1, \dots, q-1$ ,  $\alpha_q > 0$ ,  $\beta_j \geq 0$ ,  $j = 1, \dots, p-1$ , et  $\beta_p > 0$ . Ces contraintes sur les paramètres sont suffisantes pour que la variance conditionnelle  $\sigma_t^2$  soit positive. Nelson et Cao (1992) ont montré que ces contraintes peuvent être légèrement assouplies pour permettre à certains paramètres d'être négatifs alors que la variance conditionnelle reste positive. La condition supplémentaire  $\sum_{i=1}^m (\alpha_i + \beta_i) < 1$ , où  $m = \max(q, p)$  avec  $\alpha_i = 0$ , pour  $i > q$ , et  $\beta_j = 0$ , pour  $j > p$ , assure que la variance inconditionnelle  $\sigma_\varepsilon^2$  est finie.

Le modèle le plus simple et le plus largement utilisé dans cette classe est le modèle *GARCH*(1,1) où

$$\sigma_t^2 = E[\varepsilon_t^2 | \mathcal{F}_{t-1}] = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

Puisque les constantes  $\alpha_1$  et  $\beta_1$  sont positives, on constate qu'une grande valeur de  $\varepsilon_{t-1}^2$  ou  $\sigma_{t-1}^2$  donne une grande valeur de  $\sigma_t^2$ . Comme pour le processus *ARCH*, ce modèle prend donc en compte le phénomène de grappe ou clustering de la volatilité.

En supposant que  $\alpha_1 + \beta_1 < 1$ , la variance inconditionnelle de  $\varepsilon_t$  est telle que

$$\sigma_\varepsilon^2 = Var(\varepsilon_t) = \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)}$$

En outre, en supposant que les distributions conditionnelles sont normales, le moment d'ordre quatre inconditionnel de  $\varepsilon_t$  est fini à condition que  $(\alpha_1 + \beta_1)^2 + 2\alpha_1^2 < 1$  (Bollerslev, 1986). De plus, le kurtosis de la distribution marginale de  $\varepsilon_t$  est égal à

$$\kappa = \frac{E(\varepsilon_t^4)}{[E(\varepsilon_t^2)]^2} = \frac{3(1 - (\alpha_1^2 + \beta_1^2)^2)}{1 - (\alpha_1^2 + \beta_1^2)^2 - 2\alpha_1^2} > 3$$

Comme dans le cas *ARCH*, la distribution inconditionnelle de  $\varepsilon_t$  a donc une queue plus lourde que la distribution normale et devrait donner lieu à une fréquence plus élevée d'observations extrêmes ou 'aberrantes' que ce qui serait le cas sous hypothèse normale.

Maintenant, supposons  $v_t = \varepsilon_t^2 - \sigma_t^2$  de telle sorte que  $\sigma_t^2 = \varepsilon_t^2 - v_t$ , où les  $v_t$  sont de moyenne nulle et non corrélés. On remarque alors que le modèle *GARCH*(1,1) peut être réarrangé tel que  $\varepsilon_t^2 - v_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 (\varepsilon_{t-1}^2 - v_{t-1})$  ou bien

$$\varepsilon_t^2 = \alpha_0 + (\alpha_1 + \beta_1) \varepsilon_{t-1}^2 + v_t - \beta_1 v_{t-1} \quad (5.11)$$

Le processus du *carré des erreurs* a donc la forme d'un modèle *ARMA*(1,1) avec des innovations  $v_t$  non corrélées. Les  $v_t$  sont en général hétéroscédastiques. Dans le cas particulier de  $\beta_1 = 0$ , le modèle se réduit à  $\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + v_t$ , qui est la forme *AR*(1) du modèle *ARCH*(1). Pour le processus général *GARCH*( $q, p$ ), l'expression (5.11) se généralise telle que

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^m (\alpha_i + \beta_i) \varepsilon_{t-i}^2 + v_t - \sum_{i=1}^p \beta_i v_{t-i}$$

qui a la forme d'un processus *ARMA* pour  $\varepsilon_t^2$  avec un ordre *AR* égal à  $m = \max(q, p)$ .

La structure d'autocorrélation de  $\varepsilon_t^2$  imite aussi celle du processus *ARMA* à condition que le quatrième moment inconditionnel de  $\varepsilon_t^2$  soit fini (Bollerslev, 1988).

La condition nécessaire et suffisante pour la stationnarité du second ordre du processus *GARCH*( $q, p$ ) est

$$\sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i = \sum_{i=1}^m (\alpha_i + \beta_i) < 1$$

Lorsque cette condition est remplie, la variance inconditionnelle est donnée par

$$\sigma_\varepsilon^2 = \text{Var}(\varepsilon_t) = \frac{\alpha_0}{1 - \sum_{i=1}^m (\alpha_i + \beta_i)}$$



Ceci a été montré par Bollerslev (1986) qui a également donné des conditions nécessaires et suffisantes pour l'existence de tous les moments d'ordre supérieur pour le modèle  $GARCH(1, 1)$  et les moments d'ordre quatre pour les modèles  $GARCH(1, 2)$  et  $GARCH(2, 1)$ . He et Teräsvirta (1999) et Ling et McAleer (2002), entre autres, ont donné des extensions de ces résultats. Les expressions pour les moments d'ordre supérieur et les contraintes sur les paramètres nécessaires pour assurer leur existence deviennent plus complexes pour les modèles d'ordre supérieur. D'autre part, de nombreuses études ont montré que les modèles d'ordre plus bas tels que les modèles  $GARCH(1, 1)$ ,  $GARCH(2, 1)$ , et  $GARCH(1, 2)$  sont souvent adéquats en pratique, le modèle  $GARCH(1, 1)$  étant le plus populaire.