

Statistiques 2 M1  
Cours de Séries temporelles

4 mai 2020

# Table des matières

<b>1</b>	<b>Introduction aux chaînes de Markov à espace d'états fini</b>	<b>4</b>
1.1	Introduction . . . . .	4
1.2	Exemple : le hamster . . . . .	4
1.2.1	La chaîne de Markov . . . . .	4
1.2.2	La matrice de transition . . . . .	5
1.3	Classe de récurrence . . . . .	5
1.3.1	Graphe associé à une matrice de transition . . . . .	5
1.4	Mesure invariante . . . . .	8
1.5	Période, chaîne apériodique . . . . .	8
<b>2</b>	<b>Séries temporelles à observations réelles</b>	<b>10</b>
2.1	Quelques séries temporelles simples. . . . .	10
2.1.1	Modèle d'espérance nulle . . . . .	10
2.1.1.1	Bruit i.i.d. . . . .	10
2.1.1.2	Processus binaire . . . . .	10
2.1.1.3	Marche aléatoire . . . . .	11
2.1.2	Modèles avec tendance et saisonnalité . . . . .	11
2.1.2.1	Tendance déterministe . . . . .	11
2.1.2.2	Régression harmonique . . . . .	11
2.1.3	Une approche générale de la modélisation des séries temporelles . . . . .	11
2.2	Séries stationnaires et fonction d'auto-corrélation . . . . .	12
2.2.1	Quelques définitions . . . . .	12
2.2.2	Quelques exemples . . . . .	12
2.2.2.1	Bruit i.i.d. . . . .	12
2.2.2.2	La marche aléatoire . . . . .	13
2.2.2.3	Moyenne mobile d'ordre 1 . . . . .	13
2.2.2.4	Modèle auto-régressif d'ordre 1 . . . . .	13
2.2.3	La fonction d'auto-corrélation empirique . . . . .	14
2.3	Estimation et élimination de la tendance et de la saisonnalité . . . . .	14
2.3.1	Estimation et élimination de la tendance en absence de saisonnalité . . . . .	15
2.3.1.1	Estimation de la tendance . . . . .	15

2.3.1.2	Estimation et élimination de la tendance et la saisonnalité en même temps . . . . .	16
2.4	Tester les résidus . . . . .	17
<b>3</b>	<b>Equations récurrentes linéaires</b>	<b>18</b>
3.1	Suite sans second membre . . . . .	18
3.2	Suite avec second membre constant . . . . .	22
<b>4</b>	<b>Modèles ARMA</b>	<b>23</b>
4.1	Processus ARMA( $p, q$ ) . . . . .	23
4.1.1	Définition : . . . . .	23
4.1.2	Existence et unicité . . . . .	23
4.2	ACF et PACF d'un ARMA( $p, q$ ) . . . . .	24
4.2.1	Calcul de ACVF et ACF . . . . .	24
4.2.2	La fonction d'auto-corrélation partielle (PACF) . . . . .	25
4.3	Prévisions d'un processus ARMA . . . . .	26
4.3.1	Algorithme de Durbin-Levinson . . . . .	27
4.3.2	Algorithme d'innovation . . . . .	28
4.3.3	Prévision d'un ARMA . . . . .	29
4.3.4	Prédictions à $h$ pas . . . . .	30
4.3.5	Intervalle de confiance . . . . .	30
<b>5</b>	<b>Analyse spectrale</b>	<b>31</b>
5.1	Introduction . . . . .	31
5.2	Densité spectrale . . . . .	31
5.3	Le périodogramme . . . . .	36
5.4	Filtres linéaires invariants dans le temps . . . . .	38
5.5	Densité spectrale d'un processus ARMA . . . . .	40
<b>6</b>	<b>Modélisation et prévision d'un processus ARMA</b>	<b>41</b>
6.1	Estimation préliminaire . . . . .	41
6.1.1	Estimation de Yule-Walker . . . . .	41
6.1.2	Algorithme de Burg . . . . .	42
6.1.3	L'algorithme de Hannan-Rissanen . . . . .	43
6.2	Estimation par maximum de vraisemblance . . . . .	44
6.2.1	Contraste de Whittle . . . . .	45
6.2.2	Régions de confiance pour les coefficients . . . . .	45
6.3	Sélection de l'ordre . . . . .	46
6.3.1	Critère AICC . . . . .	46
6.3.2	Critère BIC . . . . .	46
<b>7</b>	<b>Modèle SARIMA</b>	<b>48</b>
7.1	Modèles ARIMA . . . . .	48
7.1.1	Définition . . . . .	48
7.1.2	Racine unitaire . . . . .	49
7.1.2.1	Racine unitaire pour la partie autorégressive . . . . .	49

*TABLE DES MATIÈRES*

3

7.1.3	Prédiction des modèles ARIMA . . . . .	50
7.1.4	Fonction de prédiction . . . . .	52
7.2	Modèles SARIMA . . . . .	52
7.2.1	Définition . . . . .	52
7.2.2	Prédiction des SARIMA . . . . .	53

# Chapitre 1

## Introduction aux chaînes de Markov à espace d'états fini

### 1.1 Introduction

On considère un ensemble  $E$  fini, par exemple  $E = \{1, \dots, N\}$ , où  $N \in \mathbb{N}$ . Une chaîne de Markov (homogène) à espace d'états fini est une suite de variables aléatoires  $(X_1, \dots, X_n, \dots) \in E^{\mathbb{N}}$ , telles que :  $P(X_{n+1} | X_n, \dots, X_0) = P(X_{n+1} | X_n)$ , i.e. la loi de  $X_{n+1}$  conditionnellement à tout le passé du processus ne dépend que du passé immédiat  $X_n$ .

Le processus  $(X_n)_{n \in \mathbb{N}}$ , est donc défini par toutes les probabilités conditionnelles définies sur  $E^2$  :  $P(X_{n+1} = i | X_n = j) := p_{ij}$ .

Si on représente tous les états possible sur le simplexe de  $\mathbb{R}^N$ , c'est-à-dire l'état  $i$  sera représenté par  $(0, \dots, 0, 1, 0, \dots, 0)$  avec le 1 en  $i$ ème position, on pourra noter  $E(X_{n+1} | X_n)^T = (P(X_{n+1} = e_i | X_n))_{1 \leq i \leq N}^T = MX_n$ , avec

$$M = \begin{pmatrix} p_{11} & \cdots & p_{1N} \\ \vdots & \ddots & \vdots \\ p_{N1} & \cdots & p_{NN} \end{pmatrix}. \text{ On appellera } M \text{ la matrice de transition de la}$$

chaîne de Markov  $(X_n)_{n \in \mathbb{N}}$ . On remarquera que  $M$  est "stochastique" c'est-à-dire que la somme sur des lignes sur une colonne vaudra toujours 1 ( $\sum_{i=1}^N p_{ij} = 1$ ), ainsi  $P(X_{n+1} | X_n = j)$  est bien une probabilité.

### 1.2 Exemple : le hamster

#### 1.2.1 La chaîne de Markov

le hamster paresseux ne connaît que 3 endroits dans sa cage :

1. Les copeaux où il dort (état  $(1, 0, 0)$ )
2. La mangeoire où il mange (état  $(0, 1, 0)$ )

3. La roue où il fait de l'exercice (état  $(0, 0, 1)$ ).

Ses journées sont semblables les unes aux autres et on représente son activité par une chaîne de Markov. Toutes les minutes, il peut soit changer d'activité, soit continuer celle qu'il était en train de faire :

- Quand il dort, il a 9 chances sur 10 de ne pas se réveiller la minute suivante
- Quand il se réveille, il y a 1 chance sur 2 qu'il aille manger et 1 chance sur 2 qu'il parte faire de l'exercice.
- Le repas ne dure qu'une minute, après il fait autre chose.
- Après avoir mangé, il y a 3 chances sur 10 qu'il parte courir dans sa roue, mais surtout 7 chances sur 10 qu'il retourne dormir.
- Courir est fatigant, il a donc 8 chances sur 10 de retourner dormir au bout d'une minute, sinon (avec 2 chances sur 10) il continue en oubliant qu'il est déjà un peu fatigué.

### 1.2.2 La matrice de transition

Si on note  $(X_n)_{n \in \mathbb{N}}$ , la suite des états du hamster, la matrice de transition sera  $M = \begin{pmatrix} 0.9 & 0.7 & 0.8 \\ 0.05 & 0 & 0 \\ 0.05 & 0.3 & 0.2 \end{pmatrix}$ . Cette matrice de transition nous permet de faire facilement des prévisions sur les états futurs. Prenons l'hypothèse que le hamster dort lors de la première minute de l'étude.  $X_1 = (1, 0, 0)^T$ , au bout d'une minute on peut prédire :

$$\hat{X}_1 = MX_0 = \begin{pmatrix} 0.9 & 0.7 & 0.8 \\ 0.05 & 0 & 0 \\ 0.05 & 0.3 & 0.2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.9 \\ 0.05 \\ 0.05 \end{pmatrix}$$

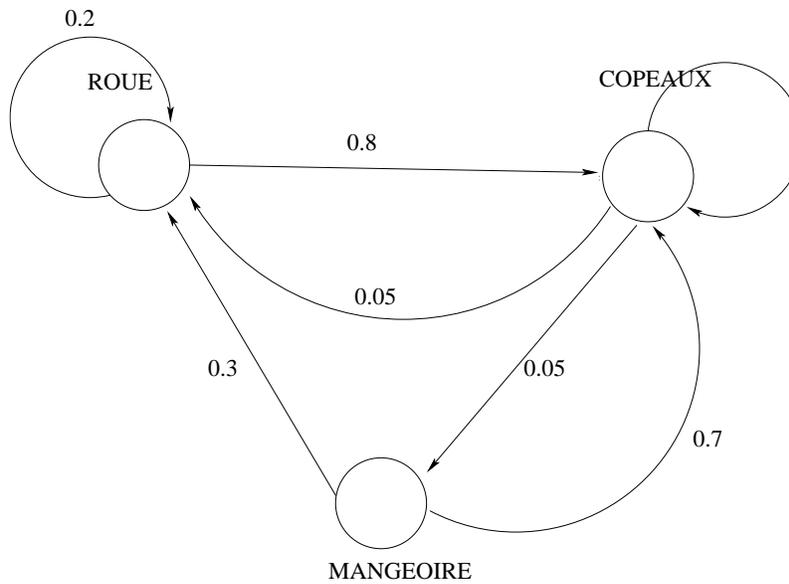
Ainsi, après une minute, on a 90% de chance que le hamster dorme encore, 5% qu'il mange et 5% qu'il coure dans sa roue. On sait de plus que  $E(X_2 | X_0) = E(E(X_2 | X_1) | X_0)$  d'où

$$\begin{aligned} \hat{X}_2 &= MX_1 = MMX_0 = \begin{pmatrix} 0.9 & 0.7 & 0.8 \\ 0.05 & 0 & 0 \\ 0.05 & 0.3 & 0.2 \end{pmatrix} \begin{pmatrix} 0.9 & 0.7 & 0.8 \\ 0.05 & 0 & 0 \\ 0.05 & 0.3 & 0.2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 0.9 & 0.7 & 0.8 \\ 0.05 & 0 & 0 \\ 0.05 & 0.3 & 0.2 \end{pmatrix} \begin{pmatrix} 0.9 \\ 0.05 \\ 0.05 \end{pmatrix} = \begin{pmatrix} 0.885 \\ 0.045 \\ 0.045 \end{pmatrix} \end{aligned}$$

## 1.3 Classe de récurrence

### 1.3.1 Graphe associé à une matrice de transition

On peut dessiner sur un graphe toutes les "liaisons" entre les états, c'est-à-dire les coordonnées strictement positive de la matrice de transition. Ainsi dans l'exemple précédent on aura :



On dira que l'état  $y$  est accessible par  $x$  si il existe une suite finie de points de  $E$  :  $x = x_0, \dots, x_n = y$  et telle que pour tout  $k \in \{0, \dots, n-1\}$ ,  $p_{x_{k+1}x_k} > 0$ , c'est-à-dire que la probabilité de passer de l'état  $x_{k+1}$  en venant de l'état  $x_k$  est strictement positive. On notera alors  $x \rightarrow y$ .

**Définition :** Nous dirons qu'un point est récurrent si  $\forall y \in E, x \rightarrow y \Rightarrow y \rightarrow x$ . Un point  $x$  non récurrent sera dit transitoire.

**Remarques :**

- Pour tout  $x \in E$ , il y a au moins un point  $y$  tel que  $x \rightarrow y$ , car  $\sum_y p_{yx} = 1$ .
- $x \rightarrow y$  si et seulement si il existe un entier  $n \geq 1$  tel que la coordonnée  $yx$  de  $M^n := M^n(x, y) > 0$ . En effet,

$$M^n(x, y) = \sum_{(x_{i_1}, \dots, x_{i_{n-1}}) \in E^{n-1}} p_{x_{i_1}x} p_{x_{i_2}x_{i_1}} \cdots p_{yx_{i_{n-1}}}.$$

Dans la somme précédente, tous les termes sont positifs ou nuls ; la somme n'est donc non nulle que si un des termes au moins ne l'est pas : c'est la définition de  $x \rightarrow y$ .

- Lorsque  $M^n(x, y) > 0$ , nous dirons qu'il existe un chemin de longueur  $n$  qui va de  $x$  à  $y$ .
- La relation  $x \rightarrow y$  est transitive. En effet, si  $M^n(x, y) > 0$  et  $M^q(y, z) > 0$ , alors  $M^{n+q}(x, z) > 0$  puisque

$$M^{n+q}(x, z) = \sum_{y_1} M^n(x, y_1) M^q(y_1, z) \geq M^n(x, y) M^q(y, z) > 0$$

- On peut avoir ou non  $x \rightarrow x$ , mais si le point  $x$  est récurrent, alors il existe au moins un point  $y$  tel que  $x \rightarrow y$  alors  $y \rightarrow x$  et par transitivité  $x \rightarrow x$ . Par contre un point transitoire peut ou non satisfaire  $x \rightarrow x$  : considérons les deux matrices de transition suivantes sur les points  $\{1, 2\}$  :

$$M_1 = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \text{ et } M_2 = \begin{pmatrix} 0.5 & 0 \\ 0.5 & 1 \end{pmatrix}$$

Dans les deux cas, le point 1 est transitoire, mais la relation  $1 \rightarrow 1$  est fautive pour  $M_1$  et vraie pour  $M_2$ .

- Si  $x \rightarrow y$  et si  $x$  est récurrent, alors  $y$  est récurrent : il suffit d'appliquer la propriété de transitivité de la relation  $\rightarrow$ .

L'ensemble des points transitoires peut être vide, mais pas l'ensemble des points récurrents si l'ensemble  $E$  est non vide. On en déduit la proposition suivante :

**Proposition :** Sur l'ensemble  $\mathcal{R}$  des points récurrents, la relation  $x \rightarrow y$  est une relation d'équivalence.

**Définition :** On appelle classe de récurrence les classes d'équivalence de  $\mathcal{R}$  pour cette relation d'équivalence. Nous dirons qu'une matrice markovienne est récurrente irréductible s'il n'y a pas de point transitoire et une seule classe de récurrence. Par abus de langage on dit qu'une chaîne de Markov homogène est récurrente irréductible si sa matrice l'est.

**Remarque :**

- Une matrice markovienne est récurrente irréductible si et seulement si pour tout couple de points  $(x, y)$  on a  $x \rightarrow y$ .
- Si  $\mathcal{R}$  est une classe de récurrence et  $x \in \mathcal{R}$ , alors  $M(x, y) = 0, \forall y \notin \mathcal{R}$ . Par conséquent la matrice  $M^{\mathcal{R}}$  définie sur  $\mathcal{R} \times \mathcal{R}$  comme la restriction de la matrice  $M$  à la classe  $\mathcal{R}$  est encore une matrice markovienne.

## 1.4 Mesure invariante

Une mesure de probabilité  $\mu$  sur  $E$  pourra toujours être représentée par un vecteur  $\mu := (\mu_1, \dots, \mu_N)^T$  tel que les  $\mu_i$  soient tous positifs ou nuls et  $\sum_{i=1}^N \mu_i = 1$  i.e.  $\mu_i = P(X = i)$ . La mesure  $\mu$  sera dite invariante pour  $M$  si et seulement si  $\mu = M\mu$ . Pour une matrice de transition, il existe toujours une mesure invariante. En effet, pour une mesure  $\mu_0$  quelconque considérons la suite  $\mu_n := \frac{1}{n+1} \sum_{i=0}^n M^i \mu_0$ , c'est une suite de probabilités, donc de points dans le compact  $[0, 1]^N$ , elle admet donc une sous-suite convergente  $\mu_{n_k} \xrightarrow{k \rightarrow \infty} \mu$ . Mais alors

$$M\mu_{n_k} = \frac{1}{n_k + 1} \sum_{i=0}^{n_k} M^{i+1} \mu_0 = \mu_{n_k} + \frac{M^{n_k+1} \mu_0 - \mu_0}{n_k + 1}$$

et lorsque  $k \rightarrow \infty$ ,  $\frac{(M^{n_k+1} \mu_0 - \mu_0)}{n_k + 1} \rightarrow 0$ , puisqu'il s'agit de la différence de deux mesures de probabilité divisée par  $n_k + 1$ . Alors, nous voyons en passant à la limite que  $M\mu = \mu$ .

On aura de plus le résultat suivant (admis)

**Théorème** Soit  $M$  une matrice markovienne; la mesure invariante est unique si et seulement si il n'y a qu'une seule classe de récurrence.

## 1.5 Période, chaîne apériodique

Dans la section précédente, nous avons vu que si la chaîne est récurrente irréductible, alors la suite  $\mu_n = \frac{\mu_0 + M\mu_0 + \dots + M^n \mu_0}{n+1}$  converge vers l'unique mesure invariante  $\mu$ . Nous nous préoccuons dans ce chapitre de la convergence de la loi de  $X_n$  vers  $\mu$ . Un premier exemple élémentaire montre que le simple fait pour  $M$  d'être récurrente irréductible n'est pas suffisant. Considérons pour cela la matrice  $M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ , matrice d'une chaîne de Markov sur deux points  $\{1, 2\}$ . Si la loi de  $X_0$  est  $\delta_{\{1\}}$ , alors la loi de  $X_{2p}$  est  $\delta_{\{1\}}$  et celle de  $X_{2p+1}$  est  $\delta_{\{2\}}$ , donc la loi de  $X_n$  ne converge pas vers l'unique mesure invariante qui est la loi uniforme.

Pour traiter ce problème de convergence, nous introduisons une nouvelle notion : la période.

**Définition :** Soit  $x$  un point de  $E$  et  $N(x) = \{n \mid M^n(x, x) > 0\}$ . La période  $p(x)$  est le plus grand diviseur commun (PGCD) de  $N(x)$ .

On a la première remarque importante suivante :

**Proposition** Dans une classe de récurrence, tous les points ont même période.

**Preuve** Soient  $x$  et  $y$  deux points de la même classe de récurrence. Il suffit de démontrer que  $p(x)$  est un diviseur de  $N(y)$  et par symétrie on aura  $p(x) = p(y)$ . Soit  $m$  un élément de  $N(y)$  c'est-à-dire que  $M^m(y, y) > 0$ .  $x$  et  $y$  étant dans la même classe de récurrence, il existe deux entiers  $n_1$  et  $n_2$  tels que  $M^{n_1}(x, y) > 0$  et  $M^{n_2}(y, x) > 0$ . Alors,  $n_1 + n_2$  et  $n_1 + n_2 + m$  appartiennent tous les deux à  $N(x)$  car

$$M^{n_1+n_2+m}(x, x) \geq M^{n_1}(x, y) M^m(y, y) M^{n_2}(y, x) > 0$$

donc  $p(x)$  divise  $n_1 + n_2 + m$ ,  $n_1 + n_2$  et par différence  $m$  ■

On peut alors énoncé alors la définition suivante :

**Définition** On dira qu'une matrice markovienne récurrente est apériodique si la période d'un point quelconque de  $E$  est égale à 1.

On déduit de la proposition précédente que si une matrice est récurrente irréductible et s'il existe un point  $x$  pour lequel  $P(x, x) > 0$  alors elle est apériodique. La propriété fondamentale des matrices récurrentes irréductibles apériodiques est la suivante :

**Proposition (admise)** Si la chaîne est irréductible récurrente et apériodique, on aura  $M^n$  qui convergera vers une matrice dont chaque ligne est l'unique distribution invariante  $\mu : \lim_{n \rightarrow \infty} M^n \mu_0 = \mu$ .

Cette importante propriété permet d'établir la loi des grands nombres et le théorème de la limite centrale pour la chaîne de Markov  $(X_n)$  ce qui permet de faire des statistiques :

**Loi des grands nombres (admise)** Soit  $(X_n)$  une chaîne de Markov définie sur un espace fini  $E$  homogène irréductible et apériodique d'unique mesure stationnaire  $\mu$ , alors quelle que soit la loi de  $X_0$  et quelle que soit la fonction  $f : E \rightarrow \mathbb{R}$ ,

$$\frac{1}{n+1} (f(X_0) + \dots + f(X_n)) \xrightarrow{p.s.} \sum_{x \in E} f(x) \mu(x)$$

qui peut se généraliser en

$$\frac{1}{n+1} (f(X_0, X_1) + \dots + f(X_{n-1}, X_n)) \xrightarrow{p.s.} \sum_{(x,y) \in E^2} f(x, y) \mu(x) M_x(y)$$

où  $M_x(\cdot)$  est la probabilité de transition  $P(X_1 = \cdot | X_0 = x)$ .

## Chapitre 2

# Séries temporelles à observations réelles

### 2.1 Quelques séries temporelles simples.

Un modèle de séries temporelles pour les observations  $(x_t)_{1 \leq t \leq n}$  est la spécification d'une probabilité (ou d'une moyenne et d'une covariance) pour les variables  $(X_t)_{t \in \mathbb{N}^*}$ , pour lesquelles  $(x_t)_{1 \leq t \leq n}$  sont supposées être une réalisation. Une détermination complète du modèle demanderait de spécifier toutes les probabilités

$$P[X_1 \leq x_1, \dots, X_n \leq x_n], \quad -\infty < x_1, \dots, x_n < \infty, \quad n \in \mathbb{N}^*$$

En générale cette question est trop difficile ou demande des hypothèses fortes (gaussianité). On s'intéressera surtout à la spécification des premiers et seconds moments de la distribution, c'est-à-dire  $(E(X_t))_{t \in \mathbb{N}^*}$  et  $(E(X_{t+h}X_t))_{t \in \mathbb{N}^*, h \in \mathbb{N}}$  ce qui reviendra à déterminer la loi jointe de  $(X_t)_{t \in \mathbb{N}^*}$ , si on suppose que le processus est gaussien.

#### 2.1.1 Modèle d'espérance nulle

##### 2.1.1.1 Bruit i.i.d.

C'est certainement le modèle le plus simple de série temporelle.

##### 2.1.1.2 Processus binaire

Comme exemple de bruit i.i.d. on peut considérer la séquence de bruit  $(X_t)_{t \in \mathbb{N}^*} \in \{-1, 1\}^{\mathbb{N}^*}$ , avec  $P(X_t = 1) = \frac{1}{2}$ .

### 2.1.1.3 Marche aléatoire

La marche aléatoire  $(S_t)_{t \in \mathbb{N}}$ , commençant à zéro, est obtenue en sommant cumulativement une variable aléatoire i.i.d. de moyenne nulle. ainsi la marche aléatoire est définie par  $S_0 = 0$  et  $S_t = \sum_{k=1}^t X_k$ .

## 2.1.2 Modèles avec tendance et saisonnalité

### 2.1.2.1 Tendance déterministe

Il arrive souvent que des séries semblent avoir une tendance haussière ou baissière (par exemple les séries de PIB). On peut alors essayer de modéliser les séries par une équation de la forme :  $X_t = m_t + Y_t$ , où  $m_t$  est une fonction variant lentement, (appelée tendance) et  $Y_t$  est de moyenne nulle. On peut, par exemple supposer que la tendance à une forme polynomiale :  $m_t = a_0 + a_1 t + a_2 t^2$  et déterminer les paramètres  $a_0, a_1, a_2$  par régression linéaire, c'est-à-dire en minimisant en  $(a_0, a_1, a_2) : \sum_{i=1}^n (x_t - m_t)^2$ .

### 2.1.2.2 Régression harmonique

Beaucoup de séries temporelles sont influencées par des facteurs saisonniers, comme le temps ou les jours de la semaine. Afin de représenter de tels modèles saisonniers on peut utiliser le modèle simple :  $X_t = s_t + Y_t$ , où  $s_t$  est une fonction périodique de  $t$  de période  $d$  ( $s_{t+d} = s_t$ ). On pourra par exemple choisir

$$s_t = a_0 + \sum_{j=1}^k a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t)$$

où  $a_0, a_1, \dots, a_k$  et  $b_1, \dots, b_k$  sont des paramètres inconnus et  $\lambda_1, \dots, \lambda_k$  sont des fréquences fixées, chacune étant un multiple entier de  $\frac{2\pi}{d}$ .

## 2.1.3 Une approche générale de la modélisation des séries temporelles

L'étude d'une série temporelle se déroule selon le processus suivante

- Affichage de la série afin de repérer :
  - Une tendance
  - Une composante saisonnière
  - Une changement brusque de comportement
  - Des observations aberrantes
- Enlever la tendance et la saison afin d'obtenir une série de résidus stationnaires. Il peut être aussi nécessaire d'appliquer des transformations au données. Par exemple, si l'amplitude des données semble fluctuer linéairement il pourra être préférable d'étudier la série  $(\ln(X_t)_{t \in \mathbb{N}})$ , ou plus généralement la transformation de Box-Cox :

$$\begin{cases} \frac{X_t^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \ln(X_t) & \text{si } \lambda = 0 \end{cases}$$

- Choisir un modèle pour modéliser les résidus. En utilisant les différentes statistiques calculables sur cette série.
- La prévision sera faite en prévoyant les résidus et en inversant les transformations décrites ci-dessus pour obtenir les prévisions de la série originale.

## 2.2 Séries stationnaires et fonction d'auto-corrélation

### 2.2.1 Quelques définitions

Une série  $(X_t)_{t \in \mathbb{Z}}$  est dite stationnaire si, pour tout entier  $h$ , elle a les mêmes propriétés statistiques que la série décalée dans le temps  $(X_{t+h})_{t \in \mathbb{Z}}$ . On peut aussi restreindre la stationnarité aux moments d'ordre 1 et 2 (on parlera alors de série faiblement stationnaire).

**Définition** Soit  $(X_t)_{t \in \mathbb{Z}}$  une série avec  $E(X_t^2) < \infty$ . La fonction moyenne de  $(X_t)_{t \in \mathbb{Z}}$  est  $\mu_X(t) = E(X_t)$ . La fonction covariance de  $(X_t)_{t \in \mathbb{Z}}$  est

$$\gamma_X(r, s) = Cov(X_r, X_s) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))]$$

pour tout entier  $r$  et  $s$ .

**Définition** On dira que  $(X_t)_{t \in \mathbb{Z}}$  est faiblement stationnaire si

1.  $\mu_X(t)$  est indépendante de  $t$ .
2.  $\gamma_X(t+h, t)$  est indépendante de  $t$  pour tout  $h$ .

On dira que  $(X_t)_{t \in \mathbb{Z}}$  est fortement stationnaire si  $(X_1, \dots, X_n)$  et  $(X_{1+h}, \dots, X_{n+h})$  ont la même loi pour tout entier  $h$  et  $n > 0$ .

**Remarques** Une série strictement stationnaire telle que  $E(X_t^2) < +\infty$ , est faiblement stationnaire. Quand on fera référence à une série stationnaire la fonction d'auto-covariance sera définie par  $\gamma_X(h) := \gamma_X(h, 0) = \gamma_X(t+h, t)$ .

**Définition** Soit  $(X_t)_{t \in \mathbb{Z}}$ , une série stationnaire. La fonction d'auto-covariance (ACVF) de  $X_t$  au pas  $h$  est

$$\gamma_X(h) = Cov(X_{t+h}, X_t)$$

La fonction d'auto-corrélation (ACF) de  $X_t$  au pas  $h$  est

$$\rho_X(h) := \frac{\gamma_X(h)}{\gamma_X(0)} = Corr(X_{t+h}, X_t)$$

### 2.2.2 Quelques exemples

#### 2.2.2.1 Bruit i.i.d.

Si  $(X_t)_{t \in \mathbb{Z}}$  est un bruit i.i.d. et  $E(X_t^2) < \infty$ , alors  $\gamma_X(h) = \sigma^2 \mathbf{1}_{\{0\}}(h)$  et  $\rho_X(h) = \mathbf{1}_{\{0\}}(h)$ .

**2.2.2.2 La marche aléatoire**

Si  $(S_t)_{t \in \mathbb{N}}$  est une marche aléatoire, alors  $E(S_t) = 0$  et  $E(S_t^2) = t\sigma^2$ . De plus  $\gamma_S(t+h, t) = Cov(S_{t+h}, S_t) = Cov(S_t, S_t) = t\sigma^2$ . Puisque  $\gamma_S(t+h, t)$  dépend de  $t$  la série n'est pas stationnaire.

**2.2.2.3 Moyenne mobile d'ordre 1**

On considère la série définie par les équations

$$X_t = \varepsilon_t + \theta\varepsilon_{t-1}, t \in \mathbb{Z}$$

où  $(\varepsilon_t)_{t \in \mathbb{Z}}$  est i.i.d.  $\mathcal{N}(0, \sigma^2)$ . On aura alors  $E(X_t) = 0$ ,  $E(X_t^2) = \sigma^2(1 + \theta^2)$  et

$$\gamma_X(t+h, t) = \begin{cases} \sigma^2(1 + \theta^2) & \text{si } h = 0 \\ \sigma^2\theta & \text{si } h = 1 \\ 0 & \text{si } |h| > 1 \end{cases}$$

la série est donc stationnaire, on aura de plus

$$\rho_X(t+h, t) = \begin{cases} 1 & \text{si } h = 0 \\ \frac{\theta}{1+\theta^2} & \text{si } h = 1 \\ 0 & \text{si } |h| > 1 \end{cases}$$

**2.2.2.4 Modèle auto-régressif d'ordre 1**

Soit  $a \in \mathbb{R}$ ,  $|a| < 1$ , supposons que  $(X_t)_{t \in \mathbb{Z}}$  soit une série stationnaire qui satisfait les équations

$$X_t = aX_{t-1} + \varepsilon_t$$

où  $(\varepsilon_t)_{t \in \mathbb{Z}}$  est i.i.d.  $\mathcal{N}(0, \sigma^2)$  et  $\varepsilon_t$  est indépendant de  $X_s$  pour  $s < t$ . On voit que

$$E(X_t) = aE(X_{t-1}) \Leftrightarrow (1-a)E(X_t) = 0 \Leftrightarrow E(X_t) = 0$$

Pour trouver la fonction de covariance, il faut multiplier chaque coté de l'équation du modèle par  $X_{t-h}$ ,  $h > 0$  et prendre l'espérance. On a alors

$$\begin{aligned} \gamma_X(h) &= Cov(X_t, X_{t-h}) = \\ &Cov(aX_{t-1}, X_{t-h}) = a\gamma_X(h-1) \\ &= \dots = a^h\gamma_X(0) \end{aligned}$$

en observant que  $\gamma_X(h) = \gamma_X(-h)$ , on trouvera aussi  $\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = a^{|h|}$ . Finalement pour trouver  $\gamma_X(0)$  on aura

$$\gamma_X(0) = Cov(X_t, X_t) = Cov(aX_{t-1} + \varepsilon_t, aX_{t-1} + \varepsilon_t) = a^2\gamma_X(0) + \sigma^2$$

Ainsi  $\gamma_X(0) = \frac{\sigma^2}{(1-a^2)}$ .

### 2.2.3 La fonction d'auto-corrélation empirique

Pour déterminer le degré de dépendance des données pour des données observées  $(x_1, \dots, x_n)$  on calcule la fonction d'auto-corrélation empirique qui nous renseignera sur les propriétés de la série. Par exemple, une fonction d'auto-corrélation empirique proche de 0 pour tous les retards différents de 0 suggère que le modèle approprié pour la série est le bruit i.i.d.

**Définition :** Soit  $x_1, \dots, x_n$  des observations d'une série temporelle.

- La moyenne empirique est  $\bar{x} := \frac{1}{n} \sum_{i=1}^n x_i$ .
- La covariance empirique est

$$\hat{\gamma}(h) := \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}), \quad -n < h < n$$

Pour que cette quantité soit proche de la covariance théorique, il faut que  $h$  soit petit devant  $n$ , par exemple  $h < 10$  et  $n > 100$ . En effet, cet estimateur est biaisé, mais on peut montrer qu'alors la matrice de covariance empirique :

$$\hat{\Gamma}_n = \begin{bmatrix} \hat{\gamma}(0) & \cdots & \hat{\gamma}(n-1) \\ \vdots & \ddots & \vdots \\ \hat{\gamma}(n-1) & \cdots & \hat{\gamma}(0) \end{bmatrix}$$

est positive, comme il se doit.

- L'auto-corrélation empirique est

$$\hat{\rho}(h) := \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad -n < h < n$$

## 2.3 Estimation et élimination de la tendance et de la saisonnalité

Le première chose à faire dans l'analyse d'une série temporelle est d'afficher les données pour repérer d'éventuelles tendances et saisonnalités. Cela peut suggérer une représentation des données de la forme

$$X_t = m_t + s_t + Y_t$$

où  $m_t$  est la fonction tendance,  $s_t$  la composante saisonnière et  $Y_t$  est un bruit stationnaire. Notre but est d'estimer et d'extraire les composantes déterministes  $m_t$  et  $s_t$  dans l'espoir que le résidu  $Y_t$  sera une série temporelle stationnaire. Une autre approche est de différencier la série jusqu'à obtenir la réalisation d'une série stationnaire  $W_t$ .

### 2.3.1 Estimation et élimination de la tendance en absence de saisonnalité

En absence de saisonnalité le modèle devient le suivant :  $X_t = m_t + Y_t$ , avec  $E(Y_t) = 0$ .

#### 2.3.1.1 Estimation de la tendance

**Lissage avec un filtre moyenne mobile** Soit  $q$  un entier strictement positif, considérons la moyenne mobile

$$W_t = \frac{1}{2q+1} \sum_{j=-q}^q X_{t-j}$$

alors pour  $q+1 \leq t \leq n-q$  :

$$W_t = \frac{1}{2q+1} \sum_{j=-q}^q m_{t-j} + \frac{1}{2q+1} \sum_{j=-q}^q Y_{t-j} \simeq m_t$$

Si on suppose que  $m_t$  est approximativement linéaire sur l'intervalle  $[t-q, t+q]$  et que la moyenne des termes d'erreur sur cet intervalle est proche de zéro, la moyenne mobile nous fournit alors l'estimation

$$\hat{m}_t = \frac{1}{2q+1} \sum_{j=-q}^q X_{t-j}, \quad q+1 \leq t \leq n-q$$

Puisque  $X_t$  n'est pas observé pour  $t \leq 0$  ou  $t > n$ , on peut définir  $X_t := X_1$  pour  $t < 1$  et  $X_t := X_n$ , pour  $t > n$ .

**Élimination de la tendance par différentiation** On essaye ici d'enlever la tendance sans l'estimer. On définit l'opérateur de différentiation d'ordre 1  $\nabla$ , par

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

où  $B$  est l'opérateur de retard :  $BX_t := X_{t-1}$ . Les puissances des opérateurs  $B$  et  $\nabla$  sont définis de manière naturelle, par exemple

$$\nabla^2 X_t = \nabla(\nabla X_t) = (1 - B)(1 - B) X_t = (1 - 2B + B^2) X_t = X_t - 2X_{t-1} + X_{t-2}$$

Si l'opérateur  $\nabla$  est appliqué à une série avec une tendance linéaire :  $m_t = c_0 + c_1 t$ , alors on obtient la fonction constante

$$\nabla m_t = m_t - m_{t-1} = c_0 + c_1 t - (c_0 + c_1(t-1)) = c_1$$

De la même manière on peut éliminer une tendance polynomiale de degré  $k$  en appliquant l'opérateur  $\nabla^k$ .

**2.3.1.2 Estimation et élimination de la tendance et la saisonnalité en même temps**

Le modèle est spécifié ainsi :

$$X_t = m_t + s_t + Y_t$$

où  $E(Y_t) = 0$ ,  $s_{t+d} = s_t$  et  $\sum_{j=1}^d s_j = 0$

**Estimation de la tendance et de la saison** Soit les observations  $x_1, \dots, x_n$  la tendance est alors estimée par une moyenne mobile choisie pour éliminer la composante saisonnière et le bruit. Si la période est paire,  $d = 2q$  alors on utilise

$$\hat{m}_t = \frac{(0.5x_{t-q} + x_{t-q+1} + \dots + x_{t+q-1} + 0.5x_{t+q})}{d}, \quad q < t \leq n - q$$

Si la période est impaire  $d = 2q + 1$  alors on utilise la moyenne mobile classique :

$$\hat{m}_t = \frac{1}{2q + 1} \sum_{j=-q}^q X_{t-j}, \quad q + 1 \leq t \leq n - q$$

La seconde étape est d'estimer la composante saisonnière. Pour  $k = 1, \dots, d$ , on calcule la moyenne  $w_k$  des déviations

$$\{(x_{k+jd} - \hat{m}_{k+jd}), \quad q < k + jd \leq n - q\}$$

Puisque les moyennes des déviations ne somment pas nécessairement à 0, on estime alors le composant  $s_k$  de la saison comme

$$\hat{s}_k = w_k - \frac{\sum_{i=1}^d w_i}{d}$$

et  $\hat{s}_k = \hat{s}_{k-d}$ ,  $k > d$ . Les données désaisonnalisées sont alors définies comme  $d_t := x_t - \hat{s}_t$ . Finalement on réestime la tendance avec la méthode de la section précédente.

**Élimination de la tendance et de la saisonnalité par différentiation**

On peut aussi différencier la série pour enlever la saisonnalité. Pour enlever une saisonnalité d'ordre  $d$  on introduit la différence d'ordre  $d$  :

$\nabla_d := (1 - B^d)$ . En appliquant cet opérateur au modèle on trouve :

$$\nabla_d X_t = m_t - m_{t-d} + Y_t - Y_{t-d}$$

Si'il est non nul, le terme  $(m_t - m_{t-d})$  peut être enlevé en utilisant une des techniques vues à la section précédente.

## 2.4 Tester les résidus

L'objectif des traitements des précédentes sections est de produire une série stationnaire. En supposant que c'est fait, l'étape suivante est de modéliser les résidus  $Y_t$ . Si il n'y a pas de dépendance parmi ces résidus alors on peut considérer qu'ils sont i.i.d. et la modélisation de la série temporelle est achevée. On peut tester si les résidus observés sont i.i.d. à l'aide des statistiques suivantes :

1. La fonction de corrélation empirique : Pour  $n$  grand, on peut montrer que les fonctions de corrélation empiriques sont approximativement i.i.d.  $\mathcal{N}(0, \frac{1}{n})$ . Ainsi 95% de ces auto-corrélations doivent être comprises entre  $-1.96\sqrt{\frac{1}{n}}$  et  $1.96\sqrt{\frac{1}{n}}$ .
2. Le test du portemanteau : On considère la statistique

$$Q = n \times \sum_{j=1}^h \hat{\rho}^2(j)$$

alors si les  $Y_t$  sont i.i.d. on aura  $Q \sim \chi^2(h)$

## Chapitre 3

# Equations récurrentes linéaires

### 3.1 Suite sans second membre

On considère la suite  $h_t, t \in \mathbb{Z}$ , solution de l'équation récurrente d'ordre  $k$  :

$$h_t + \alpha_1 h_{t-1} + \cdots + \alpha_k h_{t-k} = 0, \quad (3.1)$$

où  $\alpha_1, \dots, \alpha_k$  sont des réels constants avec  $\alpha_k \neq 0$ . En introduisant l'opérateur  $B(h_t) = h_{t-1}$ , cette équation peut se noter  $\alpha(B)h_t = 0$ , où  $\alpha(B) = 1 + \alpha_1 B + \cdots + \alpha_k B^k$ .

**Définition** Un ensemble de  $m \leq k$  solutions de l'équation (3.1)(3.1),  $\{h_t^{(1)}, \dots, h_t^{(m)}\}$  seront dites linéairement indépendantes si

$$c_1 h_t^{(1)} + \cdots + c_m h_t^{(m)} = 0 \forall t \in \{0, \dots, k-1\} \Leftrightarrow c_1 = \cdots = c_m = 0.$$

On note que si  $h_t^{(1)}$  et  $h_t^{(2)}$  deux solutions de l'équation (3.1), alors  $c_1 h_t^{(1)} + c_2 h_t^{(2)}$  sera aussi solution. De plus, toutes valeurs initiales  $h_0, \dots, h_{k-1}$  déterminera de façon unique  $h_t$ . Ainsi, si on peut trouver  $k$  solutions  $\{h_t^{(1)}, \dots, h_t^{(k)}\}$  linéairement indépendantes, il existera un unique ensemble de coefficients  $c_1, \dots, c_k$  telle que la solution

$$h_t = c_1 h_t^{(1)} + \cdots + c_k h_t^{(k)}$$

a pour valeurs initiales  $h_0, \dots, h_{k-1}$ . Il suffit donc de trouver  $k$  solutions de (3.1) linéairement indépendantes.

Pour identifier la forme des solutions on peut utiliser les deux théorèmes suivants :

**Théorème 1** Si  $h_t = (a_0 + a_1t + \dots + a_jt^j) m^t$ , avec  $a_0, \dots, a_j, m$  des constantes possiblement dans  $\mathbb{C}$ , alors il existe  $b_0, \dots, b_{j-1}$  tels que

$$(1 - mB) h_t = (b_0 + b_1t + \dots + b_{j-1}t^{j-1}) m^t.$$

**Preuve**

$$(1 - mB) h_t = (a_0 + a_1t + \dots + a_jt^j) m^t - m \left( a_0 + a_1(t-1) + \dots + a_j(t-1)^j \right) m^{t-1},$$

ainsi

$$(1 - mB) h_t = m^t \left( \sum_{r=0}^j a_r (t^r - (t-1)^r) \right)$$

qui est clairement de degré polynomial  $j - 1$  ■

On en déduit le corollaire

**Corollaire** Les fonctions  $h_t^{(j)} = t^j \xi^{-t}, j \in \{0, \dots, k-1\}$  sont  $k$  solutions linéairement indépendantes de l'équation

$$(1 - \xi^{-1}B)^k h_t = 0.$$

**Preuve** On remarque que la fonction  $h_t = 0$  est solution de l'équation. Ainsi, pour toute constante  $C$  possiblement complexe,  $Ch_t^{(1)}$  est solution de l'équation dès que  $k \geq 1$ . De plus, par le théorème précédent, on aura  $\forall j \in \{0, \dots, k-1\}$ ,  $(1 - \xi^{-1}B)^j h_t^{(j)} = C(j) h_t^{(1)}$ , où  $C(j)$  est une constante complexe et  $\forall j \in \{0, \dots, k-1\}$   $h_t^{(j)}$  sera solution de l'équation. Finalement, si

$$(c_0 + c_1t + \dots + c_{k-1}t^{k-1}) \xi^{-t} = 0, \text{ pour } t \in \{0, \dots, k-1\},$$

alors le polynôme  $(c_0 + c_1t + \dots + c_{k-1}t^{k-1})$  a pour  $k$  zéros  $z_1 = 0, \dots, z_k = k-1$ , mais comme il est de degré  $k-1$  ce n'est possible que s'il est identiquement nul, donc  $c_0 = \dots = c_{k-1} = 0$ .

**Solution de l'équation générale** Pour l'équation (3.1), l'opérateur  $\alpha(B)$  peut d'écrire  $\alpha(B) = \prod_{i=1}^j (1 - \xi_i^{-1}B)^{r_i}$ , où  $\xi_i, i \in \{1, \dots, j\}$  sont des zéros distincts de  $\alpha(z)$  et  $r_i$  est la multiplicité de la racine  $\xi_i$ . Par le corollaire précédent,  $t^n \xi_i^{-t}, n \in \{0, \dots, r_i-1\}, i \in \{1, \dots, j\}$  sont  $k$  solutions de l'équation (3.1). On peut montrer que ces solutions sont linéairement indépendantes (cf théorème et corollaire suivant), ainsi une solution générale de l'équation (3.1), sera

$$h_t = \sum_{i=1}^j \sum_{n=0}^{r_i-1} c_{in} t^n \xi_i^{-t}.$$

Pour que  $h_t$  soit réelle, les coefficients correspondant à une paire de racines complexes conjuguées doivent eux mêmes être conjugués. Plus spécifiquement,

si  $(\xi_j, \bar{\xi}_j)$  est une paire de racines conjuguées de  $\alpha(z)$  et  $\xi_j = \rho \exp(i\theta_j)$ , alors le terme correspondant dans  $h_t$  sera

$$\sum_{n=0}^{r_i-1} c_{in} t^n \xi_i^{-t} + \sum_{n=0}^{r_i-1} \bar{c}_{in} t^n \bar{\xi}_i^{-t},$$

qui peut être ré-écrit

$$\begin{aligned} & \sum_{n=0}^{r_i-1} 2 (\operatorname{Re}(c_{in}) \cos(\theta_i t) + \operatorname{Im}(c_{in}) \sin(\theta_i t)) t^n \rho^{-t} = \\ & \sum_{n=0}^{r_i-1} a_{in} (\cos(b_{in}) \cos(\theta_i t) - \sin(b_{in}) \sin(\theta_i t)) t^n \rho^{-t} \end{aligned}$$

avec  $\tan(b_{in}) = \frac{\operatorname{Im}(c_{in})}{\operatorname{Re}(c_{in})}$  et  $a_{in} = 2\sqrt{\operatorname{Re}(c_{in})^2 + \operatorname{Im}(c_{in})^2}$ . Ainsi

$$\sum_{n=0}^{r_i-1} a_{in} t^n \rho^{-t} \cos(\theta_i t + b_{in}).$$

Il reste à prouver la linéaire indépendance des solutions  $t^n \xi_i^{-t}$ ,  $n \in \{0, \dots, r_i - 1\}$ ,  $i \in \{1, \dots, j\}$ .

**Théorème** Si

$$\sum_{l=1}^q \sum_{j=0}^p c_{lj} t^j m_l^t = 0, \quad t \in \mathbb{N}, \quad (3.2)$$

où  $m_1, \dots, m_q$  sont des nombres distincts, alors  $c_{lj} = 0$  pour  $l \in \{1, \dots, q\}$ ,  $j \in \{0, \dots, p\}$ .

**Preuve** Sans perte de généralités, on peut supposer que  $|m_1| \geq |m_2| \geq \dots \geq |m_q| > 0$ . Il suffit donc de montrer que  $c_{1j} = 0$ ,  $j = 0, \dots, p$  car, dans ce cas l'équation (3.2) revient à

$$\sum_{l=2}^q \sum_{j=0}^p c_{lj} t^j m_l^t = 0, \quad t \in \mathbb{N},$$

ce qui impliquera que  $c_{2j} = 0$ ,  $j = 0, \dots, p$  et par une récurrence immédiate  $c_{lj} = 0$  pour  $l \in \{1, \dots, q\}$ ,  $j \in \{0, \dots, p\}$ . Pour prouver que  $c_{1j} = 0$ ,  $j = 0, \dots, p$ , on doit séparer deux cas :

1. Si  $|m_1| > |m_2|$  : En divisant l'équation (3.2) par  $t^p m_1^t$  et en considérant  $t \rightarrow \infty$ , on trouve que  $c_{1p} = 0$ . En remplaçant celui-ci dans l'équation (3.2), en divisant cette équation par  $t^{p-1} m_1^t$  et en considérant  $t \rightarrow \infty$ , on obtiendra  $c_{2p} = 0$  et de proche en proche, on aura  $c_{1j} = 0$ ,  $j = 0, \dots, p$ .

2. Si  $|m_1| = |m_2| = \dots = |m_s| > |m_{s+1}| > 0$ , où  $s \leq q$ . Dans ce cas, on peut écrire  $m_j = re^{i\theta_j}$ , où  $-\pi < \theta_j \leq \pi$  et  $\theta_1, \dots, \theta_s$  tous différents. En divisant l'équation (3.2) par  $t^p r^t$  et en considérant  $t \rightarrow \infty$  on trouve que

$$\sum_{l=1}^s c_{lp} e^{i\theta_l t} \rightarrow 0.$$

Montrons que ce n'est possible que si  $c_{1p} = \dots = c_{sp} = 0$ . Soit  $g_t = \sum_{l=1}^s c_{lp} e^{i\theta_l t}$  et soit  $A_n$ ,  $n \in \mathbb{N}$  la matrice

$$A_n = \begin{pmatrix} e^{i\theta_1 n} & e^{i\theta_2 n} & \dots & e^{i\theta_s n} \\ e^{i\theta_1(n+1)} & e^{i\theta_2(n+1)} & \dots & e^{i\theta_s(n+1)} \\ \vdots & \vdots & \dots & \vdots \\ e^{i\theta_1(n+s-1)} & e^{i\theta_2(n+s-1)} & \dots & e^{i\theta_s(n+s-1)} \end{pmatrix}$$

Comme le déterminant est multilinéaire, on aura  $\det A_n = e^{i(\theta_1 + \dots + \theta_s)n} \det A_0$  et  $|\det A_n| = |\det A_0| > 0$  car la matrice  $A_0$  est une matrice de Vandermonde et a un déterminant non nul. En appliquant la règle de Cramer à l'équation

$$A_n \begin{pmatrix} c_{1p} \\ \vdots \\ c_{sp} \end{pmatrix} = \begin{pmatrix} g_n \\ \vdots \\ g_{n+s-1} \end{pmatrix}$$

on aura

$$c_{1p} = \frac{\det M}{\det A_n}$$

où

$$M = \begin{pmatrix} g_n & e^{i\theta_2 n} & \dots & e^{i\theta_s n} \\ \vdots & \vdots & \dots & \vdots \\ g_{n+s-1} & e^{i\theta_2(n+s-1)} & \dots & e^{i\theta_s(n+s-1)} \end{pmatrix}$$

puisque  $g_n \rightarrow 0$  quand  $n \rightarrow \infty$ , on devra avoir  $c_{1p} = 0$ . En appliquant le même argument à  $c_{2p}, \dots, c_{sp}$  on trouvera  $c_{1p} = \dots = c_{sp} = 0$ .

Maintenant, en divisant (3.2) par  $t^{p-1} r^t$  et en répétant l'argument précédent, on trouvera  $c_{1,p-1} = \dots = c_{s,p-1} = 0$ . Finalement, de proche en proche on trouvera  $c_{lj} = 0$  pour  $l \in \{1, \dots, q\}$ ,  $j \in \{0, \dots, p\}$  ■

On en déduit alors

**Corollaire** Les  $k$  solutions  $t^n \xi_i^{-t}$ ,  $n = 0, \dots, r_i - 1$ ,  $i = 1, \dots, j$  de l'équation (3.1) sont linéairement indépendantes.

**Preuve** On doit montrer que si  $h_t = \sum_{i=1}^j \sum_{n=0}^{r_i-1} c_{in} t^n \xi_i^{-t} = 0$ , pour  $t = 0, \dots, k-1$ , alors tous les  $c_{in}$  sont nuls. Par définition, on a pour tout  $t$ ,  $\alpha(B)h_t = 0$  et  $h_0 = \dots = h_{k-1} = 0$ . Mais par les récursions  $h_t = -\alpha_1 h_{t-1} - \dots - \alpha_k h_{t-k}$ ,  $t = k, k+1, \dots$  et  $\alpha_k h_t = -h_{t+k} - \alpha_1 h_{t+k-1} - \dots - \alpha_{k-1} h_{t+1}$ ,  $t = -1, -2, \dots$  on aura nécessairement  $h_t = 0$  pour tout  $t \in \mathbb{Z}$  et il suffit d'appliquer le théorème précédent avec  $p = \max\{r_1, \dots, r_j\}$ .

### 3.2 Suite avec second membre constant

On considère maintenant la suite  $h_t$ ,  $t \in \mathbb{Z}$ , solution de l'équation récurrente d'ordre  $k$  avec second membre constant :

$$h_t + \alpha_1 h_{t-1} + \cdots + \alpha_k h_{t-k} = \alpha_0$$

On considère deux cas :

1. Si 1 n'est pas racine de  $P(z) = 1 + \alpha_1 z + \cdots + \alpha_k z^k$ , alors il suffit de trouver une constante  $\lambda$  telle que

$$\lambda + \alpha_1 \lambda + \cdots + \alpha_k \lambda = \alpha_0$$

et la solution générale sera de la forme  $h_t + \lambda = c_1 h_t^{(1)} + \cdots + c_k h_t^{(k)} + \lambda$ , où  $h_t^{(j)}$  est de la forme  $t^n \xi_i^{-t}$ .

2. Si 1 est racine d'ordre  $m$ ,  $P(z) = 1 + \alpha_1 z + \cdots + \alpha_k z^k = Q(z)(1-z)^m$ , et il suffit que  $\lambda$  soit tel que  $\lambda t^m + \alpha_1 \lambda (t-1)^m + \cdots + \alpha_k \lambda (t-k)^m = \alpha_0$  pour que la solution générale soit de la forme  $h_t + \lambda t^m = c_1 h_t^{(1)} + \cdots + c_k h_t^{(k)} + \lambda t^m$ , où  $h_t^{(j)}$  est de la forme  $t^n \xi_i^{-t}$ .

# Chapitre 4

## Modèles ARMA

Dans ce chapitre on introduit une famille importante de séries temporelles stationnaires le modèle “Autoregressive moving-average” (ARMA). Ces modèles sont relativement faciles à étudier car ils sont linéaires. De plus pour une large classe de fonction d’auto-covariance  $\gamma(\cdot)$ , il existera un processus ARMA  $(X_t)_{t \in \mathbb{Z}}$ , tel que  $\gamma_X(\cdot) = \gamma(\cdot)$ .

### 4.1 Processus ARMA( $p, q$ )

#### 4.1.1 Définition :

Un processus  $(X_t)_{t \in \mathbb{Z}}$ , est un processus ARMA( $p, q$ ) si pour tout  $t \in \mathbb{Z}$ ,

$$X_t - a_1 X_{t-1} - \dots - a_p X_{t-p} = \varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_q \varepsilon_{t-q} \quad (4.1)$$

où  $\varepsilon_t$  est i.i.d.  $\mathcal{N}(0, \sigma^2)$  et les polynômes  $(1 - a_1 z - \dots - a_p z^p)$  et  $(1 + b_1 z + \dots + b_q z^q)$  n’ont pas de facteurs communs. On dira aussi que  $(X_t)_{t \in \mathbb{Z}}$ , est un ARMA( $p, q$ ) de moyenne  $\mu$  si  $(X_t - \mu)_{t \in \mathbb{Z}}$ , est un ARMA( $p, q$ ).

#### 4.1.2 Existence et unicité

Une solution stationnaire  $(X_t)_{t \in \mathbb{Z}}$  à l’équation (4.1) existe et est unique si et seulement si

$$P(z) = 1 - a_1 z - \dots - a_p z^p \neq 0 \text{ pour tout } z, \text{ tel que } |z| = 1$$

**Causalité** Un ARMA( $p, q$ )  $(X_t)_{t \in \mathbb{Z}}$ , sera dit causal si et seulement si

$$P(z) = 1 - a_1 z - \dots - a_p z^p \neq 0 \text{ pour tout } z, \text{ tel que } |z| \leq 1$$

**Inversibilité** Un processus ARMA( $p, q$ ) sera dit inversible si et seulement si

$$P(z) = 1 + b_1z + \dots + b_qz^q \neq 0 \text{ pour tout } z, \text{ tel que } |z| \leq 1$$

Dans la suite on ne considèrera que des ARMA( $p, q$ ) inversibles et causaux. Dans ce cas la solution stationnaire peut s'écrire sous la forme  $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ . La suite  $(\psi_j)_{j \in \mathbb{N}}$  est déterminée par la relation

$$(1 - a_1z + \dots + a_pz^p) \left( \sum_{j=0}^{\infty} \psi_j z^j \right) = 1 + b_1z + \dots + b_qz^q \Rightarrow$$

$$\sum_{j=0}^{\infty} \psi_j z^j = \frac{1 + b_1z + \dots + b_qz^q}{(1 - a_1z + \dots + a_pz^p)}, \text{ pour } |z| \leq 1$$

ce qui revient à dire :

$$\begin{aligned} 1 &= \psi_0 \\ b_1 &= \psi_1 - \psi_0 a_1 \\ b_2 &= \psi_2 - \psi_1 a_1 - \psi_0 a_2 \\ &\vdots \end{aligned}$$

où de manière équivalente :

$$\psi_j - \sum_{k=1}^p a_k \psi_{j-k} = b_j, \quad j = 0, 1, \dots$$

où  $b_0 := 1$ ,  $b_j := 0$  pour  $j > q$  et  $\psi_j := 0$  pour  $j < 0$ .

D'une manière plus générale les coefficients  $\psi_n$  vérifient l'équation de récurrence pour  $n > q$  :

$$\psi_n - \sum_{k=1}^p a_k \psi_{n-k} = 0$$

Son polynôme caractéristique est

$$P(z) = z^p - a_1z^{p-1} - \dots - a_p = 0$$

Supposons que  $\lambda$  soit une racine de ce polynôme de multiplicité  $r$ , c'est -à-dire que  $(z - \lambda)^r$  divise  $p(z)$ . On aura la propriété suivante :

- chacune des suites  $\lambda^n, n\lambda^n, \dots, n^{r-1}\lambda^n$  satisfait la récurrence.
- Toute suite vérifiant la récurrence peut être écrite de façon unique comme une combinaison linéaire des solutions définies ci-dessus.
- Les coefficients de la combinaison sont donnés par les équations initiales.

## 4.2 ACF et PACF d'un ARMA( $p, q$ )

### 4.2.1 Calcul de ACVF et ACF

Comme on considère que le processus est causal on aura  $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ , d'où

$$\gamma(h) = E(X_{t+h}X_t) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|} \quad (4.2)$$

La fonction d'auto-corrélation s'en déduit par la relation  $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$ .

**Exemple** On considère le processus stationnaire vérifiant l'équation  $X_t - aX_{t-1} = \varepsilon_t + b\varepsilon_{t-1}$ , avec  $|a| < 1$  et  $|b| < 1$ . L'unique solution stationnaire vérifiera pour  $|z| \leq 1$  :

$$\sum_{j=0}^{\infty} \psi_j z^j = \frac{1+bz}{1-az} = (1+az+a^2z^2+\dots)(1+bz)$$

par conséquent  $\psi_0 = 1$  et  $\psi_j = (a+b)a^{j-1}$  pour  $j \geq 1$ . En substituant cette forme dans l'équation (4.2), on trouve que

$$\begin{aligned} \gamma(0) &= \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 \\ &= \sigma^2 \left[ 1 + (b+a)^2 \sum_{j=0}^{\infty} a^{2j} \right] \\ &= \sigma^2 \left[ 1 + \frac{(b+a)^2}{1-a^2} \right] \end{aligned}$$

De plus

$$\begin{aligned} \gamma(1) &= \sigma^2 \sum_{j=0}^{\infty} \psi_{j+1} \psi_j \\ &= \sigma^2 \left[ a+b + (b+a)^2 a \sum_{j=0}^{\infty} a^{2j} \right] \\ &= \sigma^2 \left[ b+a + \frac{a(b+a)^2}{1-a^2} \right] \end{aligned}$$

et

$$\gamma(h) = a^{h-1} \gamma(1), \quad h \geq 2$$

#### 4.2.2 La fonction d'auto-corrélation partielle (PACF)

La fonction d'auto-corrélation partielle d'un processus ARMA  $(X_t)_{t \in \mathbb{Z}}$  est la fonction  $\alpha(\cdot)$  définie par les équations  $\alpha(0) = 1$  et  $\alpha(h) = \phi_{hh}$ , où  $\phi_{hh}$  est le dernier composant de

$$\phi_h = \Gamma_h^{-1} \gamma_h,$$

$\Gamma_h = [\gamma(i-j)]_{i,j=1}^h$  et  $\gamma_h = [\gamma(1), \dots, \gamma(h)]^T$ . De même, pour les observations  $(x_1, \dots, x_n)$  avec  $x_i \neq x_j$  pour tout  $i$  et  $j$  la fonction d'auto-corrélation partielle empirique sera :  $\hat{\alpha}(0) = 1$  et  $\hat{\alpha}(h) = \hat{\phi}_{hh}$ ,  $h \geq 1$ , où  $\hat{\phi}_{hh}$  est le dernier composant de

$$\hat{\phi}_h = \hat{\Gamma}_h^{-1} \hat{\gamma}_h, \quad (4.3)$$

$\hat{\Gamma}_h = [\hat{\gamma}(i-j)]_{i,j=1}^h$  et  $\hat{\gamma}_h = [\hat{\gamma}(1), \dots, \hat{\gamma}(h)]^T$ . Cette fonction est calculée par la machine, car son calcul algébrique est plutôt compliqué, sauf quand  $q$  est nul ou  $p$  et  $q$  sont tous les deux petits. Il peut être montré que  $\alpha(h)$  est la corrélation entre l'erreur de prédiction  $X_h - E(X_h | X_1, \dots, X_{h-1})$  et  $X_0 - E(X_0 | X_1, \dots, X_{h-1})$ . Cette fonction donne une idée de l'ordre  $p$  à utiliser pour la modélisation d'un ARMA( $p, q$ ).

**Exemple : PACF d'un AR(p)** Pour un AR(p) causal définie par

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

alors le meilleur prédicteur de  $X_{h+1}$  en fonction de  $1, X_1, \dots, X_h$  est

$$\hat{X}_{h+1} = a_1 X_h + \dots + a_p X_{h+1-p}$$

comme le coefficient  $\phi_{hh}$  de  $X_1$  est  $a_p$  si  $h = p$  et 0 si  $h > p$ , on en déduit que la PACF  $\alpha(\cdot)$  du processus  $X_t$  aura la propriété  $\alpha(p) = a_p$  et  $\alpha(h) = 0$  si  $h > p$ . Pour  $h < p$  la valeur de  $\alpha(h)$  peut être facilement calculée en utilisant la relation (4.3).

On aura finalement la proposition suivante qui donne la fonction de covariance d'un processus obtenue par filtrage linéaire appliqué à une série stationnaire :

**Proposition** Soit  $Y_t$  une série stationnaire centrée ayant pour fonction de covariance  $\gamma_Y$ . Si  $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ , alors la série temporelle

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} := \Psi(B) Y_t$$

est stationnaire de moyenne 0 et de fonction d'auto-covariance

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(h+k-j) \quad (4.4)$$

**Preuve** On a

$$E(X_t) = E\left(\sum_{j=-\infty}^{\infty} \psi_j Y_{t-j}\right) = \sum_{j=-\infty}^{\infty} \psi_j E(Y_{t-j}) = 0$$

et

$$\begin{aligned} E(X_{t+h} X_t) &= E\left[\left(\sum_{j=-\infty}^{\infty} \psi_j Y_{t-j}\right) \left(\sum_{k=-\infty}^{\infty} \psi_k Y_{t-k}\right)\right] \\ &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k E(Y_{t+h-j} Y_{t-k}) \\ &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(h-j+k) \end{aligned}$$

■

### 4.3 Prévisions d'un processus ARMA

On considère le problème de prédire les valeurs  $X_{n+h}$ ,  $h > 0$  d'une série temporelle stationnaire et centrée, de fonction d'autocovariance  $\gamma_X(\cdot)$  en fonction de  $(X_1, \dots, X_n)$ . Notre but est de trouver une combinaison linéaire

de  $(X_1, \dots, X_n)$  pour prévoir  $X_{n+h}$  en minimisant l'erreur quadratique. Ce meilleur prédicteur linéaire sera noté  $P_n X_{n+h}$  et aura la forme :

$$P_n X_{n+h} = \delta_1 X_n + \dots + \delta_n X_1$$

Les coefficients  $(\delta_1, \dots, \delta_n)$  seront déterminés en minimisant

$$S(\delta_1, \dots, \delta_n) = E(X_{n+h} - \delta_1 X_n - \dots - \delta_n X_1)^2$$

En dérivant cette équation, alors les coefficients  $(\delta_1, \dots, \delta_n)$  vérifient fatalement :

$$E \left[ \left( X_{n+h} - \sum_{i=1}^n \delta_i X_{n+1-i} \right) X_{n+1-j} \right] = 0, j = 1, \dots, n$$

Ces équations se résument à

$$\Gamma_n \delta(n) = \gamma_n(h)$$

où  $\delta(n) = (\delta_1, \dots, \delta_n)^T$ ,  $\Gamma_n = [\gamma(i-j)]_{i,j=1}^n$  et  $\gamma_n(h) = (\gamma(h), \dots, \gamma(h+n-1))$ .

L'erreur de prédiction vérifiera alors

$$E(X_{n+h} - P_n X_{n+h})^2 = \gamma(0) - \delta^T(n) \gamma_n(h)$$

**Propriétés de  $P_n X_{n+h}$**

1.  $P_n X_{n+h} = \sum_{i=1}^n \delta_i X_{n+1-i}$  où  $\delta(n) = (\delta_1, \dots, \delta_n)^T$  vérifie  $\Gamma_n \delta(n) = \gamma_n(h)$ .
2.  $E(X_{n+h} - P_n X_{n+h})^2 = \gamma(0) - \delta^T(n) \gamma_n(h)$
3.  $E[(X_{n+h} - \sum_{i=1}^n \delta_i X_{n+1-i}) X_{n+1-j}] = 0, j = 1, \dots, n$

Pour résoudre le problème de la meilleure prédiction linéaire, il faut résoudre un système à  $n$  équations linéaires. Il peut aussi être intéressant de construire la meilleure prédiction  $P_n X_{n+2}$  sur celle à 1 pas :  $P_n X_{n+1}$  (algorithme récursif). C'est le cas des deux algorithmes suivants :

### 4.3.1 Algorithme de Durbin-Levinson

On sait que d'après la définition de l'autocorrélation partielle que :

$$P_n X_{n+1} = \phi_n^T X(n)$$

où  $X(n) = (X_n, \dots, X_1)$  et  $\phi_n = (\phi_{n1}, \dots, \phi_{nn})^T = \Gamma_n^{-1} \gamma_n$ .

L'erreur quadratique correspondante sera :

$$v_n = E(X_{n+1} - P_n X_{n+1})^2 = \gamma(0) - \phi_n^T \gamma_n$$

et on aura l'algorithme de Durbin-Levinson qui calculera les coefficients  $\phi_{n1}, \dots, \phi_{nn}$  de la façon suivante :

$$\begin{aligned} \phi_{nn} &= \left( \gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j} \gamma(n-j) \right) v_{n-1}^{-1} \\ \begin{bmatrix} \phi_{n1} \\ \vdots \\ \phi_{n,n-1} \end{bmatrix} &= \begin{bmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{bmatrix} - \phi_{nn} \begin{bmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{bmatrix} \\ v_n &= v_{n-1} [1 - \phi_{nn}^2] \end{aligned}$$

où  $\phi_{11} = \frac{\gamma(1)}{\gamma(0)}$  et  $v_0 = \gamma(0)$ .

### 4.3.2 Algorithme d'innovation

Cet algorithme permet d'exprimer la meilleure prédiction linéaire en fonction de la fonction d'autocovariance. Il est valide même si la série n'est pas stationnaire. Notons  $P_n X_{n+1}$  la meilleure prédiction linéaire de  $X_{n+1}$  en fonction de  $1, X_n, X_{n-1}, \dots, X_1$ .

$$P_n X_{n+1} = \alpha_0 + \alpha_1 X_n + \dots + \alpha_n X_1$$

Soit

$$\hat{X}_n = \begin{cases} 0, & \text{si } n = 1 \\ P_{n-1} X_n, & \text{si } n = 2, 3, \dots \end{cases}$$

et  $\nu_n = E(X_{n+1} - P_n X_{n+1})^2$ , les innovations sont  $U_n = X_n - \hat{X}_n$ . En vectorisant :  $\mathbb{U}_n := (U_1, \dots, U_n)^T$  et  $\mathbb{X}_n := (X_1, \dots, X_n)^T$ , on peut écrire

$$\mathbb{U}_n = A_n \mathbb{X}_n$$

où  $A_n$  est la matrice de la forme

$$A_n = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -\alpha_1 & 1 & 0 & \dots & 0 \\ -\alpha_2 & -\alpha_1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ -\alpha_{n-1} & -\alpha_{n-2} & -\alpha_{n-3} & \dots & 1 \end{bmatrix}$$

$A_n$  est non singulière et a pour inverse

$$C_n = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \theta_{11} & 1 & 0 & \dots & 0 \\ \theta_{22} & \theta_{21} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \dots & 1 \end{bmatrix} \quad (4.5)$$

Le vecteur des prévisions à un pas  $\hat{\mathbb{X}}_n := (X_1, P_1 X_2, \dots, P_{n-1} X_n)^T$  peut ainsi être exprimé :

$$\hat{\mathbb{X}}_n = \mathbb{X}_n - \mathbb{U}_n = C_n \mathbb{U}_n - \mathbb{U}_n = \Theta_n (\mathbb{X}_n - \hat{\mathbb{X}}_n) \quad (4.6)$$

où

$$\Theta_n = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \theta_{11} & 0 & 0 & \cdots & 0 \\ \theta_{22} & \theta_{21} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 0 \end{bmatrix}$$

L'équation (4.6), peut alors être réécrite :

$$\hat{X}_{n+1} = \begin{cases} 0, & \text{si } n = 0 \\ \sum_{j=1}^n \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}), & \text{si } n = 1, 2, \dots \end{cases}$$

L'algorithme d'innovation calcule les coefficients  $\theta_{ij}$ , et les erreurs quadratiques moyennes  $\nu_i = E (X_{i+1} - \hat{X}_{i+1})^2$ , en partant des covariance en partant des covariances  $\gamma(i, j)$  :

**Algorithme d'innovation pour un processus éventuellement non-stationnaire :**

Les coefficient  $\theta_{n1}, \dots, \theta_{nn}$  peuvent être calculer récursivement par les équations

$$\begin{aligned} \nu_0 &= V(X_1) \\ \theta_{n,n-k} &= \nu_k^{-1} \left( \gamma(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} \nu_j \right), \quad 0 \leq k < n \end{aligned} \quad (4.7)$$

et

$$\nu_n = \gamma(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 \nu_j$$

**4.3.3 Prévision d'un ARMA**

Introduisons un processus transformé, en notant  $m = \max(p, q)$  :

$$\begin{cases} W_t = \frac{1}{\sigma} X_t, & t = 1, \dots, m \\ W_t = \frac{1}{\sigma} (X_t - a_1 X_{t-1} - \dots - a_p X_{t-p}), & t > m \end{cases}$$

De plus, on définit  $b_0 := 1$  et  $b_j := 0$  pour  $j > q$ . On supposera aussi  $p \geq 1$  et  $q \geq 1$  quitte à choisir des coefficients égaux à 0. Supposons aussi qu'on ait calculé la fonction d'auto-covariance de  $(X_t)_{t \in \mathbb{Z}}$ , dans la pratique, on approxime cette fonction par l'auto-covariance empirique. L'auto-covariance  $\kappa(i, j) = E(W_i W_j)$ ,  $i, j \geq 1$  est alors :

$$\kappa(i, j) = \begin{cases} \frac{1}{\sigma^2} \gamma_X(i-j), & 1 \leq i, j \leq m \\ \frac{1}{\sigma^2} (\gamma_X(i-j) - \sum_{r=1}^p a_r \gamma_X(r - |i-j|)), & \min(i, j) \leq m < \max(i, j) \leq 2m \\ \sum_{r=0}^q b_r b_{r+|i-j|}, & \min(i, j) > m \\ 0, & \text{sinon} \end{cases} \quad (4.8)$$

Appliquant l'algorithme d'innovation à  $W_t$ , on trouve

$$\begin{cases} \hat{W}_{n+1} = \sum_{j=1}^n \theta_{nj} \left( W_{n+1-j} - \hat{W}_{n+1-j} \right), & 1 \leq n < m \\ \hat{W}_{n+1} = \sum_{j=1}^q \theta_{nj} \left( W_{n+1-j} - \hat{W}_{n+1-j} \right), & n \geq m \end{cases}$$

où les coefficients  $\theta_{nj}$  et les erreurs quadratiques moyennes  $r_n = E \left( W_{n+1} - \hat{W}_{n+1} \right)^2$  sont calculés récursivement par l'algorithme d'innovation. On notera que  $\theta_{nj} = 0$  si  $n \geq m$  et  $j > q$ . C'est la conséquence de  $\kappa(r, s) = 0$  si  $r > m$  et  $|r - s| > q$ .

Maintenant, comme  $X_t - \hat{X}_t = \sigma \left( W_t - \hat{W}_t \right)$ , pour tout  $t \geq 1$ , on aura finalement

$$\begin{cases} \hat{X}_{n+1} = \sum_{j=1}^n \theta_{nj} \left( X_{n+1-j} - \hat{X}_{n+1-j} \right), & 1 \leq n < m \\ \hat{X}_{n+1} = a_1 X_n + \dots + a_p X_{n+1-p} + \sum_{j=1}^q \theta_{nj} \left( X_{n+1-j} - \hat{X}_{n+1-j} \right), & n \geq m \end{cases}$$

#### 4.3.4 Prédiction à $h$ pas

On peut montrer que le prédicteur linéaire à  $h$  pas  $P_n X_{n+h}$  en fonction de  $1, X_n, \dots, X_1$  sera

$$\begin{cases} P_n X_{n+h} = \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} \left( X_{n+h-j} - \hat{X}_{n+h-j} \right), & 1 \leq h \leq m - n \\ P_n X_{n+h} = \sum_{i=1}^p a_i P_n X_{n+h-i} + \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} \left( X_{n+h-j} - \hat{X}_{n+h-j} \right), & h \geq m - n \end{cases}$$

Si, comme c'est souvent le cas,  $n > m = \max(p, q)$ , alors pour  $h \geq 1$  :

$$P_n X_{n+h} = \sum_{i=1}^p a_i P_n X_{n+h-i} + \sum_{j=h}^q \theta_{n+h-1,j} \left( X_{n+h-j} - \hat{X}_{n+h-j} \right)$$

#### 4.3.5 Intervalle de confiance

En considérant la décomposition  $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$  et en notant  $\sigma^2(h) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2$ , on aura l'erreur de prédiction  $X_{n+h} - P_n X_{n+h}$  qui sera distribuée selon  $\mathcal{N}(0, \sigma^2(h))$ , ce qui permet de construire des intervalles de confiance.

# Chapitre 5

## Analyse spectrale

### 5.1 Introduction

La représentation spectrale de la série stationnaire  $(X_t)_{t \in \mathbb{Z}}$ , décompose cette série en une somme de composantes sinusoïdales avec des coefficients aléatoires non-corrélés. Il y a une décomposition correspondante pour la fonction d'autocovariance de  $(X_t)_{t \in \mathbb{Z}}$ . La décomposition spectrale est l'analogue, pour les processus stationnaires, de la représentation de Fourier des fonctions déterministes.

### 5.2 Densité spectrale

Soit  $(X_t)_{t \in \mathbb{Z}}$ , un processus stationnaire centré avec pour fonction d'autocovariance  $\gamma(\cdot)$ , satisfaisant  $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$ . La densité spectrale de  $(X_t)_{t \in \mathbb{Z}}$  est la fonction  $f(\cdot)$  définie par

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \gamma(h), \quad -\infty < \lambda < \infty \quad (5.1)$$

où  $e^{i\lambda} = \cos(\lambda) + i \sin(\lambda)$ . La sommabilité de  $|\gamma(\cdot)|$ , implique que la série converge absolument.

**Propriété de  $f$  :**

- $f$  est paire, i.e.  $f(\lambda) = f(-\lambda)$
- $f(\lambda) \geq 0$  pour tout  $\lambda \in ]-\pi, \pi]$

$$\gamma(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda = \int_{-\pi}^{\pi} \cos(k\lambda) f(\lambda) d\lambda \quad (5.2)$$

**Preuve** Puisque  $\sin(\cdot)$  est une fonction impaire et que  $\cos(\cdot)$  et  $\gamma(\cdot)$  sont des fonctions paires, on a

$$\begin{aligned} f(\lambda) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} (\cos(h\lambda) - i \sin(h\lambda)) \gamma(h) \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \cos(h\lambda) \gamma(h) + 0 \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \cos(-h\lambda) \gamma(h) = f(-\lambda) \end{aligned}$$

Pour chaque entier positif  $N$  définissons :

$$f_N(\lambda) = \frac{1}{2\pi N} E \left( \left| \sum_{r=1}^N X_r e^{-ir\lambda} \right|^2 \right) \geq 0$$

On aura

$$\begin{aligned} f_N(\lambda) &= \frac{1}{2\pi N} E \left( \left| \sum_{r=1}^N X_r e^{-ir\lambda} \right|^2 \right) \\ &= \frac{1}{2\pi N} E \left( \sum_{r=1}^N X_r e^{-ir\lambda} \sum_{s=1}^N X_s e^{is\lambda} \right) \\ &= \frac{1}{2\pi N} \sum_{|h| < N} (N - |h|) e^{-ih\lambda} \gamma(h) \end{aligned}$$

comme  $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$ , pour tout  $\frac{\varepsilon}{2}$  il existera  $h_0$  tel que  $\sum_{|h| > h_0} |\gamma(h)| < \frac{\varepsilon}{2}$  et pour  $N > N_0$  on aura

$$\left| \frac{1}{2\pi N} \sum_{|h| < h_0} (N - |h|) e^{-ih\lambda} \gamma(h) - \frac{1}{2\pi} \sum_{|h| < h_0} e^{-ih\lambda} \gamma(h) \right| < \frac{\varepsilon}{2}$$

donc pour  $N > N_0$

$$\left| \frac{1}{2\pi N} \sum_{|h| < h_0} (N - |h|) e^{-ih\lambda} \gamma(h) - \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \gamma(h) \right| < \varepsilon$$

ce qui prouve que  $f_N(\lambda) \rightarrow \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{ih\lambda} \gamma(h) = f(\lambda) \geq 0$ . Finalement, on aura

$$\begin{aligned} \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda &= \int_{-\pi}^{\pi} \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{i(k-h)\lambda} \gamma(h) d\lambda \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \int_{-\pi}^{\pi} e^{i(k-h)\lambda} d\lambda = \gamma(k) \end{aligned}$$

puisque les intégrales  $\int_{-\pi}^{\pi} e^{i(k-h)\lambda} d\lambda$  sont toutes nulles, sauf pour  $k = h$ .

L'équation (5.2) exprime les auto-covariances d'une série stationnaire avec des auto-covariances absolument sommable comme les coefficients de Fourier d'une fonction paire et positive sur  $]-\pi; \pi]$ . Cependant, même si  $\sum_{h=-\infty}^{\infty} |\gamma(h)| = \infty$ , il peut exister une densité spectrale définit ainsi :

**Définition** Une fonction  $f$  est une densité spectrale d'une série stationnaire  $(X_t)$  avec une fonction d'auto-covariance  $\gamma(\cdot)$  si

1.  $f(\lambda) \geq 0$  pour tout  $\lambda \in ]0, \pi]$
2.  $\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda) d\lambda$  pour tout entier  $h$ .

**Remarque** La densité spectrale est unique pour le sens suivant : Si  $f$  et  $g$  sont deux densités spectrales correspondants à une fonction d'auto-covariance  $\gamma(\cdot) : \gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda) d\lambda = \int_{-\pi}^{\pi} e^{ih\lambda} g(\lambda) d\lambda$ , pour tout entier  $h$ . alors  $f$  et  $g$  ont les mêmes coefficients de Fourier et ainsi, sont égales.

La proposition suivante caractérise les densités spectrales :

**Proposition** Une fonction à valeurs réelles, définie sur  $] -\pi, \pi]$ , est la densité spectrale d'une processus stationnaire si et seulement si

1.  $f(\lambda) = f(-\lambda)$
2.  $f(\lambda) \geq 0$
3.  $\int_{-\pi}^{\pi} f(\lambda) d\lambda < \infty$

**Preuve** Si  $\gamma(\cdot)$  est absolument sommable, alors 1), 2) et 3) est une conséquence des propriétés basiques de  $f$ , on peut aussi le montrer dans un cas plus général (cf TSTM).

Réciproquement, supposons que  $f$  satisfait 1), 2) et 3).

En utilisant 1) la fonction définie par  $\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda) d\lambda$  est paire. De plus, si  $a_r \in \mathbb{R}$ ,  $r = 1, \dots, n$ , alors

$$\begin{aligned} \sum_{r,s=1}^n a_r \gamma(r-s) a_s &= \int_{-\pi}^{\pi} \sum_{r,s=1}^n a_r a_s e^{i\lambda(r-s)} f(\lambda) d\lambda \\ &= \int_{-\pi}^{\pi} \left| \sum_{r=1}^n a_r e^{i\lambda r} \right|^2 f(\lambda) d\lambda \geq 0 \end{aligned}$$

Ainsi  $\gamma(\cdot)$  est positive et peut donc être vue comme une fonction de covariance.

**Corollaire** Une fonction absolument sommable  $\gamma(\cdot)$  est la fonction d'auto-covariance d'une série stationnaire si et seulement si elle est paire et

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \gamma(h) \geq 0, \text{ pour tout } \lambda \in ] -\pi, \pi]$$

$f(\lambda)$  est alors la densité spectrale de  $\gamma(\cdot)$ .

**Preuve** La nécessité de la positivité de  $f$  a déjà été établie. Il suffit d'appliquer la proposition précédente pour conclure que  $f$  est la densité spectrale d'une fonction d'auto-covariance. Mais cette fonction doit être  $\gamma(\cdot)$  puisque

$$\gamma(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda$$

pour tout entier  $k$ .

**Exemple** En utilisant le corollaire précédent, il est simple de montrer que la fonction définie par

$$\kappa(h) = \begin{cases} 1, & \text{si } h = 0 \\ \rho, & \text{si } h = 1 \text{ ou } -1 \\ 0 & \text{sinon} \end{cases}$$

est une fonction d'auto-covariance si et seulement si  $|\rho| \leq \frac{1}{2}$ . En effet  $\kappa(\cdot)$  est une ACVF si et seulement si la fonction

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ih\lambda} \gamma(h) = \frac{1}{2\pi} [1 + 2\rho \cos \lambda]$$

est positive pour tout  $\lambda \in ]-\pi, \pi]$  si et seulement si  $|\rho| \leq \frac{1}{2}$ .

Ainsi, le corollaire, précédent fournit un outil puissant pour vérifier si une fonction absolument sommable est une fonction d'auto-covariance.

**Théorème (Représentation spectrale d'une fonction d'auto-corrélation)**

Une fonction  $\gamma(\cdot)$  définie sur les entiers est la fonction d'auto-covariance d'une série stationnaire, si et seulement si il existe une fonction bornée, croissante et continue à droite sur  $] -\pi, \pi]$  avec  $F(-\pi) = 0$  telle que

$$\gamma(h) = \int_{]-\pi; \pi]} e^{ih\lambda} dF(\lambda)$$

pour tout entier  $h$ .  $F$  est symétrique dans le sens où  $\int_{]a, b]} dF(x) = \int_{[-b, -a[} dF(x)$  pour tout  $a$  et  $b$  tels que  $0 < a < b$ .

**Remarque** La fonction  $F$  est une fonction de distribution généralisée dans le sens où  $G(\lambda) = \frac{F(\lambda)}{F(\pi)}$ , est une distribution de probabilité sur  $] -\pi, \pi]$ . Comme  $F(\pi) = \gamma(0) = Var(X_1)$ , la fonction d'auto-corrélation aura la représentation spectrale

$$\rho(h) = \int_{]-\pi; \pi]} e^{ih\lambda} dG(\lambda)$$

La fonction  $F$  est appelée la distribution spectrale de  $\gamma(\cdot)$ . Si on peut écrire  $F(\lambda) = \int_{-\pi}^{\lambda} f(\lambda) d\lambda$ , pour tout  $\lambda \in [-\pi, \pi]$ , alors  $f$  est la densité spectrale et la série temporelle a un spectre continue. Si  $F$  est une distribution discrète (i.e.  $G$  est la distribution d'une loi de probabilité discrète) alors la série temporelle est dites avoir un spectre discret. Par exemple, la série temporelle  $X_t = A \cos(\omega t) + B \sin(\omega t)$ , où  $A$  et  $B$  sont non-corrélés avec moyenne 0 et variance 1 a pour distribution spectrale :

$$F(\lambda) = \begin{cases} 0 & \text{si } \lambda < -\omega \\ 0.5 & \text{si } -\omega \leq \lambda < \omega \\ 1 & \text{si } \lambda \geq \omega \end{cases}$$

**Exemple : Combinaison linéaire de sinusoides** Considérons le processus :

$$X_t = \sum_{j=1}^k (A_j \cos(\omega_j t) + B_j \sin(\omega_j t)), \quad 0 < \omega_1 < \dots < \omega_k < \pi$$

où les variables  $A_1, B_1, \dots, A_k, B_k$  sont non-corrélées,  $E(A_j) = E(B_j) = 0$  et  $Var(A_j) = Var(B_j) = \sigma^2$ . La fonction d'auto-covariance de cette série temporelle est  $\gamma(h) = \sum_{j=1}^k \sigma_j^2 \cos(\omega_j h)$ . Malheureusement, on ne peut pas calculer directement sa densité spectrale car en général  $\sum_{h=-\infty}^{\infty} |\gamma(h)| = \infty$ . Cependant, en utilisant la théorie des distributions (qui dépasse largement le cadre de ce cours) on peut définir sa fonction de distribution spectrale  $F(\lambda) = \sum_{j=1}^k \sigma_j^2 F_j(\lambda)$ , où

$$F_j(\lambda) = \begin{cases} 0 & \text{si } \lambda < -\omega_j \\ 0.5 & \text{si } -\omega_j \leq \lambda < \omega_j \\ 1.0 & \text{si } \omega_j \leq \lambda \end{cases}$$

Et on aura bien

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} dF(\lambda) = \sum_{j=1}^k \sigma_j^2 (0.5e^{-ih\omega_j} + 0.5e^{ih\omega_j}) = \sum_{j=1}^k \sigma_j^2 \cos(\omega_j h).$$

On peut donc généraliser la définition de la densité spectrale à des fonctions d'auto-covariance non sommables, en posant pour la cas des combinaisons élémentaires de sinusoides :

$$f(\lambda) = \sum_{j=1}^k \sigma_j^2 (0.5\delta_{\{-\omega_j\}}(\lambda) + 0.5\delta_{\{\omega_j\}}(\lambda)).$$

En fait, tout processus centré stationnaire peut s'écrire comme la superposition de sinusoides non corrélées, mais en général il s'agit d'un nombre infini de sinusoides plutôt qu'un nombre fini. La généralisation de cette exemple est l'"intégrale stochastique" suivante :

$$X_t = \int_{]-\pi; \pi]} e^{it\lambda} dZ(\lambda)$$

où  $\{Z(\lambda), -\pi < \lambda \leq \pi\}$  est un processus complexe avec des incréments non-corrélés. Cette représentation est appelée représentation spectrale de  $X_t$ . Les intégrales stochastiques sont au-delà du programme de ce cours, mais pour l'exemple des combinaisons linéaire de sinusoides, on peut l'écrire de la façon suivante :

$$dZ(\lambda) = \begin{cases} \frac{A_j + iB_j}{2}, & \text{si } \lambda = -\omega_j \text{ et } j \in \{1, \dots, k\} \\ \frac{A_j - iB_j}{2}, & \text{si } \lambda = \omega_j \text{ et } j \in \{1, \dots, k\} \\ 0, & \text{sinon} \end{cases}$$

On aura ainsi

$$E(dZ(\lambda) dZ^-(\lambda)) = \begin{cases} \frac{\sigma_j^2}{2}, & \text{si } \lambda = \pm\omega_j \\ 0, & \text{sinon} \end{cases}$$

En général, le lien entre  $dZ(\lambda)$  et la fonction de distribution spectrale peut être écrite :

$$E(dZ(\lambda) dZ^-(\lambda)) = \begin{cases} F(\lambda) - F(\lambda-) & \text{pour un spectre discret} \\ f(\lambda) d\lambda & \text{pour un spectre continu} \end{cases}$$

Cette relation montre qu'un grand saut de la fonction de distribution spectrale (ou un grand pic dans la densité spectrale) à une fréquence  $\pm\omega$ , indique la présence dans la série temporelle d'une composante sinusoïdale forte proche de  $\omega$ .

**Exemple : Bruit blanc** Si  $X_t \sim \mathcal{N}(0, \sigma^2)$ , i.i.d. alors  $\gamma(0) = \sigma^2$  et  $\gamma(h) = 0$  pour  $|h| > 0$ . On montre alors que  $f(\lambda) = \frac{\sigma^2}{2\pi}$ ,  $-\pi \leq \lambda \leq \pi$ . Un tel processus est appelé un bruit blanc, car chaque fréquence du spectre contribue également à la variance.

**Exemple : AR(1)** Si  $X_t = aX_{t-1} + \varepsilon_t$ , où  $\varepsilon_t$  est i.i.d.  $\mathcal{N}(0, \sigma^2)$ , alors par la formule (5.1), on aura

$$\begin{aligned} f(\lambda) &= \frac{\sigma^2}{2\pi(1-a^2)} \left(1 + \sum_{h=1}^{\infty} a^h (e^{-ih\lambda} + e^{ih\lambda})\right) \\ &= \frac{\sigma^2}{2\pi(1-a^2)} \left(1 + \frac{ae^{i\lambda}}{1-ae^{i\lambda}} + \frac{ae^{i\lambda}}{1-ae^{-i\lambda}}\right) \\ &= \frac{\sigma^2}{2\pi} (1 - 2a \cos \lambda + a^2)^{-1} \end{aligned}$$

**Exemple : MA(1)** Si  $X_t = \varepsilon_t + b\varepsilon_{t-1}$ , où  $\varepsilon_t$  est i.i.d.  $\mathcal{N}(0, \sigma^2)$ , alors par la formule (5.1), on aura

$$\begin{aligned} f(\lambda) &= \frac{\sigma^2}{2\pi} (1 + b^2 + b(e^{-i\lambda} + e^{i\lambda})) \\ &= \frac{\sigma^2}{2\pi} (1 + 2b \cos \lambda + b^2) \end{aligned}$$

### 5.3 Le périodogramme

Si  $X_t$  est une série stationnaire, avec pour fonction de covariance  $\gamma(\cdot)$  et une densité spectrale  $f(\cdot)$  alors le périodogramme  $I_n(\cdot)$ , fonction des observations  $x_1, \dots, x_n$ , estime  $2\pi f(\cdot)$ . Considérons le vecteur de  $\mathbb{C}^n$  :  $x = (x_1, \dots, x_n)^T$ . Soit  $\omega_k = \frac{2\pi k}{n}$ ,  $k = -\lfloor \frac{n-1}{2} \rfloor, \dots, \lfloor \frac{n}{2} \rfloor$ , on dira que  $e_1, \dots, e_n$  sont orthonormaux si

$$\bar{e}_j^T e_i = \begin{cases} 1, & \text{si } j = k \\ 0, & \text{sinon} \end{cases}$$

Cela implique que  $(e_1, \dots, e_n)$  est une base pour  $\mathbb{C}^n$ , donc le vecteur  $x$  peut s'écrire  $x = \sum_{k=-\lfloor \frac{n}{2} \rfloor}^{\lfloor \frac{n}{2} \rfloor} a_k e_k$ . On peut facilement trouver les coefficients  $a_k$  en

écrivait  $a_k = \bar{e}_k^T x$ . On aura alors

$$a_k = \bar{e}_k^T x = \frac{1}{\sqrt{n}} \sum_{t=1}^n x_t e^{-it\omega_k}$$

La suite  $(a_k)_{1 \leq k \leq n}$  est appelée la transformé de Fourier discrète de la suite  $x_1, \dots, x_n$ .

**Définition** Le périodogramme de  $\{x_1, \dots, x_n\}$  est la fonction

$$I_n(\lambda) = \frac{1}{n} \left| \sum_{t=1}^n x_t e^{-it\lambda} \right|^2$$

**Remarque** Si  $\lambda$  est une des séquence de Fourier  $\omega_k$  alors  $I_n(\omega_k) = |a_k|^2$  et la longueur au carré de  $x$  sera

$$\sum_{t=1}^n |x_t|^2 = \bar{x}^T x = \sum_{k=-\lfloor \frac{n-1}{2} \rfloor}^{\lfloor \frac{n-1}{2} \rfloor} |a_k|^2 = \sum_{k=-\lfloor \frac{n-1}{2} \rfloor}^{\lfloor \frac{n-1}{2} \rfloor} I_n(\omega_k)$$

La valeur du périodogramme à la fréquence  $\omega_k$  est donc la contribution de la fréquence  $\omega_k$  à cette somme des carrés.

La proposition suivante montre que  $I_n(\lambda)$  peut être vu comme l'estimateur de  $2\pi f(\lambda) = \sum_{-\infty}^{\infty} \gamma(h) e^{-ih\lambda}$ .

**Proposition** Si  $x_1, \dots, x_n$  sont des nombres réels et  $\omega_k = \frac{2\pi k}{n}$  les fréquences de Fourier non nulles alors

$$I_n(\omega_k) = \sum_{|h| < n} \hat{\gamma}(h) e^{-ih\omega_k}$$

où  $\hat{\gamma}(h)$  est la fonction de covariance empirique de  $x_1, \dots, x_n$ .

**Preuve** Puisque  $\sum_{t=1}^n e^{-it\omega_k} = 0$  si  $\omega_k \neq 0$ , on peut soustraire la moyenne empirique de  $x$  dans la définition de  $I_n(\lambda)$  :

$$I_n(\omega_k) = \frac{1}{n} \left| \sum_{t=1}^n (x_t - \bar{x}) e^{-it\omega_k} \right|^2 = \frac{1}{n} \sum_{s=1}^n \sum_{t=1}^n (x_s - \bar{x})(x_t - \bar{x}) e^{-i(s-t)\omega_k}$$

d'où

$$I_n(\omega_k) = \sum_{|h| < n} \hat{\gamma}(h) e^{-ih\omega_k}$$

■

On peut montrer que  $I_n(\lambda)$  n'est malheureusement pas un estimateur consistant de  $2\pi f(\lambda)$ , cependant on peut obtenir un estimateur consistant de  $f(\lambda)$  en moyennant le périodogramme dans un petit intervalle de fréquence contenant  $\lambda$ . On aura ainsi

**Définition** Un estimateur moyennés de la densité spectrale  $f(\lambda)$  a la forme

$$\hat{f}(\lambda) = \frac{1}{2\pi} \sum_{|j| \leq m_n} W_n(j) I_n \left( g(n, \lambda) + \frac{2\pi j}{n} \right)$$

ou la fenêtre  $m_n$  satisfait, pour  $n \rightarrow \infty$ ,  $m_n \rightarrow \infty$  et  $\frac{m_n}{n} \rightarrow 0$ . Les fonctions de poids  $W_n(\cdot)$  satisfont par contre

$$W_n(j) = W_n(-j) \geq 0, \text{ pour tout } j,$$

$$\sum_{|j| \leq m_n} W_n(j) = 1$$

et

$$\sum_{|j| \leq m_n} W_n^2(j) \rightarrow 0 \text{ pour } n \rightarrow \infty$$

**Remarque** Les conditions imposées sur les suites  $m_n$  et  $W_n$  assurent la consistance de  $\hat{f}(\lambda)$ . Les conditions sur  $m_n$  signifient simplement que le nombre de termes tend vers l'infini lorsque  $n \rightarrow \infty$ , mais que la largeur de l'intervalle des fréquences sur lequel on moyenne tend vers 0. Les conditions sur  $W_n$  assure que  $\hat{f}_n(\lambda)$  est asymptotiquement sans biais et que sa variance converge vers 0.

**Exemple** Un estimateur vérifiant les conditions précédentes sera

$$\tilde{f}(\lambda) = \frac{1}{2\pi} \sum_{|j| \leq m} \frac{1}{2m+1} I_n \left( g(n, \lambda) + \frac{2\pi j}{n} \right)$$

avec  $m = \sqrt{n}$  et  $g(n, \lambda)$  le multiple de  $\frac{2\pi}{n}$  le plus proche de  $\lambda$ .

## 5.4 Filtres linéaires invariants dans le temps

On a déjà vu l'utilité des Filtres linéaires invariants dans le temps (TLF) pour lisser les données, estimer les tendances ou les composante saisonnière. Un processus linéaire est le résultat d'un TLF appliqué à une bruit blanc. Plus généralement, on dit que  $Y_t$  est le résultat du filtre linéaire  $C = \{c_{t,t-k}, t, k = 0 \pm 1, \dots\}$  appliqué au processus  $X_t$  si

$$Y_t = \sum_{k=-\infty}^{\infty} c_{t,t-k} X_k, \quad t = 0 \pm 1, \dots$$

Le filtre  $C$  est dit être invariant dans le temps si les poids  $c_{t,t-k}$  sont indépendants de  $t$  i.e.  $c_{t,t-k} = \psi_k$ . De plus, le filtre est dit causal si  $\psi_j = 0$  pour  $j < 0$ .

**Exemple**

— Le filtre définie par

$$Y_t = aX_{-t}$$

est linéaire mais n'est pas invariant dans le temps car  $c_{t,t-k} = 0$  sauf si  $2t = k$ , donc  $c_{t,t-k}$  dépend de  $t$ .

— Le filtre

$$Y_t = \frac{1}{2q+1} \sum_{|j| \leq q} X_{t-j}$$

est invariant dans le temps. En effet  $\psi_j = \frac{1}{2q+1}$  si  $j = -q, \dots, q$  et  $\psi_j = 0$  sinon.

La méthode spectrale est particulièrement utile pour décrire le comportement des TLF, grâce à la proposition suivante :

**Proposition** Soit  $X_t$  un série temporelle stationnaire centrée de densité spectrale  $f_X(\lambda)$ . Soit  $\Psi = \{\psi_j, j = 0 \pm 1, \dots\}$ , un TLF tel que  $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ , alors la série temporelle

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j}$$

est stationnaire centrée et a pour densité spectrale

$$f_Y(\lambda) = |\Psi(e^{-i\lambda})|^2 f_X(\lambda) = \Psi(e^{-i\lambda}) \Psi(e^{i\lambda}) f_X(\lambda)$$

où  $\Psi(e^{i\lambda}) = \sum_{j=-\infty}^{\infty} \psi_j e^{-ij\lambda}$ . La fonction  $\Psi(e^{i\lambda})$  est appelée : fonction de transfert du filtre et le module au carré  $|\Psi(e^{i\lambda})|^2$  la fonction puissance du filtre.

**Preuve** En appliquant la formule (4.4), on voit que  $Y_t$  est stationnaire centré et de fonction d'auto-covariance

$$\gamma_Y(h) = \sum_{j,k=-\infty}^{\infty} \psi_j \psi_k \gamma_X(h+k-j)$$

Puisque  $X_t$  a la densité spectrale  $f_X(\lambda)$ , on a

$$\gamma_X(h+k-j) = \int_{-\pi}^{\pi} e^{i(h-j+k)\lambda} f_X(\lambda) d\lambda$$

en substituant cette égalité dans l'égalité précédente, on aura

$$\begin{aligned} \gamma_Y(h) &= \sum_{j,k=-\infty}^{\infty} \psi_j \psi_k \int_{-\pi}^{\pi} e^{i(h-j+k)\lambda} f_X(\lambda) d\lambda \\ &= \int_{-\pi}^{\pi} \left( \sum_{j=-\infty}^{\infty} \psi_j e^{-ij\lambda} \right) \left( \sum_{k=-\infty}^{\infty} \psi_k e^{-ik\lambda} \right) e^{ih\lambda} f_X(\lambda) d\lambda \\ &= \int_{-\pi}^{\pi} e^{ih\lambda} \left| \sum_{j=-\infty}^{\infty} \psi_j e^{-ij\lambda} \right|^2 f_X(\lambda) d\lambda \end{aligned}$$

Cette dernière expression identifie la densité spectrale de  $Y_t$  comme

$$f_Y(\lambda) = |\Psi(e^{-i\lambda})|^2 f_X(\lambda) = \Psi(e^{-i\lambda}) \Psi(e^{i\lambda}) f_X(\lambda)$$

■

**Remarque** La proposition précédente permet d'analyser facilement l'effet de la composition des filtres. Par exemple si  $X_t$  ayant pour densité spectrale  $f_X$ , est transformé par les TLF  $\Psi_1$  et  $\Psi_2$  alors l'effet sera le même que l'application d'un TLF ayant pour fonction de transfert  $\Psi_1(e^{-i\lambda}) \Psi_2(e^{-i\lambda})$  et la densité spectrale du processus  $W_t = \Psi_1(B) \Psi_2(B) X_t$  sera  $|\Psi_1(e^{-i\lambda}) \Psi_2(e^{-i\lambda})|^2 f_X(\lambda)$

**Exemple** On a vu que la différenciation au rang  $d$  permettait de retirer une tendance linéaire et une saison d'ordre  $d$ . La fonction de transfert pour ce filtre est  $1 - e^{-id\lambda}$  qui est 0 pour toutes fréquences qui sont multiples de  $\frac{2\pi}{d}$ . Les composantes avec périodes  $d$  seront donc retirées.

## 5.5 Densité spectrale d'un processus ARMA

La densité spectrale d'un processus ARMA s'obtient facilement à partir de la proposition précédente :

**Proposition** Si  $X_t$  est un ARMA( $p, q$ ) causal vérifiant  $\phi(B) X_t = \theta(B) \varepsilon_t$  alors

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2}, \quad -\pi < \lambda \leq \pi$$

**Preuve**  $X_t$  est obtenue de  $\varepsilon_t$  par application du TLF ayant pour fonction de transfert  $\Psi(e^{-i\lambda}) = \frac{\theta(e^{-i\lambda})}{\phi(e^{-i\lambda})}$ , puisque  $\varepsilon_t$  a pour densité spectrale  $f_\varepsilon(\lambda) = \frac{\sigma^2}{2\pi}$  le résultat suit.

## Chapitre 6

# Modélisation et prévision d'un processus ARMA

L'estimation des paramètres est faite par des méthodes proches du maximum de vraisemblance. Elles sont proches dans le sens où leurs propriétés asymptotiques sont les mêmes. Cependant comme la vraisemblance est une fonction non-linéaire on doit initialiser cette estimation à l'aide d'une estimation préliminaire des paramètres.

La détermination d'un modèle ARMA( $p, q$ ) approprié est une question complexe. Cependant, on a vu que si  $P$  et  $Q$  ont des facteurs en commun le modèle peut ne pas être identifiable. La convergence de la statistique du rapport vers un sup de carré de processus Gaussien n'a été résolu qu'en 1999 par Dacunha-Castelle et Gassiat. La sélection de l'ordre du modèle peut être automatisée grâce à un critère d'information comme BIC (Bayesian Information Criteria) même si les hypothèses pour avoir un bon comportement asymptotique ne sont pas totalement réalistes. Dans ce chapitre on supposera que  $p$  et  $q$  sont connus.

### 6.1 Estimation préliminaire

Dans cette section on considère des techniques préliminaires pour l'estimation des paramètres  $\sigma^2, a_1, \dots, a_p$  et  $b_1, \dots, b_q$ . D'un ARMA( $p, q$ ) causal défini par l'équation :

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

#### 6.1.1 Estimation de Yule-Walker

Comme le processus est causal on peut écrire

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

En multipliant l'équation ARMA par  $X_{t-j}$ ,  $j = 0, 1, \dots, p$ , en prenant l'espérance et en utilisant le développement  $ma(\infty)$ , on obtient :

$$\Gamma_p \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} = \gamma_p := \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(p) \end{pmatrix}$$

et

$$\sigma^2 = \gamma(0) - (a_1, \dots, a_p) \gamma_p$$

où  $\Gamma_p = (\gamma(i-j))_{i,j=1}^p$ . Dans la pratique on calculera ces estimations à partir des covariances empiriques  $\hat{\gamma}(j)$ ,  $j = 0, \dots, p$ , et on obtiendra les équations

$$\hat{\Gamma}_p \begin{pmatrix} \hat{a}_1 \\ \vdots \\ \hat{a}_p \end{pmatrix} = \hat{\gamma}_p := \begin{pmatrix} \hat{\gamma}(1) \\ \vdots \\ \hat{\gamma}(p) \end{pmatrix}$$

et

$$\hat{\sigma}^2 = \hat{\gamma}(0) - (a_1, \dots, a_p) \hat{\gamma}_p$$

Si le modèle est en fait un AR( $p$ ), on aura de plus :

$$\begin{pmatrix} \hat{a}_1 \\ \vdots \\ \hat{a}_p \end{pmatrix} \simeq \mathcal{N} \left( \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix}, \frac{1}{n} \sigma^2 \Gamma_p^{-1} \right)$$

ce qui permet de construire des intervalles de confiance pour les différents paramètres.

### 6.1.2 Algorithme de Burg

Les coefficients de Yules-Walker sont ceux de la meilleure prévision linéaire de  $X_{p+1}$  en fonction de  $X_1, \dots, X_p$ . L'algorithme de Burg estime les autocorrélation partielles (PACF)  $\{\phi_{11}, \phi_{22}, \dots\}$ . Soit une réalisation d'une série stationnaire centrée  $(x_1, \dots, x_n)$ , on définit  $u_i(t)$ ,  $t = i+1, \dots, n$ ,  $0 \leq i \leq n$  comme la différence entre  $x_{n+1-t}$  et la meilleure prévision linéaire de  $x_{n+1-t}$  en fonction des  $i$  observations précédentes. De la même manière, on définit  $v_i(t)$ ,  $t = i+1, \dots, n$ ,  $0 \leq i \leq n$  comme la différence entre  $x_{n+1-t}$  et la meilleure prévision linéaire en fonction des  $i$  observations suivantes. Alors on peut montrer que les erreurs de prédiction forward et backward  $\{u_i(t)\}$  et  $\{v_i(t)\}$  satisfont les récursions :

$$\begin{aligned} u_0(t) &= v_0(t) = x_{n+1-t} \\ u_i(t) &= u_{i-1}(t-1) - \phi_{ii} v_{i-1}(t) \\ v_i(t) &= v_{i-1}(t) - \phi_{ii} u_{i-1}(t-1) \end{aligned}$$

L'estimation de Burg  $\phi_{11}^B$  de  $\phi_{11}$  est faite en minimisant

$$\sigma_1^2 := \frac{1}{2(n-1)} \sum_{t=2}^n (u_1^2(t) + v_1^2(t))$$

par rapport à  $\phi_{11}$  et en utilisant les équations de récurrence. Cela donne des valeurs numériques pour  $u_1(t)$  et  $v_1(t)$  qui peuvent être alors substituer dans les équations de récurrence avec  $i = 2$ . On minimise alors

$$\sigma_2^2 := \frac{1}{2(n-2)} \sum_{t=3}^n (u_2^2(t) + v_2^2(t))$$

par rapport à  $\phi_{22}$  pour obtenir l'estimation de Burg  $\phi_{22}^B$  de  $\phi_{22}$ . On peut montrer que les estimations précédentes sont équivalentes à la résolution des équations suivantes :

$$\begin{aligned} d(1) &= \sum_{t=2}^n (u_0^2(t-1) + v_0^2(t)) \\ \phi_{ii}^B &= \frac{2}{d(i)} \sum_{t=i+1}^n v_{i-1}(t) u_{i-1}(t-1) \\ d(i+1) &= \left(1 - (\phi_{ii}^B)^2\right) d(i) - v_i^2(i+1) - u_i^2(n) \\ (\sigma_i^B)^2 &= \left[ \left(1 - (\phi_{ii}^B)^2\right) d(i) \right] / (2(n-i)) \end{aligned}$$

Les propriétés asymptotiques des coefficients estimés par l'algorithme de Burg pour un processus AR(p) sont les mêmes que celles des estimateurs de Yule-Walker.

### 6.1.3 L'algorithme de Hannan-Rissanen

L'équation d'un AR(p) à la forme d'une régression linéaire, mais pour un ARMA(p,q)  $X_t$  doit être régresser non seulement sur les retard de  $X_t$  mais aussi sur les retards des résidus qui sont inconnus. Néanmoins, il est possible d'appliquer la régression linéaire pour estimer  $(a_1, \dots, a_p)$  puis  $(b_1, \dots, b_q)$  en faisant une régression sur les résidus estimés. C'est la procédure de Hannan-Rissanen :

**Pas 1** Un AR(m), avec  $m$  grand ( $m > \max(p, q)$ ) est estimé par Yule-Walker,  $(\hat{a}_{m1}, \dots, \hat{a}_{mm})^T$  est le vecteur des coefficients estimés. On calcule les résidus grâce aux équations :

$$\hat{\varepsilon}_t = X_t - \hat{a}_{m1} - \dots - \hat{a}_{mm} X_{t-m}, \quad t = m+1, \dots, n$$

**Pas 2** Une fois que les résidus  $\hat{\varepsilon}_t, t = m+1, \dots, n$  sont estimés, le vecteur paramètre  $\beta = (a_1, \dots, a_p, b_1, \dots, b_q)^T$  est estimé par la méthode des moindres carrés, en régressant  $X_t$  sur  $(X_{t-1}, \dots, X_{t-p}, \hat{\varepsilon}_{t-1}, \dots, \hat{\varepsilon}_{t-q}), t = m+1+q, \dots, n$ . Cela nous fournit une estimation  $\hat{\beta}$  qui fournira l'initialisation de l'algorithme de recherche du maximum de vraisemblance. L'estimation de Hannan-Rissanen du bruit blanc sera alors

$$\sigma_{HR}^2 = \frac{\sum_{t=m+1+q}^n (X_t - a_1 X_{t-1} - \dots - a_p X_{t-p} - b_1 \hat{\varepsilon}_{t-1} - \dots - b_q \hat{\varepsilon}_{t-q})^2}{n - m - q}$$

## 6.2 Estimation par maximum de vraisemblance

Soit  $(X_t)_{t \in \mathbb{Z}}$  une série temporelle gaussienne. Soit  $\mathbb{X}_n = (X_1, \dots, X_n)^T$  et  $\hat{\mathbb{X}}_n = (\hat{X}_1, \dots, \hat{X}_n)^T$ , où  $\hat{X}_j = E(X_j | X_1, \dots, X_{j-1})$ . Soit  $\Gamma_n = E[\mathbb{X}_n \mathbb{X}_n^T]$ , en supposant que  $\Gamma_n$  est inversible, la vraisemblance de  $\mathbb{X}_n$  est :

$$L(\mathbb{X}_n) = \frac{1}{2\pi} \frac{1}{\sqrt{\det \Gamma_n}} \exp\left(-\frac{1}{2} \mathbb{X}_n^T \Gamma_n^{-1} \mathbb{X}_n\right)$$

On va montrer que le calcul direct de  $\det \Gamma_n$  et  $\Gamma_n^{-1}$  peut être évité en utilisant l'erreur de prédiction à 1 pas  $X_t - \hat{X}_t$  et sa variance  $\nu_{t-1}$ ,  $t = 1, \dots, n$  qui sont calculés par l'algorithme d'innovation (4.7).

Soit  $\theta_{ij}$ ,  $j = 1, \dots, i$  et  $i = 1, 2, \dots$  les coefficients obtenus quand l'algorithme d'innovation est appliqué à la fonction d'autocovariance normalisée  $\kappa$  de  $(X_t)_{t \in \mathbb{Z}}$  cf (4.8) et soit  $\mathbb{C}_n$  la matrice définie par (4.5). On a l'égalité

$$\mathbb{X}_n = \mathbb{C}_n (\mathbb{X}_n - \hat{\mathbb{X}}_n)$$

Par définition, les composantes de  $\mathbb{X}_n - \hat{\mathbb{X}}_n$  sont non corrélées, ainsi  $\mathbb{X}_n - \hat{\mathbb{X}}_n$  a une matrice de covariance diagonale

$$\mathbb{D}_n = \text{diag}(\nu_0, \dots, \nu_{n-1})$$

on en déduit que

$$\Gamma_n = \mathbb{C}_n \mathbb{D}_n \mathbb{C}_n^T$$

on aura donc

$$\mathbb{X}_n^T \Gamma_n^{-1} \mathbb{X}_n = (\mathbb{X}_n - \hat{\mathbb{X}}_n)^T \mathbb{D}_n^{-1} (\mathbb{X}_n - \hat{\mathbb{X}}_n) = \sum_{j=1}^n \frac{(X_j - \hat{X}_j)^2}{\nu_{j-1}}$$

et

$$\det \Gamma_n^{-1} = (\det \mathbb{C}_n)^2 \det \mathbb{D}_n = \nu_0 \nu_1 \cdots \nu_{n-1}$$

On en déduit la vraisemblance du vecteur  $\mathbb{X}_n$  :

$$L(\mathbb{X}_n) = \frac{1}{\sqrt{(2\pi)^n \nu_0 \nu_1 \cdots \nu_{n-1}}} \exp\left(-\frac{1}{2} \sum_{t=1}^n \frac{(X_t - \hat{X}_t)^2}{\nu_{t-1}}\right)$$

En notant  $m = \max(p, q)$ , la vraisemblance d'un ARMA( $p, q$ ) peut être facilement calculable par l'algorithme d'innovation (cf section 3.3.2) :

$$\begin{cases} P_n X_{n+h} = \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j}), & 1 \leq h \leq m-n \\ P_n X_{n+h} = \sum_{i=1}^p a_i P_n X_{n+h-i} + \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j}), & h \geq m-n \end{cases}$$

et

$$E \left( X_{n+1} - \hat{X}_{n+1} \right)^2 = \sigma^2 E \left( W_{n+1} - \hat{W}_{n+1} \right)^2 = \sigma^2 r_n$$

où  $\theta_{ij}$  et  $r_n$  sont déterminé par l'algorithme d'innovation et la fonction d'auto-covariance normalisé  $\kappa$  de  $(X_t)_{t \in \mathbb{Z}}$  cf (4.8). On obtient finalement

$$L_{a,b,\sigma^2}(\mathbb{X}_n) = \frac{1}{\sqrt{(2\pi\sigma^2)^n r_0 \cdots r_{n-1}}} \exp \left( -\frac{1}{2} \sum_{t=1}^n \frac{(X_t - \hat{X}_t)^2}{r_{j-1}} \right)$$

On trouve alors les estimateurs du maximum de vraisemblance par :

$$\begin{aligned} \hat{\sigma}^2 &= \frac{S(\hat{a}, \hat{b})}{n} \text{ où } S(\hat{a}, \hat{b}) = \sum_{j=1}^n \frac{(X_j - \hat{X}_j)^2}{r_{j-1}} \\ (\hat{a}, \hat{b}) &= \arg \min \ln \left( \frac{S(a,b)}{n} \right) + \frac{1}{n} \sum_{j=1}^n \ln(r_{j-1}) \end{aligned}$$

En notant  $\beta := (a_1, \dots, a_p, b_1, \dots, b_q)^T$ , la minimisation de  $l(\beta) = \ln \left( \frac{S(a,b)}{n} \right) + \frac{1}{n} \sum_{j=1}^n \ln(r_{j-1})$  est faite numériquement (généralement par une méthode du gradient).

### 6.2.1 Contraste de Whittle

En notant  $\theta = (a_1, \dots, a_p, b_1, \dots, b_q, \sigma^2)^T$  Le contraste est Whittle est défini par

$$C(\theta) = \frac{n}{2} \ln(\sigma^2) + \frac{1}{\sigma^2} \sum_j \frac{I_n(\omega_j)}{f_\theta(\omega_j)} + \sigma^2 f \sum_j \ln(f_\theta(\omega_j))$$

où  $I_n$  est le périodogramme et  $\frac{\sigma^2 f_\theta}{2\pi}$  est la densité spectrale du modèle. Les sommes  $\sum_j$  se font sur toutes les fréquences de Fourier  $\omega_j = \frac{\pi j}{n}$ ,  $j \in \left\{ -\frac{int(n-1)}{n}, -\frac{int(n-1)}{n} + \frac{1}{n}, \dots, \frac{int(n)}{n} \right\}$ , où  $int(n)$  est la partie entière de  $n$ . Le calcul de  $C(\theta)$  est très rapide grâce à la transformée de Fourier rapide. On cherche  $\hat{\theta} = \arg \max_{\theta \in \Theta} C(\theta)$  par optimisation numérique. On peut montrer que les propriétés asymptotique de  $\hat{\theta}$  sont les mêmes que celles du maximum de vraisemblance (asymptotique sans biais et asymptotiquement efficace).

### 6.2.2 Régions de confiance pour les coefficients

Pour un nombre d'observation assez grand l'estimateur du maximum de vraisemblance  $\hat{\beta}$  de  $\beta := (a_1, \dots, a_p, b_1, \dots, b_q)^T$  est approximativement gaussien. On aura ainsi :

$$\hat{\beta} \simeq \mathcal{N} \left( \beta, \frac{1}{n} V(\beta) \right)$$

la matrice  $\frac{1}{n} V(\beta)$  peut être approximée par  $2H^{-1}(\beta)$ , où  $H$  est la matrice Hessienne :  $\left[ \frac{\partial^2 l(\beta)}{\partial \beta_i \partial \beta_j} \right]_{i,j=1}^{p+q}$ .

### 6.3 Sélection de l'ordre

Généralement, on choisit l'ordre du modèle grâce au principe de parcimonie (c'est un compromis entre la valeur de la vraisemblance et du nombre de paramètres utilisés pour le modèle).

#### 6.3.1 Critère AICC

On choisit  $p$  et  $q$  de manière à minimiser

$$AICC = -2 \ln L_{a,b,\sigma^2}(X_1, \dots, X_n) + 2 \frac{(p+q+1)n}{n-p-q-2}$$

AICC est introduit pour donner une estimation non biaisé de la distance de Kulback en la vraisemblance du vrai modèle et  $L_{a,b,\sigma^2}(X_1, \dots, X_n)$ . Souvent, l'ordre choisi en minimisant AICC surestime l'ordre du vrai modèle.

#### 6.3.2 Critère BIC

La justification théorique des critères d'information pour la sélection de modèle est un problème très difficile, cependant Dacunha-Castelle et Gassiat (1999) ont montré la statistique de rapport de vraisemblance était bornée en probabilité, en peut alors montrer que un critère de type BIC :

$$T_n(p, q) = -\ln L_{\theta_{(p,q)}}(X_1, \dots, X_n) + (p+q) \ln(n)$$

est consistant (converge en probabilité vers le vrai ordre de l'ARMA). Cela repose sur le résultat suivant :

**Théoreme** Soit  $(\Theta_{(p,q)})_{1 \leq p \leq D, 1 \leq q \leq D}$  est une famille finie, emboîtée de modèles ARMA( $p, q$ ) causaux et stationnaire. Supposons que  $(a_n)_{n \in \mathbb{N}}$  soit une suite croissante qui converge vers l'infini. Supposons de plus que  $\frac{a_n}{n} \xrightarrow[n \rightarrow \infty]{} 0$ . Alors  $(\hat{p}, \hat{q})$  maximisant le critère pénalisé  $T_n(p, q)$  convergera en probabilité vers le vrai modèle de taille minimale

$$(\hat{p}, \hat{q}) \xrightarrow{P} (p^0, q^0)$$

**Preuve** Prouvons d'abord que  $(\hat{p}, \hat{q})$  ne sur-estime pas  $(p^0, q^0)$ .

$$\begin{aligned} \mathbb{P}(\hat{p} > p^0 \text{ et } \hat{q} > q^0) &\leq \sum_{p > p^0, q > q^0} \mathbb{P}(T_n(p, q) > T_n(p^0, q^0)) \\ &= \sum_{p > p^0, q > q^0} \mathbb{P}\left(l_{\hat{\theta}_{(p,q)}}(X_1, \dots, X_n) - a_n \times (p+q) > l_{\hat{\theta}_{(p^0, q^0)}}(X_1, \dots, X_n) - a_n \times (p^0 + q^0)\right) \\ &= \sum_{p > p^0, q > q^0} \mathbb{P}\left(l_{\hat{\theta}_{(p,q)}}(X_1, \dots, X_n) - l_{\hat{\theta}_{(p^0, q^0)}}(X_1, \dots, X_n) > a_n \times (p+q - (p^0 + q^0))\right) \end{aligned}$$

Si on note  $\lambda_n = l_{\hat{\theta}_{(p,q)}}(X_1, \dots, X_n) - l_{\hat{\theta}_{(p^0, q^0)}}(X_1, \dots, X_n)$ , Dacunha-Castelle et Gassiat on montré que,  $\lambda_n \xrightarrow{\mathcal{L}} \mathcal{V}(p, q, p^0, q^0)$ , une variable aléatoire réelle

dont la loi dépend de  $(p, q, p^0, q^0)$  et comme  $a_n \times (p + q - (p^0 + q^0)) \xrightarrow{n \rightarrow \infty} \infty$

$$\mathbb{P}(\lambda_n > a_n \times (p + q - (p^0 + q^0))) \rightarrow 0$$

et  $\mathbb{P}(\hat{p} > p^0 \text{ et } \hat{q} > q^0) \rightarrow 0$ .

Prouvons maintenant que ni  $\hat{p}$  ne sous-estime pas  $p^0$  et ni  $\hat{q}$  ne sous-estime  $q^0$ . Si l'ARMA( $p, q$ ) ne contient pas le vrai modèle de dimension  $(p^0, q^0)$ .

$$\begin{aligned} \mathbb{P}(\hat{p} < p^0) + \mathbb{P}(\hat{q} < q^0) &\leq \sum_q \sum_{p < p^0} \mathbb{P}(T_n(p, q) > T_n(p^0, q^0)) + \sum_p \sum_{q < q^0} \mathbb{P}(T_n(p, q) > T_n(p^0, q^0)) = \\ &\sum_q \sum_{p < p^0} \mathbb{P}\left(l_{\hat{\theta}_{(p, q)}}(X_1, \dots, X_n) - a_n \times (p + q) > l_{\hat{\theta}_{(p^0, q^0)}}(X_1, \dots, X_n) - a_n \times (p^0 + q^0)\right) + \\ &\sum_p \sum_{q < q^0} \mathbb{P}\left(l_{\hat{\theta}_{(p, q)}}(X_1, \dots, X_n) - a_n \times (p + q) > l_{\hat{\theta}_{(p^0, q^0)}}(X_1, \dots, X_n) - a_n \times (p^0 + q^0)\right) = \\ &\sum_q \sum_{p < p^0} \mathbb{P}\left(l_{\hat{\theta}_{(p, q)}}(X_1, \dots, X_n) - l_{\hat{\theta}_{(p^0, q^0)}}(X_1, \dots, X_n) > a_n \times (p + q - (p^0 + q^0))\right) + \\ &\sum_p \sum_{q < q^0} \mathbb{P}\left(l_{\hat{\theta}_{(p, q)}}(X_1, \dots, X_n) - l_{\hat{\theta}_{(p^0, q^0)}}(X_1, \dots, X_n) > a_n \times (p + q - (p^0 + q^0))\right) = \\ &\sum_q \sum_{p < p^0} \mathbb{P}\left(\frac{1}{n} \left(l_{\hat{\theta}_{(p, q)}}(X_1, \dots, X_n) - l_{\hat{\theta}_{(p^0, q^0)}}(X_1, \dots, X_n)\right) > \frac{1}{n} a_n \times (p + q - (p^0 + q^0))\right) + \\ &\sum_p \sum_{q < q^0} \mathbb{P}\left(\frac{1}{n} \left(l_{\hat{\theta}_{(p, q)}}(X_1, \dots, X_n) - l_{\hat{\theta}_{(p^0, q^0)}}(X_1, \dots, X_n)\right) > \frac{1}{n} a_n \times (p + q - (p^0 + q^0))\right) \end{aligned}$$

Si  $p < p^0$  ou  $q < q^0$  le vrai modèle ne peut pas appartenir à l'ARMA( $p, q$ ), ainsi

$$\frac{1}{n} \left(l_{\hat{\theta}_{(p, q)}}(X_1, \dots, X_n) - l_{\hat{\theta}_{(p^0, q^0)}}(X_1, \dots, X_n)\right) \xrightarrow{p.s.} -K(\hat{\theta}_{(p, q)}, \theta^0) < 0.$$

Et comme

$$\frac{1}{n} (a_n \times (p + q - (p^0 + q^0))) \xrightarrow{P} 0,$$

on aura

$$\mathbb{P}(\hat{p} < p^0) \xrightarrow{P} 0 \text{ et } (\hat{q} < q^0) \xrightarrow{P} 0.$$

■

# Chapitre 7

## Modèle SARIMA

On cherche à modéliser un ensemble d'observations  $(x_1, \dots, x_n)$  qui ne semblent plus exactement stationnaires. C'est modèle peuvent avoir une tendance (ARIMA) ou bien aussi une saison (SARIMA). Une fois stationnarisé leur traitement est similaire aux modèles ARMA.

### 7.1 Modèles ARIMA

Une généralisation des modèles ARMA sont les modèles ARIMA : Des modèles qui se réduise à des ARMA une fois différenciés un nombre fini de fois.

#### 7.1.1 Définition

Si  $d$  est un entier non négatif, alors  $(X_t)_{t \in \mathbb{Z}}$  est un modèle ARIMA( $p, d, q$ ) si

$$Y_t := (1 - B)^d X_t$$

est un processus ARMA( $p, q$ ) causal.

Cette définition signifie que  $(X_t)_{t \in \mathbb{Z}}$  vérifie une équation de la forme :

$$\phi^*(B) X_t = \phi(B) (1 - B)^d X_t = \theta(B) \varepsilon_t, (\varepsilon_t)_{t \in \mathbb{Z}} \text{ i. i. d. } \mathcal{N}(0, \sigma^2)$$

où  $\phi(z)$  et  $\theta(z)$  sont des polynômes de degrés  $p$  et  $q$  tels que  $\phi(z) \neq 0$  pour  $|z| < 1$ . Le polynôme  $\phi^*(z)$  a un zéro d'ordre  $d$  à  $z = 1$ .

#### Remarque

- Si  $d \geq 1$ , on peut ajouter une tendance polynomiale de degré  $d - 1$  arbitraire à  $(X_t)$  sans que l'égalité  $\phi^*(B) X_t = \theta(B) \varepsilon_t$  soit modifié. Les modèles ARIMA sont donc très utiles pour représenté des données avec des tendances polynomiales.
- Une caractéristique des modèles ARIMA est que la fonction d'autocorrélation est positive et lentement décroissante, alors que celle d'un ARMA décroît exponentiellement vite.

### 7.1.2 Racine unitaire

Le problème de la racine unitaire est quand le polynôme autorégressif ou moyenne mobile à une racine de norme 1 ou très proche de 1. Par exemple une racine unitaire pour le polynôme autorégressif suggère qu'il faut d'abord différencier la série (marche aléatoire). Une racine unitaire pour la partie moyenne mobile indique que le processus a été trop différencié.

#### 7.1.2.1 Racine unitaire pour la partie autorégressive

Lorsque la fonction d'autocovariance décroît lentement, on peut penser qu'il y a une racine unitaire, il faut alors différencier  $(I - B)^d X_t$ , jusqu'à ce que la fonction d'autocovariance décroisse vite. La série différenciée pourra alors être modélisée par un ARMA(p,q) et le modèle résultant sera un ARIMA(p,d,q) avec  $d$  racines unitaires. Une autre méthode peut être celle de Dickey et Fuller.

Soit  $x_1, \dots, x_n$  des observations de modèle  $AR(1)$  :

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

lorsque  $|a| < 1$  l'estimateur du maximum de vraisemblance de  $a$ , pour  $n$  grand, suit à peu près la loi  $\mathcal{N}\left(a, \frac{1-a^2}{n}\right)$ . En cas de racine unitaire cette approximation n'est plus valable. Pour pouvoir tester :

$$H_0 : a = 1 \text{ contre } H_1 |a| < 1$$

on écrit le modèle comme :

$$\nabla X_t = X_t - X_{t-1} = a_0^* + a_1^* X_{t-1} + \varepsilon_t$$

avec  $a_0^* = \mu(1 - a_1)$  et  $a_1^* = a_1 - 1$ . Soit  $\hat{a}_1^*$  l'estimateur des moindres carrés de  $a_1^*$ . L'écart-type estimé de  $\hat{a}_1^*$  sera

$$\hat{\sigma}(\hat{a}_1^*) = S \sqrt{\sum_{t=2}^n (X_{t-1} - \bar{X})^2}$$

où  $S^2 = \frac{\sum_{t=2}^n (\nabla X_t - a_0^* + a_1^* X_{t-1})}{n-3}$  et  $\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t$ . Dickey et Fuller ont calculé la loi limite quand  $n \rightarrow \infty$  du ratio :

$$\hat{\tau}_\mu := \frac{\hat{a}_1^*}{\hat{\sigma}(\hat{a}_1^*)}$$

les régions de rejets seront les suivantes :

$$\begin{aligned} \alpha = 0.01 & : \hat{\tau}_\mu < -3.43 \\ \alpha = 0.05 & : \hat{\tau}_\mu < -2.86 \\ \alpha = 0.1 & : \hat{\tau}_\mu < -2.57 \end{aligned}$$

Cette procédure peut être étendue au cas où  $(X_t)_{t \in \mathbb{Z}}$  suit un processus AR(p) :

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

Ce modèle peut être réécrit :

$$\nabla X_t = a_0^* + a_1^* X_{t-1} + a_2^* \nabla X_{t-1} + \dots + a_p^* \nabla X_{t-p+1} + \varepsilon_t$$

où  $a_0 = \mu(1 - a_1 - \dots - a_p)$ ,  $a_1^* = \sum_{i=1}^p a_i - 1$  et  $a_j^* = -\sum_{i=j}^p a_i$ ,  $j = 2, \dots, p$ . Si le polynôme associé à la partie autorégressive à une racine unitaire alors  $a_1^* = 0$  et la série différenciée est un AR(p-1). Pour tester la présence d'une racine unitaire il suffit de tester  $a_1^* = 0$ . Comme dans le cas du AR(1)  $a_1^*$  peut être estimé par la méthode des moindres carrés et pour  $n$  grand

$$\hat{\tau}_\mu := \frac{\hat{a}_1^*}{\hat{\sigma}(\hat{a}_1^*)}$$

aura la même distribution que dans le cas AR(1).

### 7.1.3 Prédiction des modèles ARIMA

Si  $d \geq 1$  les premiers moments  $E(X_t)$  et  $E(X_{t+h}X_t)$  ne sont pas déterminés par l'équation différenciée. Il faut des hypothèses supplémentaires pour trouver le meilleur prédicteur linéaire de  $X_{t+1}$ . Par exemple considérons l'ARMA(p,q) causal  $Y_t$  centré et soit  $X_0$  une variable aléatoire. Définissons :

$$X_t = X_0 + \sum_{j=1}^t Y_j, \quad t = 1, 2, \dots$$

alors  $(X_t)_{t \in \mathbb{N}^*}$  est un ARIMA(p, 1, q) de moyenne  $E(X_t) = X_0$  et fonction d'autocovariance  $E(X_{t+h}X_t) - E(X_0)^2$  qui dépend de  $V(X_0)$  et  $Cov(X_0, Y_j)$ ,  $j = 1, 2, \dots$ . Le meilleur prédicteur linéaire de  $X_{n+1}$  en fonction de  $(1, X_0, \dots, X_n)$  est le même que le meilleur prédicteur linéaire en fonction de  $(1, X_0, Y_1, \dots, Y_n)$ . On peut écrire

$$P_n X_{n+1} = P_n (X_0 + Y_1 + \dots + Y_{n+1}) = P_n (X_n + Y_{n+1}) = X_n + P_n Y_{n+1}$$

Pour évaluer  $P_n Y_{n+1}$  il est nécessaire de connaître  $E(X_0 Y_j)$ ,  $j = 1, \dots, n+1$ . Cependant si on suppose que  $X_0$  n'est pas corrélé avec  $(Y_t)_{t \in \mathbb{N}^*}$  alors  $P_n Y_{n+1}$  est uniquement fonction de  $(Y_1, \dots, Y_n)$ .

Nous supposons donc que le processus  $X_t$  satisfait l'équation

$$(I - B)^d X_t = Y_t$$

où  $(Y_t)_{t \in \mathbb{N}^*}$  est un ARMA(p,q) causal et que le vecteur aléatoire  $(X_{1-d}, \dots, X_0)$  n'est pas corrélé avec  $Y_t$ ,  $t > 0$ .

L'équation différenciée peut être réécrite sous la forme :

$$X_t + \sum_{j=1}^d C_d^j (-B)^j X_t = Y_t, \quad t = 1, 2, \dots$$

d'où

$$X_t = Y_t - \sum_{j=1}^d C_d^j (-1)^j X_{t-j} \quad t = 1, 2, \dots$$

Quitte à renuméroter les observations, on supposera que l'on dispose de la série  $X_{1-d}, X_{2-d}, \dots, X_n$  les observations différenciées seront  $Y_1, \dots, Y_n$ . Notre but est de prévoir  $P_n X_{n+h}$ , grâce à la précédente équation, on aura :

$$P_n X_{n+h} = P_n Y_{n+h} - \sum_{j=1}^d C_d^j (-1)^j X_{n+h-j}$$

Comme, par hypothèse,  $(X_{1-d}, \dots, X_0)$  n'est pas corrélé avec  $Y_t$ ,  $t > 0$ , la meilleure prédiction linéaire de  $Y_{n+h}$  peut être calculée en fonction de  $(Y_1, \dots, Y_n)$  et le prédicteur  $P_n X_{n+1}$  peut être obtenu directement par l'équation précédente en notant que  $P_n X_{n+1-j} = X_{n+1-j}$ , pour  $j \geq 1$ . Le prédicteur  $P_n X_{n+2}$  pourra être trouver aussi par l'équation précédente en connaissant  $P_n X_{n+1}$  et de proche en proche  $P_n X_{n+h}$ .

Pour trouver l'erreur quadratique moyenne de la prédiction, il faut exprimer  $P_n Y_{n+h}$  en fonction des  $X_j$ . Pour  $n \geq 0$  on notera la prédiction à un pas  $\hat{Y}_{n+1} = P_n Y_{n+1}$  et  $\hat{X}_{n+1} = P_n X_{n+1}$ , des équations précédentes on aura

$$X_{n+1} - \hat{X}_{n+1} = Y_{n+1} - \hat{Y}_{n+1}, \quad n \geq 1$$

et par l'algorithme d'innovation, si  $n > \max(p, q)$  et  $h \geq 1$  :

$$P_n Y_{n+h} = \sum_{i=1}^p a_i P_n Y_{n+h-i} + \sum_{j=h}^q \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j})$$

Notons

$$a^*(z) = (1-z)^d a(z) = (1-z)^d (1 - a_1 z - \dots - a_p z^p) = 1 - a_1^* z - \dots - a_{p+d}^* z^{p+d}$$

on aura alors :

$$P_n X_{n+h} = \sum_{j=1}^{p+d} a_j^* P_n X_{n+h-j} + \sum_{j=h}^q \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j})$$

l'erreur de prédiction de la prédiction à  $h$ -pas sera :

$$\sigma_n^2(h) = E(X_{n+h} - P_n X_{n+h})^2 = \sum_{j=0}^{h-1} \left( \sum_{r=0}^j \chi_r \theta_{n+h-r-1,j-r} \right)^2 v_{n+h-j-1}$$

où  $\theta_{n0} = 1$ ,  $\sum_{r=0}^{\infty} \chi_r z^r = \frac{1}{(1 - a_1^* z - \dots - a_{p+d}^* z^{p+d})}$  et

$$v_{n+h-j-1} = E(X_{n+h-j} - \hat{X}_{n+h-j})^2 = E(Y_{n+h-j} - \hat{Y}_{n+h-j})^2$$

Pour  $n$  grand on peut approximer l'erreur de prédiction de la prédiction à  $h$ -pas par

$$\sigma_n^2(h) \simeq \sum_{j=0}^{h-1} \psi_j^2 \sigma^2$$

où

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{1 + b_1 z + \dots + b_q(z)}{1 - a_1^* z - \dots - a_{p+d}^* z^{p+d}}$$

### 7.1.4 Fonction de prédiction

Pour  $n > \max(p, q)$ , notons  $g(h) := P_n X_{n+h}$  la fonction de prédiction. Cette fonction vérifiera l'équation :

$$g(h) - a_1^* g(h-1) - \dots - a_{p+d}^* g(h-p-d) = 0, \quad h > q$$

Comme la racine 1 est d'ordre de multiplicité  $d$ , cette équation linéaire de récurrence aura pour solution

$$g(h) = \alpha_0 + \alpha_1 h + \dots + \alpha_{d-1} h^{d-1} + \beta_1 \xi_1^{-h} + \dots + \beta_p \xi_p^{-h}, \quad h > q - p - d$$

où  $(\xi_1, \dots, \xi_p)$  sont les racines (ici supposées distinctes) du polynôme  $1 - a_1 z - \dots - a_p z^p$  et les coefficients  $\alpha_1, \dots, \alpha_d, \beta_1, \dots, \beta_p$  sont déterminés par les  $p+d$  équations initiales qui peuvent être calculée numériquement grâce au équations précédentes.

## 7.2 Modèles SARIMA

Les modèles SARIMA sont des modèles que l'on doit élaguer au pas  $s$  pour les rendre stationnaires. Dans le cas général la définition d'un tel modèle sera :

### 7.2.1 Définition

Si  $d$  et  $D$  sont des entiers positifs, alors  $(X_t)_{t \in \mathbb{N}}$  sera un SARIMA( $p, d, q$ )  $\times$  ( $P, D, Q$ ) $_s$  avec période  $s$  si la série différenciée  $Y_t = (1 - B)^d (1 - B^s)^D X_t$  est un ARMA causal défini par

$$a(B) \phi(B^s) Y_t = b(B) \theta(B^s) \varepsilon_t$$

où  $a(z) = 1 - a_1 z - \dots - a_p z^p$ ,  $b(z) = 1 + b_1 z + \dots + b_q z^q$ ,  $\phi(z) = 1 - \phi_1 z - \dots - \phi_P z^P$ ,  $\theta(z) = 1 + \theta_1 z + \dots + \theta_Q z^Q$ .

**remarque 1** Le processus  $Y_t$  sera causal uniquement si  $a(z) \neq 0$  et  $\phi(z) \neq 0$  pour  $|z| \leq 1$ .

**Remarque 2** L'équation ARMA du modèle pourra être réécrite

$$a^*(B)Y_t = b^*(B)\varepsilon_t$$

où  $a^*(z)$  et  $b^*(z)$  sont des polynômes de degrés  $p + sP$  et  $q + sQ$ . Tous les coefficients de  $a^*(z)$  et  $b^*(z)$  peuvent être exprimés en fonction de ceux de  $a(z)$ ,  $b(z)$ ,  $\phi(z)$  et  $\theta(z)$ . Si  $p < s$  et  $q < s$  les contraintes sur les coefficients de  $a^*(z)$  et  $b^*(z)$  peuvent être exprimer par :

$$\begin{cases} a_{is+j}^* = a_{is}^* a_j^*, i = 1, 2, \dots; j = 1, \dots, s-1 \\ \text{et} \\ b_{is+j}^* = b_{is}^* b_j^*, i = 1, \dots; j = 1, \dots, s-1 \end{cases}$$

### 7.2.2 Prédiction des SARIMA

Les prédictions des SARIMA est tout à fait analogue à celle des ARIMA. En développant l'opérateur  $(1 - B)^d (1 - B^s)^D$  et en réarrangeant l'équation

$$(1 - B)^d (1 - B^s)^D X_t = Y_t$$

et en posant  $t = n + h$  on obtiendra

$$X_{n+h} = Y_{n+h} + \sum_{j=1}^{d+Ds} a_j X_{n+h-j}$$

En supposant que les  $d + Ds$  premières observations  $X_{-d-Ds+1}, \dots, X_0$  ne sont pas corrélées avec  $Y_t$ ,  $t \geq 1$ , on peut déterminer la meilleure prédiction linéaire  $P_n X_{n+h}$  en fonction de  $(X_{-d-Ds+1}, \dots, X_n)$  en appliquant  $P_n$  à chaque coté de l'équation précédente pour obtenir

$$P_n X_{n+h} = P_n Y_{n+h} + \sum_{j=1}^{d+Ds} a_j P_n X_{n+h-j}$$

Le premier des termes de droite est juste la meilleure prédiction linéaire du processus ARMA en fonction de  $(Y_1, \dots, Y_n)$  et les prédictions  $P_n X_{n+h}$  peuvent être calculée de proche en proche à partir de l'équation précédente, en remarquant que  $P_n X_{n+1-j} = X_{n+1-j}$ ,  $j \geq 1$ .

Un argument similaire à la section précédente nous donnera l'erreur de prédiction moyenne :

$$\sigma_n^2(h) = E(X_{n+h} - P_n X_{n+h})^2 = \sum_{j=0}^{h-1} \left( \sum_{r=0}^j \chi_r \theta_{n+h-r-1, j-r} \right)^2 v_{n+h-j-1}$$

où  $\theta_{nj}$  et  $v_n$  sont obtenus en appliquant l'algorithme d'innovation à la série différenciée  $Y_t$  et

$$\sum_{r=0}^{\infty} \chi_r z^r = \frac{1}{(a(z)\phi(z)(1-z)^d(1-z^s)^D)}, |z| < 1$$

Pour  $n$  grand on peut approximer l'expression de l'erreur de prédiction moyenne par :

$$\sigma_n^2(h) = \sum_{j=0}^{h-1} \psi_j^2 \sigma^2$$

où

$$\sum_{j=0}^{\infty} \psi_j z^j = \frac{b(z) \theta(z^s)}{a(z) \phi(z^s) (1-z)^d (1-z^s)^D}, \quad |z| < 1$$