Analyse spectrale et séries chronologiques
M1 BIBS Orsay – 2010-2011

Cours: Raphaël Rossignol
TP et TD: Thierry Dumont

19 mai 2011

Table des matières

Table des matières iii

Introduction 1

I Description temporelle des séries chronologiques 3

1 Covariance, stationnarité 7
1.1 Rappels . . . 7
1.2 Autococorrélation et corrélation croisée . . . 9
1.3 Stationnarité empirique . . . 10

2 Tendance et saisonnalité 13
2.1 Notion de tendance, lissage d'une série chronologique . . . 13
2.2 Phénomènes saisonniers . . . 13

3 Stationnarisation 15
3.1 Modèle additif . . . 15
3.2 Transformations de Box-Cox . . . 15
3.3 Différentiation . . . 16
3.4 Première description d'une série chronologique . . . 16

II Les ARMA : des bruits stationnaires autocorrélés 17

4 Processus stationnaires 21
4.1 Bruit blanc . . . 21
4.2 Série chronologique aléatoire stationnaire . . . 22
4.3 Retour sur la notion de stationnarité empirique . . . 23

5 Filtrage linéaire et processus ARMA 25
5.1 Filtrage linéaire . . . 25
5.2 Les processus ARMA . . . 26
5.3 Rappels sur les nombres complexes . . . 27
5.4 Manipuler les séries d'opérateurs retard . . . 28
5.5 Existence, causalité, inversibilité . . . 29
5.6 Calcul de la covariance théorique . . . 31
5.7 Simulation . . . 32
5.8 Moyenne, autocovariance et autocorrélation empiriques . . . 33

6 Prédiction linéaire **35**
6.1 Rappels sur la projection orthogonale dans $\mathbb{R}^d$ . . . 35
6.2 Prédiction linéaire . . . 35
6.3 Application aux processus SLC . . . 38
6.3.1 Prédiction à 1 pas . . . 38
6.3.2 Prédiction à h pas . . . 39
6.4 Projection sur tout le passé pour un ARMA causal et inversible . . . 40
6.5 Autocorrélation partielle . . . 41
6.6 Caractérisation des AR et des MA . . . 42

7 Statistique sur les ARMA **45**
7.1 Moyenne, autocovariance et autocorrélation empiriques . . . 45
7.1.1 Moyenne empirique . . . 45
7.1.2 Autocovariance et autocorrélation empiriques . . . 45
7.2 Estimation des paramètres d'un ARMA d'ordre connu . . . 46
7.2.1 Méthode des moments pour un AR . . . 46
7.2.2 Estimation d'un ARMA gaussien par maximum de vraisemblance . . . 47
7.2.3 Pseudo-vraisemblance gaussienne . . . 47
7.3 Validation . . . 48
7.4 Détermination de l'ordre d'un ARMA . . . 48
7.4.1 Détermination de l'ordre d'un AR . . . 48
7.4.2 Détermination de l'ordre d'un MA . . . 49
7.4.3 Détermination de l'ordre d'un ARMA . . . 49
7.5 Prédiction de valeurs futures . . . 49

III Modèles de séries chronologiques non stationnaires **51**

8 Modèles d'état et filtre de Kalman **55**
8.1 Modèle d'état . . . 55
8.2 Le filtre de Kalman . . . 56
8.3 Filtre de Kalman et vraisemblance gaussienne . . . 59

9 Variables explicatives et modèle à fonction de transfert **61**
9.1 Régression à résidu ARMA . . . 61
9.2 Modèle à fonction de transfert . . . 62
9.2.1 Modèle d'état . . . 63
9.2.2 Identification . . . 64

10 ARIMA et SARIMA **65**
10.1 ARIMA . . . 65
10.1.1 Définition . . . 65
10.1.2 Identification . . . 67
10.2 SARIMA . . . 67
10.2.1 Définition . . . 67
10.2.2 Identification . . . 68
10.3 Pour aller plus loin . . . 70

IV Analyse spectrale **71**

11 Signaux numériques et changement de base **75**
11.1 Quelques notions d'algèbre linéaire 75
11.1.1 Signaux réels ou complexes, changement de base 75
11.1.2 Bases orthonormées 76
11.2 Effet d'un changement de base orthonormale sur un bruit blanc 76

12 Bases de Fourier finies **77**
12.1 Transformée de Fourier Discrète (TFD) 77
12.1.1 Rappels sur les complexes 77
12.1.2 Définition de la TFD 77
12.1.3 Propriétés de la TFD 79
12.2 Effet de la TFD sur les ARMA 79
12.3 Phénomène de Gibbs et effets de bord 80
12.4 Transformée en cosinus discret 82

13 Transformée de Fourier à court terme **85**
13.1 Transformée de Fourier à court terme 85
13.2 Transformée de Gabor 86
13.2.1 Notion de repère 87
13.3 Pour aller plus loin 88

Bibliographie **89**

Index **90**

Introduction

Ce cours est une introduction à l'étude mathématique des *séries chronologiques*, c'est à dire des suites finies de mesures régulièrement espacées dans le temps, à valeurs réelles.

Décrivons quelques exemples :

- la croissance d'un arbre est fonction de son environnement, notamment de la température, de l'humidité etc. mais aussi de facteurs individuels. Peut-on modéliser le lien entre la largeur de ses *cernes* successives et ces facteurs environnementaux ?
- selon l'état du patient (éveil, sommeil lent léger, sommeil lent profond, sommeil paradoxal, état pathologique), les caractéristiques de son *électroencéphalogramme* (EEG) varient. La Figure 1 montre un extrait d'un enregistrement d'EEG, présentant une transition entre deux états au cours d'une crise d'épilepsie. Le partitionnement de l'EEG peut être fait

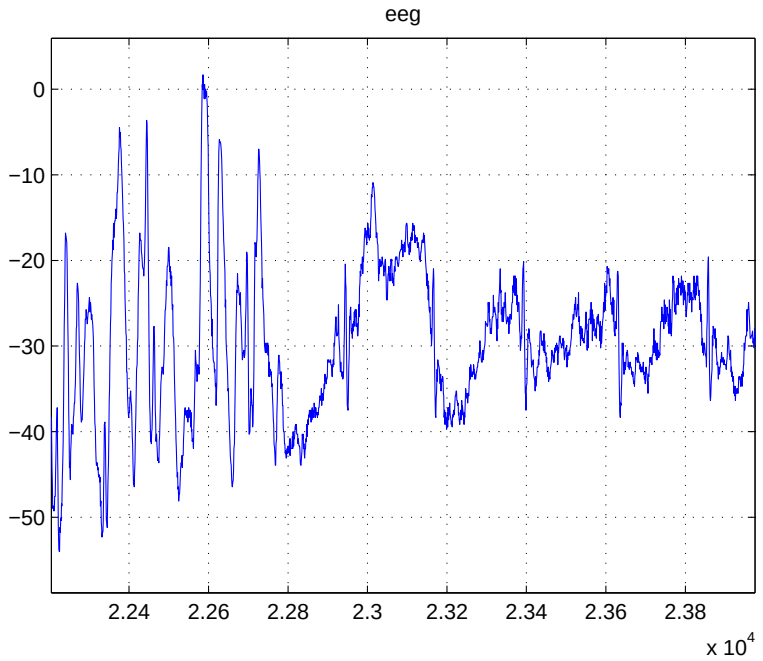

Figure 1 – Transition dans un EEG

manuellement par un opérateur spécialiste, mais cette opération est très fastidieuse pour des enregistrements longs (par exemple un partitionnement de huit heures d'EEG par tranches de cinq secondes) et peut conduire à des erreurs. On cherche donc à mettre en place un partitionnement automatique basé sur le fait que le modèle de l'EEG change de paramètres lorsque le patient change d'état, voir par exemple [1]. Ce dernier exemple illustre la parenté entre le traitement du signal et l'étude des séries chronologiques, un signal, électrique ici, pouvant être considéré comme une série chronologique.

Dans tous ces exemples intervient le *hasard* : il nous manque toujours de l'information, et la compréhension qe l'on peut avoir du phénomène est entachée d'un *bruit*. Ceci explique

l'*approche probabiliste et statistique* adoptée dans ce cours. Pour fixer les idées, on aura donc souvent en tête une modélisation de nos observations $x_0, \ldots, x_{n-1}$ sous la forme :

$$x_t = f(t) + \varepsilon_t \ ,$$

avec f une fonction "régulière" et ε un "bruit", c'est à dire une composante aléatoire, irrégulière. Les objectifs classiques sont en général d'*estimer* la fonction f ou de *prédire* les valeurs futures de x.

Ceci nécessite de contrôler statistiquement le bruit ε, ce qui n'est faisable qu'au prix d'hypothèses sur sa nature aléatoire. Rappelons-nous l'hypothèse du modèle linéaire gaussien classique : les valeurs successives du bruit ε sont alors indépendantes est identiquement distribuées (et gaussiennes, mais ceci est secondaire pour l'instant). Cette hypothèse, raisonnable lorsqu'on observe des résultats répétés d'expériences indépendantes, est très forte et en général irréaliste lorsqu'on observe une série chronologique : il y a en général *corrélation* entre différentes valeurs successives du bruit. Néanmoins, cette corrélation est souvent plus faible entre deux instants éloignés. Par ailleurs, il est nécessaire d'apprendre quelles sont les *formes* fréquentes pour f. La partie II de ce cours sera consacrée à la description temporelle des séries chronologiques, en cherchant à décrire la partie régulière en terme de *tendance* (une évolution lente et régulière) et *saisonnalité* (un phénomène périodique), et à décrire le bruit en terme de *stationnarité* (régularité statistique dans le temps), et de *corrélation*.

Dans la partie II, on accomplira un objectif majeur du cours qui est d'introduire un modèle de bruit stationnaire et *dépendant* appelé $ARMA$. Notamment, un $ARMA$ est facilement contrôlable d'un point de vue statistique.

Ce modèle de bruit servira de brique élémentaire à la construction de modèles plus généraux de séries chonologiques aléatoires non stationnaires dans la partie III.

Enfin, dans la partie IV, on changera légèrement de point de vue pour introduire une idée très importante en traitement du signal : l'idée de changement de base. Les signaux (ou séries chronologiques) de longueur n sont, d'un point de vue mathématique, des vecteurs dans un espace de grande dimension, $\mathbb{R}^n$, et pour "comprendre" ces signaux, les apprivoiser, les examiner, il faut pouvoir "tourner autour", comme on tournerai autour d'une sculpture pour en saisir tous les aspects. Tourner autour d'une sculpture, c'est changer de repère dans l'espace $\mathbb{R}^3$. Tourner autour d'un signal, c'est aussi changer de repère (ou simplement de base), mais dans $\mathbb{R}^n$. Certaines bases sont plus utiles que d'autres, et on verra l'exemple d'une base très importante : la *base de Fourier*. Examiner un signal ou une série chronologique dans la base de Fourier, c'est ce qu'on appelle faire son *analyse spectrale*.

Pour finir cette introduction, soulignons que l'objectif principal de ce cours est de comprendre et de savoir manipuler mathématiquement les notions suivantes :

- la stationnarité,
- les bruits corrélés,
- les modèles d'états,
- le changement de base pour "observer un signal sous un angle différent" avec la transformée de Fourier.

Enfin, citons quelques références. Le livre de Brockwell et Davis [3] contient toute la théorie nécessaire, et même plus. Son "frère" [2] est souvent à examiner en premier car il est plus simple d'abord. L'ouvrage de Hipel et Mc Leod [5] est orienté vers les applications environnementales, surtout hydrologiques. Il est très clair et donne beaucoup de cas d'études à des données réelles. Concernant la partie "analyse spectrale" du cours, on pourra consulter le chapitre 1 du cours de Bruno Torrésani [8]. Enfin, quelques notes de cours en français trainent sur le web, par exemple celles de Xavier Guyon [4] ou de Michel Prenat[6].

Première partie

Description temporelle des séries chronologiques

Lorsqu'on observe une série chronologique pour la première fois, il faut bien arriver à la décrire, en un sens *à décrire sa forme*. Cette étape de description est assez informelle, mais cruciale pour la suite de l'analyse. C'est en effet le "degré zéro" de la modélisation.

Le but de cette partie est de donner quelques notions fondamentales sur cette étape, ainsi que quelques outils mathématiques bien utiles pour aider nos yeux et préparant l'introduction des principaux modèles qui seront étudiés dans ce cours.

Comme le titre de cette partie l'indique, on restera dans le domaine temporel. Les idées de changement de base, avec l'analyse de Fourier, feront l'objet de la partie IV de ce cours.

Chapitre 1

Covariance, stationnarité

1.1 Rappels

Dans la suite, on supposera données deux séries d'observations de même longueur : $x = (x_1, \ldots, x_n)$ et $y = (y_1, \ldots, y_n)$.

On considère le nuage de points $(x_t, y_t)_{t=1,\ldots,n}$. Si les observations x et y n'ont aucun lien entre elles, on obtient un nuage de points dispersés tout à fait au hasard entre les axes. Au contraire, s'il existe une dépendance entre x et y, le nuage de points prend une forme particulière (elliptique par exemple) et le nuage est d'autant plus étroit que la liaison statistique entre x et y est plus serrée (dépendance forte). Introduisons à présent les outils nécessaires pour mesurer cette dépendance.

Définition 1.1 *La moyenne empirique d'une série x est :*

$$\overline{x} = \frac{1}{n}\sum_{t=1}^{n} x_t.$$

La variance empirique de la série x est :

$$s_{x,n}^2 = \frac{1}{n}\sum_{t=1}^{n}(x_t - \overline{x})^2 = \frac{1}{n}\sum_{t=1}^{n} x_t^2 - \overline{x}^2,$$

l'écart type empirique de la série est $s_{x,n}$.

Remarque 1.2 *La moyenne empirique est une notion de "tendance centrale". Une autre mesure de tendance centrale très utile est la* médiane empirique*, plus robuste aux valeurs extrêmes. Il y a en fait plusieurs définition voisines de* la *médiane empirique. On pourra utiliser la suivante. On note $x_{(1)} \leqslant \ldots \leqslant x_{(n)}$ la série réordonnée. Si $n = 2m+1$ est impair,*

$$\mathrm{med}_n = x_{(m)} \ .$$

Si $n = 2m$ est pair,

$$\mathrm{med}_n = \frac{1}{2}(x_{(m)} + x_{(m+1)} \ .$$

R*: mean() et median(), éventuellement avec l'option "na.rm=TRUE".*

Remarque 1.3 *Souvent, la variance empirique est définie par :*

$$\sigma_{x,n}^2 = \frac{1}{n-1}\sum_{t=1}^{n}(x_t - \overline{x})^2 \ .$$

L'intérêt de cette définition est que cela donne un estimateur non biaisé *de la variance σ^2 lorsque les x_i sont des réalisations de n variables aléatoires décorrélées de variance σ^2. Dans un contexte descriptif, il n'y a pas d'intérêt majeur à utiliser $\sigma^2_{x,n}$ plutôt que $s^2_{x,n}$.*

R*: var(x) donne la valeur de $\sigma^2_{x,n}$, et sd(x) la valeur de $\sigma_{x,n}$.*

On définit de la même manière $\overline{y}$ et $s_{y,n}$. On rappelle que l'écart-type empirique indique la dispersion des observations autour de leur moyenne. On définit alors :

Définition 1.4 *la covariance empirique de x et y :*

$$\mathsf{cov}(x,y) = \frac{1}{n}\sum_{t=1}^{n} x_t y_t - \overline{x}\,\overline{y} = \frac{1}{n}\sum_{t=1}^{n}(x_t-\overline{x})(y_t-\overline{y})$$

et le coefficient de corrélation de x et y :

$$\mathrm{corr}(x,y) = \frac{\mathsf{cov}(x,y)}{s_{x,n}s_{y,n}}.$$

Remarque 1.5 *De même, la covariance empirique est parfois définie par :*

$$\mathsf{cov}(x,y) = \frac{1}{n-1}\sum_{t=1}^{n} x_t y_t - \overline{x}\,\overline{y} = \frac{1}{n-1}\sum_{t=1}^{n}(x_t-\overline{x})(y_t-\overline{y})\ .$$

C'est ce que retourne R lorsqu'on lui demande `cov(x,y)`*. Mais dans tous les cas, le coefficient de corrélation est donné par la formule :*

$$= \frac{\mathsf{cov}(x,y)}{s_{x,n}s_{y,n}}.$$

R*: cor(x,y) donne la valeur de* $\mathrm{corr}(x,y)$.

On a la proposition suivante :

Proposition 1.6 *Pour toutes séries x et y,*
- $|\mathrm{corr}(x,y)| \leqslant 1$,
- $|\mathrm{corr}(x,y)| = 1 \Leftrightarrow x$ *et y sont liés (i.e. il existe λ et μ tels que soit pour tout t, $x_t = \lambda y_t + \mu$ soit pour tout t, $y_t = \lambda x_t + \mu$).*

Preuve : il suffit d'appliquer l'inégalité de Cauchy-Schwartz pour le produit scalaire (forme bilinéaire réelle symétrique définie positive)

$$\langle a,b\rangle = \frac{1}{n}\sum_{t=1}^{n} a_t b_t. \tag{1.1}$$

Rappelons que l'inégalité de Cauchy-Schwartz se démontre en développant pour tout réel λ, $\|\lambda a + b\|^2 \geqslant 0$. □

Etant donné notre nuage de point, il peut parfois être utile de connaître la droite qui "approche le mieux" ce nuage de points. Pour nous, elle sera définie de la manière suivante et appelée **droite de régression**.

Définition 1.7 *La droite de régression est la droite $y = ax + b$ qui approche le mieux le nuage de points au sens des moindres carrés, c'est à dire que a et b sont choisis pour que*

$$\sum_{t=1}^{n}(ax_t + b - y_t)^2 \tag{1.2}$$

soit minimale.

On a alors le résultat suivant :

Proposition 1.8 *Si $y = ax + b$ est la droite de régression de $(x_t, y_t)_{t=1,\dots,n}$, alors*

$$a = \frac{\mathsf{cov}(x,y)}{s_{x,n}^2}, \quad b = \overline{y} - a\overline{x}.$$

Preuve : Il suffit de dériver par rapport à a et b l'expression (1.2). □

1.2 Autococorrélation et corrélation croisée

On va maintenant utiliser la notion de covariance d'une manière spécifique aux séries chronologiques. Tout d'abord en mesurant la dépendance à l'intérieur d'une série d'observations avec l'*autocovariance empirique*, puis entre deux séries chronologiques avec la *corrélation croisée*.

Pour mesurer la "dépendance" entre les observations d'une même série $x = (x_1, \dots, x_n)$, on peut utiliser la *fonction d'autocovariance empirique* :

Définition 1.9 *La* fonction d'autocovariance empirique *de la série x est :*

$$\forall\, h \in \{0, 1, \dots, n-1\}, \quad g_{x,n}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(x_t - \overline{x})(x_{t+h} - \overline{x}).$$

Remarque 1.10 *On remarque que $g_{x,n}(0) = s_{x,n}^2$.*

Enfin, on donne la définition de la *fonction d'autocorrélation empirique.*

Définition 1.11 *La* fonction d'autocorrélation empirique *(ACF) pour la série x est :*

$$\forall\, h \in \{0, 1, \dots, n-1\}, \quad \rho_{x,n}(h) = \frac{g_{x,n}(h)}{g_{x,n}(0)} = \frac{\sum_{t=1}^{n-h}(x_t - \overline{x})(x_{t+h} - \overline{x})}{\sum_{t=1}^{n}(x_t - \overline{x})^2}.$$

R: C'est la fonction acf() qui se charge du calcul de l'autocovariance (option "type=cov") et de l'autocorrélation empiriques. Bien évidemment, $\rho_{x,n}(0) = 1$, si bien qu'en passant des fonctions d'autocovariance empirique aux fonctions d'autocorrélation empirique, on perd une information : la dispersion de la série autour de sa moyenne. Ceci dit, ce sont les fonctions d'autocorrélation empirique qui caractérisent les dépendances.

Remarque 1.12 *Pour voir ce dernier aspect, il faut remarquer que quand $p << n$, sous des conditions très générales, $\rho_{x,n}(p)$ est presque égal au coefficient de corrélation entre la série $(x_1, x_2, \dots, x_{n-p})$ et $(x_{p+1}, x_{p+2} \dots, x_n)$.*

Il faut noter que pour les indices p élevés, la valeur de $\rho_{x,n}(p)$ n'a pas de signification statistique. Le bon sens ainsi que des considérations de convergence dépassant les limites de ce cours, recommandent de ne prendre en considération que les autocovariances empiriques $\rho_{x,n}(1), \rho_{x,n}(2), \dots, \rho_{x,n}(h)$ pour un h pas trop "grand" par rapport à la longueur de la série. La même remarque s'applique pour la fonction d'autocorrélation empirique.

Définition 1.13 *La* fonction d'autocorrélation croisée *(CCF) pour les séries x et y est :*

$$\forall\, h \in \{0, 1, \dots, n-1\}, \quad \rho_{x,y,n}(h) = \frac{\sum_{t=1}^{n-h}(y_t - \overline{y})(x_{t+h} - \overline{x})}{\sqrt{\sum_{t=1}^{n}(y_t - \overline{y})^2}\sqrt{\sum_{t=1}^{n}(x_t - \overline{x})^2}}.$$

R: ccf() : la valeur en h de ccf(x,y) correspond à l'estimation de la corrélation entre x_{t+h} et y_t

1.3 Stationnarité empirique

On dit qu'une série chronologique est stationnaire lorsque ses caractéristiques statistiques au second ordre (autocovariance et moyenne) ne dépendent pas (ou pas trop) de la fenêtre de temps dans laquelle on les observe. Ce genre de série est en général irrégulière (au sens "désordonnée") mais *statistiquement régulière*. C'est exactement ce qu'on attend d'un *bruit stationnaire*.

Pour définir cela de manière un peu plus précise, nous allons développer deux idées.

La première est la suivante : on se dit qu'il doit exister trois échelles de temps dans l'intervalle de temps $[0, n-1]$ d'observation de la série.

- le court terme qui correspond à la longueur de la mémoire de la série : deux valeurs dont les dates diffèrent d'un court terme sont corrélées,
- le moyen terme à l'échelle duquel la mémoire s'efface (environ un dixième de la longueur de la série en pratique)
- le long terme, qui est l'ordre de la longueur de la série, et qui doit être grand devant le moyen terme.

On peut alors déplacer mentalement une fenêtre de largeur le moyen terme sur toute la longueur de la série. A chaque position de cette fenêtre, on observe l'échantillon qui se trouve à l'intérieur. Si tous ces échantillons semblent similaires au second ordre, au sens où leurs moyennes et covariances empiriques à court terme semblent à peu près les mêmes, alors on en conclura que la série semble stationnaire. On peut implémenter cela en calculant l'autocovariance et la moyenne empiriques sur toutes les sous-fenêtres de largeur le moyen terme contenues dans $[0, n-1]$. On peut parler d'*autocovariance et moyenne empiriques à court terme*.

Une autre façon de mettre en évidence la non-stationnarité, plus proche de la définition théorique (cf. la Définition 4.4), est d'observer la *matrice de covariance empirique* définie par :

$$\forall h_1, h_2 \in \{1, \ldots, \lfloor n/2 \rfloor\}, \quad \widehat{R}(h_1, h_2) = \frac{1}{n - \max\{h_1, h_2\} + 1} \sum_{t=0}^{n-\max\{h_1,h_2\}} x_{t+h_1} x_{t+h_2} \ .$$

Remarquons qu'elle n'est (volontairement) pas centrée. Si le signal est stationnaire, $\widehat{R}(h_1, h_2)$ doit être proche d'une fonction de $h_2 - h_1$.

Remarque 1.14 *Pour les séries stationnaires, $\rho_{x,n}(h)$ devient généralement proche de 0 rapidement quand h augmente. Lorsque l'on étudie une série brute comportant une ou plusieurs tendances ou saisonnalités, le comportement de la fonction d'autocorrélation empirique est très différent. Par exemple, on peut montrer le résultat suivant.*

Proposition 1.15 *Si on considère une tendance polynomiale pure :*

$$x_t = a_0 + a_1 t + \ldots + a_p t^p, \quad t = 1, 2, \ldots, n$$

alors, pour h fixé, $\rho_{x,n}(h) \underset{n\infty}{\longrightarrow} 1$.
Si on considère une série sinusoïdale pure :

$$x_t = a \cos(2t\pi/T), \quad t = 1, 2, \ldots, n$$

alors, pour h fixé, $\rho_{x,n}(h) \underset{n\infty}{\longrightarrow} \cos(\frac{2h\pi}{T})$.

Ce résultat indique que si la série est assez longue, la période se voit aussi bien sur la fonction d'autocorrélation empirique que sur la série elle-même. En réalité, on n'a jamais affaire à une tendance et/ou une saisonnalité pures, mais la proposition précédente donne une bonne indication de ce qui se passe pour une série présentant une tendance ou une saisonnalité additive. En résumé et pour parler sommairement, pour une longue série présentant une tendance polynomiale, la fonction d'autocorrélation empirique est positive et assez proche de 1 lorsque p évolue

tout en restant petit devant n. Pour une série présentant une périodicité, celle-ci se voit bien sur la fonction d'autocorrélation empirique.

Dans ce cours, lorsque l'on étudie une série chronologique, la première démarche consiste à stationnariser la série (cf. chapitres 2 et 3). On en conclut que si la fonction d'autocorrélation empirique du résidu stationnarisé présente une des propriétés de la Proposition 1.15, c'est sans doute parce que l'on a oublié d'enlever une tendance ou que l'on a mal désaisonnalisé.

Chapitre 2

Tendance et saisonnalité

2.1 Notion de tendance, lissage d'une série chronologique

Dans une série chronologique, une tendance désigne une évolution à long terme, relativement lente. Mathématiquement, cela se formalise (un peu) par une fonction *régulière, variant lentement*, ou plus précisément *dont les variations significatives sont à long terme.*

Donnons quelques exemples de fonctions du temps pouvant présenter ce genre de comportement.

- Les polynômes de bas degré.
- Plus généralement, les fonctions puissances de la forme t^α, $\alpha \in \mathbb{R}$ pour $t > 0$, et leurs combinaisons linéaires.
- Les fonctions exponentielles, de la forme $t \mapsto \alpha \exp \beta t$, avec α et β dans $\mathbb{R}$.
- Les fonctions logarithmes, de la forme $t \mapsto \alpha \ln(t) + \beta$, avec α et β dans $\mathbb{R}$.

Dans les applications, il est rarement agréable d'utiliser ce genre de fonctions explicitement. En effet, il est rare que des formes si nettes soient justifiées par un argument *naturel* (physique, biologique, etc.). D'un autre côté, ces fonctions sont loin d'être aussi malléables qu'on aimerait, et sont donc des outils limités pour estimer une tendance. Par contre, elles fournissent de bons vocables pour décrire rapidement l'allure d'une tendance : tout le monde comprend aisément ce que signifie "une tendance linéaire, croissante". On les utilisera essentiellement de cette façon.

Pour estimer plus précisément une tendance en fonction du temps, on utilisera de préférence une méthode de *régression linéaire locale pondérée* comme *loess* (locally weighted scatterplot smoothing). L'idée de loess (linéaire) est simple : en un instant donné t, on calcule une droite de régression en considérant une fenêtre centrée en t, contenant une proportion α de tous les points. La régression est une régression pondérée : les points les plus éloignés de t dans la fenêtre ont un poids (beaucoup) plus faible que les points les plus proches de t. Puis on estime la tendance en t par l'ordonnée de cette droite de régression en t. Ceci est fait en chaque instant t.

R: Pour une série temporelle x, utiliser loess(x ~ time(x)). Le paramètre alpha peut être changé (défaut à 0.75). Le résultat est un objet de classe 'loess'. La tendance est dans loess$fitted.

2.2 Phénomènes saisonniers

De très nombreux phénomènes temporels présentent un, ou plusieurs, caractères périodiques. Les cycles peuvent être annuels (influences des saisons), mensuels (penser au salaire), lunaires, journaliers etc. Le cycle le plus fréquemment rencontré est sans doute le cycle annuel, nous allons donc nous placer dans ce cadre, le transfert à une autre période étant immédiat.

Lorsqu'on observe pendant plusieurs années une grandeur influencée par un cycle annuel, on dit qu'il y a un phénomène saisonnier. On peut espérer estimer un *motif périodique* qui représenterait sa dépendance au rythme annuel. Une façon de faire cela est de calculer pour

chaque série mensuelle (la série des mois de janvier, celle des mois de février etc.) la moyenne et l'écart-type empiriques. La *série désaisonnalisée* est alors la série obtenue en retranchant à la série initiale les moyennes mensuelles, puis en divisant cette différence par les écart-types mensuelles.

Plus formellement, pour une série mensuelle avec un rythme annuel, on obtient 12 moyennes mensuelles, $m_{Jan}, \ldots, m_{Dec}$ et 12 écart-types mensuels $\sigma_{Jan}, \ldots, \sigma_{Dec}$. Si la série initiale est x, on obtient la série désaisonnalisé en un mois t qui est, par exemple, un mois de juin, par :

$$u_t = \frac{x_t - m_{Juin}}{\sigma_{Juin}} \ .$$

Remarque 2.1 *Il est nécessaire d'avoir préalablement extrait la tendance.*

Lorsque cette opération réussit, au sens où la série désaisonnalisée ne présente plus de caractère saisonnier, le motif répété des moyennes mensuelles est appelé *saisonnalité*. C'est une série périodique, de même longueur que la longueur de la série initiale.

R: Un outil bien utile : monthplot(), qui trace les séries mensuelles.

Chapitre 3

Stationnarisation

3.1 Modèle additif

Lorsqu'on a identifié (éventuellement) une tendance puis (éventuellement) une saisonnalité, et qu'on a éliminé ces composantes de notre série, on obtient quelque chose, un *résidu* qui n'a plus ni tendance ni saisonnalité, et qui est de ce fait relativement *informe*. En réalité, il y a souvent encore beaucoup à dire, au moins d'*un point de vue statistique*. L'idéal (car cela mène à une représentation plus simple) est lorsque ce résidu est *stationnaire* (cf. section 1.3). On pourra alors essayer d'ajuster un modèle de bruit stationnaire comme les $ARMA$ (cf. partie II), et on aura un joli *modèle additif* :

$$x_t = m_t + s_t + \varepsilon_t \ ,$$

où :
- m est la tendance,
- s la saisonnalité,
- ε est un bruit stationnaire.

Ce type de modèle est assez souhaitable : il est simple, linéaire, et on se représente bien visuellement ce qui se passe lorsqu'on ajoute deux séries. Néanmoins, il est rare qu'une série nous tombe toute cuite dans les mains, prête à être décomposée en modèle additif. Souvent, cette décomposition peut coincer pour les raisons suivantes :
- les variations saisonnières ont une amplitude variable,
- le bruit restant a une amplitude variable

On peut alors essayer de transformer la série de manière simple. C'est l'objet des transformations de Box-Cox, cf. section 3.2.

Par ailleurs, une autre limitation du modèle additif est l'accent porté sur la modélisation de m et s, la tendance et la saisonnalité. On peut être intéressé à transformer une série afin de la rendre stationnaire sans avoir à estimer tendance et saisonnalité. Une façon de faire cela est de stationnariser par différentiation, c'est l'objet de la section 3.3.

3.2 Transformations de Box-Cox

Il arrive fréquemment que des séries chronologiques aient une variance de bruit et une amplitude de variations saisonnières qui sont des fonctions croissantes de m. Pour y rémédier, une solution est d'utiliser une *transformation de Box-Cox* pour les séries strictement positives. Cela consiste à transformer une série x_t en $x_t^{(a)}$ où :

$$x_t^{(a)} = \begin{cases} \frac{(x_t)^a - 1}{a} & \text{si } a \neq 0 \\ \log(x_t) & \text{si } a = 0 \end{cases}$$

où a est une constante à choisir. En pratique, un bon choix de a permet souvent de stabiliser les fluctuations des résidus et de la saisonnalité, voire de les rendre à peu près gaussiens. On peut également appliquer cette transformation à $x_t + c$ avec c une constante à choisir. Cela permet de rendre la série positive, mais également, parfois, d'améliorer encore la stabilité de la variance ou le caractère à peu près gaussien des résidus.

3.3 Différentiation

Une approche complètement différente est fournie par les remarques suivantes :

- Une tendance est une fonction régulière, donc si on la dérive suffisamment, elle doit devenir moins régulière, jusqu'à ressembler à un bruit,
- Une saisonnalité s est une fonction périodique de période p, donc si on calcule $s_t - s_{t-p}$, on la supprime.

Dans le premier cas, on parle de différentiation à l'échelle de l'échantillon. Comme nos séries vivent en des instants discrets, cela correspond à tranformer x_t en $x_t - x_{t-1}$.

Dans le second cas, on parle de différentiation à l'échelle de la période. Cela correspond à tranformer x_t en $x_t - x_{t-p}$.

Cela revient à considérer un tout autre type de modèle que le modèle additif : on cherche désormais à différentier x à l'échelle de la période et de l'échantillon jusqu'à ce que la série obtenue soit stationnaire. On verra plus tard que lorsqu'on obtient comme résidu un bruit $ARMA$, alors la série initiale est ce qu'on appelle un $SARIMA$, et que ce genre de modèle est très bien adapté à la prédiction.

R: la fonction diff() permet de différentier. L'option "lag" permet de préciser la valeur de p dans $x_t - x_{t-p}$, et l'option "differences" le nombre de fois que l'on veut différentier.

3.4 Première description d'une série chronologique

Pour clore ce chapitre, résumons ce qu'on entend par "décrire une série chronologique". Cela signifie tracer la série, l'observer, puis commenter les aspects suivants :

- tendance (linéaire ? exponentielle ? etc.)
- phénomène saisonnier (amplitude variable ? motif variable ?)
- bruit (amplitude variable ? Corrélation variable ?)

Enfin, cela signifie transformer de la façon la plus simple possible la série de manière à obtenir un résidu stationnaire, et être capable de résumer les transformations effectuées.

Deuxième partie

Les ARMA : des bruits stationnaires autocorrélés

Nous avons vu dans la partie I que les résidus dans la décomposition d'une série chronologique sont rarement décorrélés, même s'ils sont stationnaires. L'objectif de cette partie est de présenter une classe de modèles, sans doute les plus simples, de bruits stationnaires autocorrélés : les $ARMA$. L'introduction de tels bruit représentent non seulement un progrès dans le réalisme de la modélisation, mais permettra aussi d'améliorer la qualité de la prédiction.

Pour illustrer cette possibilité d'améliorer la prédiction, nous allons partir d'un exemple théorique simple. Soit $\phi \in]-1,1[$ un paramétre supposé connu. On considère une suite de variables aléatoires définies récursivement comme suit :

$$\begin{cases} Y_0 = 0 \\ Y_n = \phi Y_{n-1} + Z_n \quad \forall n \geqslant 1 \end{cases}$$

où $(Z_n)_{n\geqslant 1}$ est une suite i.i.d de variables aléatoires gaussiennes centrées de variance $\sigma^2 > 0$. On suppose maintenant qu'on observe $Y_0, \ldots, Y_{n-1}$, et qu'on veut prédire la valeur de Y_n. Au vu de la récurrence, et Z_n étant indépendant de $Y_0, \ldots, Y_{n-1}$, il semble logique de prédire Y_n par ϕY_{n-1}. On trouve alors que l'*erreur quadratique moyenne de prédiction* est :

$$\mathbb{E}[(Y_n - \phi Y_{n-1})^2] = \mathbb{E}[Z_n^2] = \sigma^2 \ .$$

Et si on avait oublié la dépendance de Y_n en Y_{n-1} ? On aurait naturellement prédit Y_n par son espérance, c'est à dire par 0. On aurait alors une erreur quadratique moyenne de prédiction de :

$$\mathbb{E}[Y_n^2] = \mathbb{E}[(\phi Y_{n-1} + Z_n)^2] = \sigma^2 + \phi^2 \mathbb{E}[Y_{n-1}^2] = \ldots = \sigma^2 \sum_{i=0}^{n-1} \phi^{2i} = \sigma^2 \frac{1-\phi^{2n}}{1-\phi^2} \ .$$

On a ainsi une erreur quadratique moyenne de prédiction qui est strictement supérieur à σ^2, l'erreur quadratique obtenue en tenant compte de la dépendance de Y_n en Y_{n-1}. Remarquons qu'en général, on ne connait pas le paramètre ϕ, et on ne connait même pas la forme exacte de la dépendance de Y_n en fonction de $Y_0, \ldots, Y_{n-1}$. Il faut alors *estimer* cette dépendance, c'est le rôle de la statistique.

Le chapitre 4 sera dédié à la définition théorique des *processus stationnaires*. Les *processus* $ARMA$ et leurs propriétés élémentaires seront introduits dans le chapitre 5. Le chapitre 6 visera à formuler rigoureusement la notion de *prédiction linéaire* telle qu'on l'a mise en oeuvre ci-dessus. Enfin, le chapitre 7 traitera de la *statistique des processus* $ARMA$.

Chapitre 4

Processus stationnaires

Un processus aléatoire $(X_t)_{t\in T}$, indexé par un ensemble T est simplement une collection de variables aléatoires définies sur le même espace de probabilité. Dans ce cours, T désigne une succession d'instants. Plus précisément :

Dans tout ce cours, T désignera un ensemble d'instants et sera soit $\mathbb{Z}$ soit un sous-ensemble fini de $\mathbb{Z}$, typiquement $\{0, \ldots, n-1\}$.

4.1 Bruit blanc

L'exemple le plus simple de processus aléatoire est celui du *bruit blanc fort discret*. C'est tout simplement une suite de n variables aléatoires centrées indépendantes et identiquement distribuées $W_0, \ldots, W_{n-1}$. Lorsque ces variables sont gaussiennes, on parle de bruit blanc gaussien discret. Sur la Figure 4.1, on voit deux exemples de bruits blancs, l'un gaussien et l'autre de loi de Pareto (standardisée) de paramètres $(1, 3)$. Noter que cette loi de Pareto a une queue de distribution assez lourde, et que cela a une incidence sur les valeurs extrêmes prises par le bruit blanc.

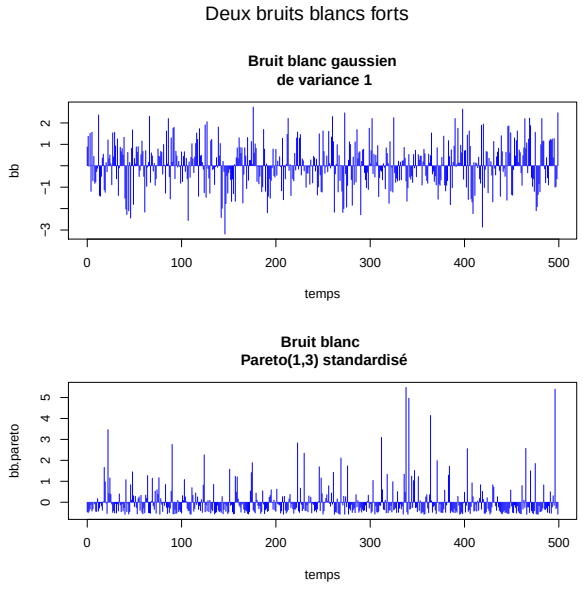

FIGURE 4.1 – Deux réalisations de bruits blancs

4.2 Série chronologique aléatoire stationnaire

Le point de vue probabiliste adopté dans ce cours est qu'*une série chronologique observée est la réalisation d'un processus aléatoire*, qu'on appellera *série chronologique aléatoire.*

Définition 4.1 (Série chronologique aléatoire) *Soit $T \subset \mathbb{Z}$ un ensemble d'instants. Une* série chronologique aléatoire *est une collection $(X_t)_{t\in T}$ de variables aléatoires réelles définies sur un même espace de probabilité.*

Définition 4.2 *Une série chronologique aléatoire $(X_t)_{t\in T}$ est dite* du second ordre *si :*

$$\forall t \in T,\ \mathbb{E}(X_t^2) < \infty\ .$$

Une série chronologique aléatoire $(X_t)_{t\in T}$ est dite centrée *si :*

$$\forall t \in T,\ \mathbb{E}(X_t) = 0\ .$$

Définition 4.3 (Fonction de covariance) *On définit la* fonction de covariance Γ_X *d'une série chronologique aléatoire du second ordre $(X_t)_{t\in T}$ par :*

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

Définition 4.4 (Stationnarité) *On dit qu'une série chronologique aléatoire $(X_t)_{t\in\mathbb{Z}}$ indexée par $\mathbb{Z}$ est* stationnaire en moyenne d'ordre deux *(ou stationnaire au sens large) si elle est du second ordre, et si :*

$$\begin{cases} t \mapsto \quad \mathbb{E}(X_t) \text{ est constante}\ , \\ \forall h, s, t \in \mathbb{Z},\ \Gamma_X(t,s) = \Gamma_X(t+h, s+h)\ . \end{cases}$$

Pour un processus stationnaire, on peut réécrire la fonction de covariance comme une fonction d'une seule variable :

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

On définit alors la fonction d'autocorrélation ρ_X *par :*

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}\ ,$$

avec

$$\gamma_X(0) = \mathsf{Var}(X_0)\ .$$

Par définition, on a donc $\rho_X(0) = 1$.

Définition 4.5 (Processus SLC) *On dit qu'une série chronologique aléatoire est un* processus SLC *si elle est stationnaire au sens large et centrée.*

Exemple 4.6 *Le* bruit blanc fort discret *de la section 4.1 peut s'étendre en un processus SLC si l'on considère une suite $(W_t)_{t\in\mathbb{Z}}$ de variables aléatoires i.i.d de carré intégrable.*

Par définition, un bruit blanc faible *est un processus SLC X tel que pour tout h non nul, $\gamma_X(h)$ soit nul : à deux instants distincts, les valeurs du processus sont décorrélées.* Dans la suite de ce cours, le terme "bruit blanc" fera référence à un bruit blanc faible, sauf mention explicite du contraire.

Remarque 4.7 *On peut définir aussi la* stationnarité au sens strict *: un processus X est strictement stationnaire si les vecteurs $(X_t, \ldots, X_{t+n-1})$ et $(X_0, \ldots, X_{n-1})$ ont même loi quels que soient t dans $\mathbb{Z}$ et n dans $\mathbb{N}^*$. La stationnarité au sens strict implique la stationnarité au sens large si le processus est du second ordre, mais la réciproque n'est pas vraie, sauf lorsque X est un processus gaussien, c'est à dire lorsque $(X_t, \ldots, X_{t+n-1})$ est un vecteur gaussien pour tous t dans $\mathbb{Z}$ et n dans $\mathbb{N}^*$.*

Proposition 4.8 *Soit X une série chronologique aléatoire non constante et stationnaire.*
- $\forall t \in \mathbb{Z}, \quad \mathsf{Var}(X_t) = \gamma_X(0)$,
- $\forall t \in \mathbb{Z}, \quad |\gamma_X(t)| \leqslant |\gamma_X(0)|$ *et* $|\rho_X(t)| \leqslant 1$,
- $\forall t \in \mathbb{Z}, \quad \gamma_X(-t) = \gamma_X(t)$ *et* $\rho_X(-t) = \rho_X(t)$.

Exemple 4.9 *Voir l'Exercice 2.1 du TD.*

4.3 Retour sur la notion de stationnarité empirique

Dans la pratique, on observe une *réalisation* d'une série chronologique aléatoire. On a déjà, dans la section 1.3, donné une définition (vague) de stationnarité *empirique*. Y a-t-il un lien entre la définition de stationnarité théorique définie dans ce chapitre et la définition (vague) de stationnarité empirique ? On aimerait pouvoir qualifier une série chronologique de stationnaire (d'un point de vue empirique) s'il est vraisemblable qu'elle soit la réalisation d'un processus stationnaire. Ceci suppose que l'on puisse dire quelque chose sur la loi du processus à partir d'une seule réalisation, et fait donc l'hypothèse d'une certaine *ergodicité* du processus, et même *d'une mémoire courte par rapport à la longueur de la série* puisque la réalisation est de longueur finie.

Pour faire le lien entre les deux notions, on dira donc dans ce cours qu'*une série chronologique semble stationnaire lorsqu'il est vraisemblable qu'elle soit la réalisation d'un processus stationnaire dont la mémoire est courte devant la longueur de la série.*

Chapitre 5

Filtrage linéaire et processus ARMA

Pour manipuler les séries chronologiques aléatoires, on utilisera essentiellement trois opérations : addition de deux séries chronologiques, multiplication d'une série chronologique par un nombre réel et image d'une série chronologique par l'*opérateur retard*. Cet opérateur, noté B, consiste simplement à transformer une série par la même série, mais retardée dans le temps d'une période d'échantillonnage. Autrement dit, si X est une série chronologique indexée par $\mathbb{Z}$, BX est encore une série chronologique indexée par $\mathbb{Z}$, définie par :

$$(BX)_t = X_{t-1} \ .$$

On note aussi B^{-1} l'opérateur inverse de B :

$$(B^{-1}X)_t = X_{t+1} \ .$$

On a bien sûr $B^{-1}(BX) = B(B^{-1}X) = X$. Pour $k \in \mathbb{N}$, on note B^k l'opérateur qui consiste à appliquer B k fois de suite. Notamment, B^0 ne fait rien : $B^0X = X$. Pour k dans $\mathbb{N}$, B^{-k} désigne l'opérateur qui consiste à appliquer B^{-1} k fois de suite. Remarquons qu'alors :

$$(B^kX)_t = X_{t-k} \ .$$

En combinant les opérations décrites ci-dessus (addition de deux séries chronologiques, multiplication d'une série chronologique par un nombre réel et image d'une série chronologique par l'*opérateur retard*), on obtient le *filtrage linéaire*.

5.1 Filtrage linéaire

Une façon simple d'obtenir de nouveaux processus SLC à partir de processus SLC est de les *filtrer linéairement*.

Théorème 5.1 (Filtrage linéaire) *Soit X un processus SLC et (ψ_k) une suite sommable de nombres réels :*

$$\sum_{k\in\mathbb{Z}} |\psi_k| < \infty \ .$$

On note $\psi(B)$ l'opérateur de filtrage défini par :

$$\psi(B) = \sum_{k\in\mathbb{Z}} \psi_k B^k \ .$$

Alors, le processus Y défini par :

$$\forall t \in \mathbb{Z}, \quad Y_t = \sum_{k\in\mathbb{Z}} \psi_k X_{t-k} \ ,$$

et noté $\psi(B)X$, est un processus SLC dont la covariance vaut :

$$\forall h \in \mathbb{Z}, \quad \gamma_Y(h) = \sum_{j,k\in\mathbb{Z}} \psi_j \psi_k \gamma_X(h-j+k) \ .$$

Démonstration : On passe sous silence la démonstration de la convergence dans la définition de Y ... Comme la série $\sum_{i\in\mathbb{Z}} |\psi_i|$ converge, on peut inverser l'espérance et la somme dans ce qui suit :

$$\mathbb{E}(Y_t) = \sum_{k\in\mathbb{Z}} \psi_k \mathbb{E}(X_{t-k}) = 0,$$

donc Y est centré, et :

$$\begin{aligned} \mathbb{E}(Y_{t+h}Y_t) &= \sum_{j,k\in\mathbb{Z}} \psi_j \psi_k \mathbb{E}(X_{t+h-j}X_{t-k}) \ , \\ &= \sum_{j,k\in\mathbb{Z}} \psi_j \psi_k \gamma_X(h-j+k) \ , \end{aligned}$$

qui ne dépend pas de t. Donc Y est stationnaire. □

Exemple 5.2 *Le filtrage linéaire d'un bruit blanc donne un bruit "coloré" en général, c'est à dire non blanc. Donnons un exemple. Soient $\theta_1, \ldots, \theta_q$ des nombres réels, Z un bruit blanc de variance σ^2, et X un processus défini par :*

$$X_t = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q} \ .$$

X est ce qu'on appelle un processus à moyenne mobile d'ordre q, en abrégé un $MA(q)$ (MA signifie moving average*, i.e moyenne mobile). La covariance de X vaut, d'après le Théorème 5.1 :*

$$\forall h \geqslant 0, \quad \gamma_X(h) = \sigma^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h} \ ,$$

en posant $\theta_0 = 1$.

R: filter()

5.2 Les processus ARMA

Les modèles $ARMA$ sont des modèles stationnaires de dimension finie dont la covariance approche une large famille de covariances.

Définition 5.3 (Processus ARMA) *On dit qu'une série chronologique aléatoire X est un processus $ARMA(p,q)$ si X est un processus SLC et s'il existe un bruit blanc (faible) Z_t et des réels $\phi, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ tels que :*

$$\forall t \in \mathbb{Z}, \quad X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q} \ , \tag{5.1}$$

relation que l'on notera : $\phi(B)X_t = \theta(B)Z_t$, où ϕ et θ sont les polynômes de degrés respectifs p et q :

$$\begin{aligned} \phi(z) &= 1 - \phi_1 z - \ldots - \phi_p z^p \ , \\ \theta(z) &= 1 + \theta_1 z + \ldots + \theta_q z^q \ , \end{aligned}$$

et B est l'opérateur "retard" défini par :

$$B^k X_t = X_{t-k} \ .$$

Enfin, on dit que X est un $ARMA(p,q)$ de moyenne m si le processus $(X_t - m)_{t\in\mathbb{Z}}$ est un $ARMA(p,q)$ de moyenne nulle.

Exemple 5.4 *On a déjà vu les processus $MA(q)$ dans l'Exemple 5.2. Ce sont donc ceux obtenus en prenant $\phi(z) = 1$ dans la Définition 5.3. On définit aussi les processus autorégressifs d'ordre p, en abrégé $AR(p)$, comme les $ARMA(p,q)$ pour lesquels $\theta(z) = 1$. Autrement dit, les processus stationnaires, centrés tels qu'il existe un bruit blanc faible Z et des réels $\phi, \dots, \phi_p$ tels que :*

$$\forall t \in \mathbb{Z}, \quad X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t \ .$$

5.3 Rappels sur les nombres complexes

Les nombres complexes sont une extension des nombres réels (ils contiennent donc les nombres réels). Leur ensemble est noté $\mathbb{C}$. La construction des complexes se fait principalement par l'adjonction d'un nombre "imaginaire" i tel que $i^2 = -1$. Tout nombre complexe z s'écrit alors de manière unique sous la forme $a + ib$ où a et b sont deux *réels*. Le nombre a est la *partie réelle* de z, notée $Re(z)$ et b est la *partie imaginaire* de z, notée $Im(z)$. Lorsque $Im(z) = 0$, z est en fait un nombre réel. On définit aussi le *module* de $z = a + ib$, noté $|z|$, par :

$$|z| = \sqrt{a^2 + b^2} \ .$$

Cela correspond à la distance entre z et l'origine lorsqu'on représente le complexe $a + ib$ par le point de coordonnées (a, b) dans un plan, muni d'un repère orthonormé (on parle alors de représentation dans le plan complexe). Lorsque $z \in \mathbb{R}$, module et valeur absolue coïncident.

Le point crucial que nous allons utiliser pour le moment est que $\mathbb{C}$ est *algébriquement clos*, c'est à dire que tout polynôme de degré d à coefficients dans $\mathbb{C}$ (et donc notamment tout polynôme de degré d à coefficients dans $\mathbb{R}$) admet d racines dans $\mathbb{C}$. Cela signifie que tout polynôme de degré d à coefficients dans $\mathbb{C}$ se factorise en (i.e est égal à) un produit de d polynômes de degré 1 à coefficients dans $\mathbb{C}$.

Rappelons un résultat concernant la somme des séries géométriques, valable aussi pour les nombres complexes : pour tout $z \in \mathbb{C}$, si $|z| < 1$,

$$\sum_{k=0}^{\infty} z^k = \frac{1}{1-z} \tag{5.2}$$

Enfin, remarquons qu'on peut étendre la notion de processus SLC à des processus qui sont à valeurs complexes au lieu d'être à valeurs réelles. Pour cela, on utilise la notion suivante de covariance :

$$\Gamma_X(t,s) = \mathsf{Cov}(X_t, \overline{X_s}) = \mathbb{E}[(X_t - \mathbb{E}(X_t))\overline{(X_t - \mathbb{E}(X_t))}] \ .$$

On peut ainsi définir le filtrage linéaire avec des coefficients complexes.

R: Re(), Im(), abs() ou Mod().
le nombre i s'obtient avec 1i.
Pour factoriser un polynôme : polyroot()

5.4 Manipuler les séries d'opérateurs retard

Voici quelques règles qui permettront de manipuler sans trop de douleur les séries d'opérateurs retard apparaissant dans les filtres linéaires que l'on rencontrera.

Soient $\alpha = (\alpha_k)_{k\in\mathbb{Z}}$ et $\beta = (\beta_k)_{k\in\mathbb{Z}}$ deux suites sommables de nombres complexes. Alors, pour toute série chronologique X,

$$\alpha(B)(\beta(B)X) = (\alpha(B)\beta(B))X \ ,$$

avec :

$$\begin{aligned}
\alpha(B)\beta(B) &= \sum_{k\in\mathbb{Z}} \alpha_k B^k \sum_{l\in\mathbb{Z}} \beta_l B^l \ , \\
&= \sum_{k,l\in\mathbb{Z}} \alpha_k \beta_l B^{k+l} \ , \\
&= \sum_{n\in\mathbb{Z}} \sum_{k,l\in\mathbb{Z}} \mathbb{1}_{k+l=n} \alpha_k \beta_l B^{k+l} \ , \\
&= \sum_{n\in\mathbb{Z}} B^n \sum_{k,l\in\mathbb{Z}} \mathbb{1}_{k+l=n} \alpha_k \beta_l \ , \\
&= \sum_{n\in\mathbb{Z}} B^n \sum_{k\in\mathbb{Z}} \alpha_k \beta_{n-k} \ .
\end{aligned}$$

Enfin, pour un $ARMA$ X défini par :

$$\phi(B)X_t = \theta(B)Z_t \ ,$$

avec Z un bruit blanc, on aimerait comprendre comment s'obtient la représentation de X en filtrage linéaire de Z évoqué dans le Théorème 5.5. Pour comprendre ce qui peut se passer, prenons un $AR(1)$:

$$X_t - \phi X_{t-1} = Z_t \ ,$$

avec $|\phi| < 1$. On a alors :

$$X_t = \sum_{k\geqslant 0} \phi^k Z_{t-k} \ .$$

Formellement, cela revient à écrire :

$$X_t = \frac{1}{1-\phi B} Z_t = (\sum_{k\geqslant 0} \phi^k B^k) Z_t \ .$$

Mais si on a :

$$X_t - \phi X_{t-1} = Z_t \ ,$$

avec $|\phi| > 1$, alors la série $\sum_{k\geqslant 0} \phi^k$ diverge. On a néanmoins :

$$X_t = \frac{1}{\phi} X_{t+1} - \frac{1}{\phi} Z_{t+1} \ .$$

D'où :

$$X_t = -\sum_{k\leqslant -1} \phi^k Z_{t-k} \ .$$

Formellement, cela revient à écrire :

$$X_t = \frac{1}{1-\phi B} Z_t = \frac{-\phi B}{1-\frac{1}{\phi B}} Z_t = -\phi B \left[\sum_{k\geqslant 0} \left(\frac{1}{\phi B} \right)^k \right] Z_t \ .$$

On voit donc qu'il faut tenir compte de la valeur de ϕ, ou plus précisément des racines du polynôme $\phi(B)$.

Par exemple, pour un $AR(2)$ défini par :

$$\phi(B)X_t = \theta(B)Z_t \ ,$$

avec $\phi(B) = (1 - B/z_1)(1 - B/z_2)$, où $|z_1| > 1$ et $|z_2| > 1$ et $z_1 \neq z_2$. On a :

$$\begin{aligned}
\frac{1}{\phi(B)} &= \frac{1}{1 - B/z_1}\frac{1}{1 - B/z_2} \ , \\
&= \left(\sum_{k\geqslant 0} \frac{B^k}{z_1^k}\right)\left(\sum_{l\geqslant 0} \frac{B^l}{z_1^l}\right) \ , \\
&= \sum_{n\in\mathbb{Z}} B^n \sum_{k=0}^{n} \frac{1}{z_1^k z_2^{n-k}} \ , \\
&= \sum_{n\in\mathbb{Z}} B^n \sum_{k=0}^{n} \frac{1}{z_1^k z_2^{n-k}} \ , \\
&= \sum_{n\in\mathbb{Z}} B^n \frac{1}{z_2^n} \sum_{k=0}^{n} \left(\frac{z_2}{z_1}\right)^k \ , \\
&= \sum_{n\in\mathbb{Z}} B^n \frac{1}{z_2^n} \frac{\left(\frac{z_2}{z_1}\right)^{n+1} - 1}{\frac{z_2}{z_1} - 1} \ , \\
&= \sum_{n\in\mathbb{Z}} B^n \frac{z_1^{n+1} - z_2^{n+1}}{z_1 - z_2} \ ,
\end{aligned}$$

5.5 Existence, causalité, inversibilité

Le théorème suivant montre que l'existence des processus $ARMA$ est reliée aux racines (dans $\mathbb{C}$, voir le paragraphe 5.3) du polynôme $z \mapsto \phi(z)$.

Théorème 5.5 (Existence d'un ARMA) *Soient Z un bruit blanc, et :*

$$\begin{aligned}
\phi(B) &= 1 - \phi_1 B - ... - \phi_p B^p \ , \\
\theta(B) &= 1 + \theta_1 B + ... + \theta_q B^q \ ,
\end{aligned}$$

tels que ϕ n'ait pas de racine de module 1. Alors, il existe un unique processus SLC X tel que :

$$\forall t \in \mathbb{Z}, \quad \phi(B)X_t = \theta(B)Z_t \ .$$

De plus, X est obtenu à partir de Z par un filtrage linéaire.

Démonstration : Admise. □

Deux notions liées à la relation (5.1) sont importantes en pratique. La première est celle de *causalité*, et signifie que X_t peut s'obtenir uniquement à partir des valeurs de $(Z_s)_{s\leqslant t}$. D'un point de vue logique, c'est quelque chose de souhaitable si le bruit Z_s représente un aléa ayant lieu au temps s. L'autre notion est celle d'*inversibilité*, et signifie que l'on peut reconstruire Z_t uniquement à partir des valeurs de $(X_s)_{s\leqslant t}$. L'inversibilité est nécessaire pour avoir une chance de pouvoir reconstruire (de manière approchée) le bruit blanc Z après avoir observé l'$ARMA$ jusqu'à un certain instant.

Dans la suite, X est un $ARMA(p, q)$ et Z un bruit blanc vérifiant (5.1), c'est à dire $\phi(B)X_t = \theta(B)Z_t$.

Définition 5.6 (Causalité) *X est dit causal (par rapport à Z) s'il existe une suite sommable de réels $(\psi_k)_{k\in\mathbb{N}}$ telle que :*

$$\forall t \in \mathbb{Z}, \quad X_t = \sum_{k=0}^{+\infty} \psi_k Z_{t-k} \ .$$

La proposition suivante montre qu'un $ARMA$ est causal si et seulement si les racines du polynôme $z \mapsto \phi(z)$ sont en dehors du disque unité.

Proposition 5.7 (Critère de Causalité) *X est causal (par rapport à Z) si et seulement si :*

$$\forall z \in \mathbb{C}, \quad |z| \leqslant 1 \Rightarrow \phi(z) \neq 0 \ .$$

Dans ce cas,

$$X_t = \sum_{k=0}^{+\infty} \psi_k Z_{t-k} \ ,$$

avec, pour tout z complexe tel que $|z| \leqslant 1$:

$$\psi(z) = \sum_{k=0}^{+\infty} \psi_k z^k = \frac{\theta(z)}{\phi(z)} \ .$$

Démonstration : On esquisse juste un sens, de manière non rigoureuse (cf. Théorème 3.1.1 page 85 de [3] pour une démonstration complète). Supposons que

$$\forall z \in \mathbb{C}, \quad |z| \leqslant 1 \Rightarrow \phi(z) \neq 0 \ .$$

Cela signifie qu'il existe $z_1, \ldots, z_p$ des nombres complexes de module strictement supérieur à 1 tels que :

$$\phi(z) = \prod_{i=1}^{p} (1 - \frac{z}{z_i}) \ .$$

Donc :

$$\frac{\theta(z)}{\phi(z)} = \theta(z) \prod_{i=1}^{p} \frac{1}{1 - \frac{z}{z_i}} \ .$$

Maintenant, d'après (5.2), pour tout z tel que $|z| < |z_i|$,

$$\frac{1}{1 - \frac{z}{z_i}} = \sum_{k=0}^{\infty} \left(\frac{z}{z_i}\right)^k \ .$$

Soit $\rho = \min\{|z_i|, \ i = 1, \ldots, p\}$ et z un nombre complexe tel que $|z| < \rho$. Alors,

$$\frac{\theta(z)}{\phi(z)} = \theta(z) \prod_{i=1}^{p} \left(\sum_{k=0}^{\infty} \left(\frac{z}{z_i}\right)^k \right) \ .$$

En développant ce produit, on peut montrer qu'il existe alors une suite de réels $(\psi_k)_{k\geqslant 0}$ telle que, pour tout z vérifiant $|z| < \rho$:

$$\frac{\theta(z)}{\phi(z)} = \sum_{k=0}^{\infty} \psi_k z^k \ ,$$

et telle que la convergence de la somme soit valable en module. Comme $\rho < 1$, on en déduit que ψ est absolument sommable, et on peut formellement écrire :

$$X_t = \frac{\theta(B)}{\phi(B)} Z_t = \sum_{k=0}^{\infty} \psi_k B^k Z_t = \sum_{k=0}^{\infty} \psi_k Z_{t-k} \ ,$$

qui a bien un sens grâce au Théorème 5.1. □

Remarque 5.8 *Pour un ARMA causal, les coefficients ψ_k peuvent être calculés récursivement en développant l'identité :*

$$(\psi_0 + \psi_1 z + \ldots)(1 - \phi_1 z - \ldots - \phi_p z^p) = 1 + \theta_1 z + \ldots + \theta_q z^q \ ,$$

et en identifiant les coefficiants des monômes de même degré. On obtient :

$$\begin{cases} \psi_0 & = & 1 \\ \psi_j & = & \sum_{k=1}^{\min\{p,j\}} \phi_k \psi_{j-k} + \theta_j \quad \forall j \in \mathbb{N}^* \ , \end{cases}$$

où on a posé $\theta_j = 0$ si $j \geqslant q+1$.

Définition 5.9 (Inversibilité) *X est dit inversible (par rapport à Z) s'il existe une suite sommable de réels $(\lambda_k)_{k\in\mathbb{N}}$ telle que :*

$$\forall t \in \mathbb{Z}, \quad Z_t = \sum_{k=0}^{+\infty} \lambda_k X_{t-k} \ .$$

La proposition suivante montre qu'un $ARMA$ est inversible si et seulement si les racines du polynôme $z \mapsto \theta(z)$ sont en dehors du disque unité.

Proposition 5.10 (Critère d'inversibilité) *X est inversible si et seulement si :*

$$\forall z \in \mathbb{C}, \quad |z| \leqslant 1 \Rightarrow \theta(z) \neq 0 \ .$$

Dans ce cas,

$$Z_t = \sum_{k=0}^{+\infty} \lambda_k X_{t-k} \ ,$$

avec, pour tout z complexe tel que $|z| \leqslant 1$:

$$\lambda(z) = \sum_{k=0}^{+\infty} \lambda_k z^k = \frac{\phi(z)}{\theta(z)} \ .$$

Démonstration : Admise (cf. Théorème 3.1.2 page 86 de [3]). □

5.6 Calcul de la covariance théorique

On peut utiliser la méthode des *équations aux différences* pour calculer la fonction de covariance théorique d'un $ARMA$ causal.

Exemple 5.11 $AR(1)$

Proposition 5.12 (Méthode des équations aux différences) *Soit X un $ARMA(p,q)$ causal, de covariance γ, piloté par un bruit blanc de variance σ^2. On a alors :*

$$\begin{array}{rcll} \gamma(k) - \phi_1\gamma(k-1) - \ldots - \phi_p\gamma(k-p) & = & \sigma^2 \sum_{j=k}^{q} \theta_j \psi_{j-k} & \forall 0 \leqslant k \leqslant q \\ & = & 0 & \forall k \geqslant q+1 \end{array}$$

où les $(\psi_k)_{k\in\mathbb{Z}}$ *sont tels que :*

$$\forall |z| \leqslant 1, \quad \sum_{k=0}^{+\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)} \ ,$$

c'est à dire :

$$\left\{ \begin{array}{lcl} \psi_0 & = & 1 \\ \psi_j & = & \sum_{k=1}^{\min\{p,j\}} \phi_k \psi_{j-k} + \theta_j \quad \forall j \in \mathbb{N}^* \ , \end{array} \right.$$

où on a posé $\theta_j = 0$ *si* $j \geqslant q+1$.

Démonstration : On part de l'équation de l'$ARMA$, valable pour tout $t \in \mathbb{Z}$:

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q} \ .$$

On la multiplie par X_{t-k} :

$$X_{t-k}X_t - \phi_1 X_{t-k}X_{t-1} - \ldots - \phi_p X_{t-k}X_{t-p} = X_{t-k}Z_t + \theta_1 X_{t-k}Z_{t-1} + \ldots + \theta_q X_{t-k}Z_{t-q} \ ,$$

puis on prend l'espérance :

$$\begin{array}{rl} & \mathbb{E}(X_{t-k}X_t) - \phi_1 \mathbb{E}(X_{t-k}X_{t-1}) - \ldots - \phi_p \mathbb{E}(X_{t-k}X_{t-p}) \\ = & \mathbb{E}(X_{t-k}Z_t) + \theta_1 \mathbb{E}(X_{t-k}Z_{t-1}) + \ldots + \theta_q \mathbb{E}(X_{t-k}Z_{t-q}) \ . \end{array}$$

Pour les termes de droite, on remarque, en utilisant la Proposition 5.7 que :

$$\mathbb{E}(X_{t-k}Z_{t-j}) = \sum_{l\geqslant 0} \psi_l \mathbb{E}(Z_{t-k-l}Z_{t-j}) = \left\{ \begin{array}{ll} \psi_{j-k}\sigma^2 & \text{si } k \leqslant j \\ 0 & \text{sinon} \end{array} \right. \ .$$

En remarquant enfin que $\mathbb{E}(X_{t-k}X_{t-j}) = \gamma_X(j-k)$, on obtient exactement le résultat annoncé. □

R: ARMAacf(ar=,ma=,lag.max=) calcule l'autocorrélation théorique d'un ARMA jusqu'au retard lag.max

5.7 Simulation

On veut simuler un morceau de trajectoire $X_0, \ldots, X_{n-1}$ d'un $ARMA(p,q)$ causal X donné par :

$$\forall t \in \mathbb{Z}, \quad X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q} \ .$$

Le problème est celui de l'initialisation de $X_{-p}, \ldots, X_{-1}$ lorsque $p \neq 0$. On adoptera la règle suivante : on part de $X_{-p} = \ldots = X_{-1} = 0$, et on "grille" les k premières valeurs simulées, avec :

$$k = p + \left\lceil \frac{6}{\log r^*} \right\rceil \ ,$$

où on a défini :

$$r^* = \min\{|\rho| \text{ t.q. } \rho \text{ racine de } \phi\} \ .$$

R: arima.sim()

5.8 Moyenne, autocovariance et autocorrélation empiriques

A partir d'une réalisation $(x_0, \ldots, x_{n-1})$ d'un $ARMA$ de moyenne m, peut-on accéder à la moyenne m et à la fonction de covariance du processus ? Oui, au sens où on peut les estimer, grâce aux propriétés de stationnarité et de mémoire courte de l'$ARMA$. Plus précisément, on estime la moyenne m par la moyenne empirique

$$\overline{X}_n = \frac{1}{n} \sum_{t=0}^{n-1} X_t \ ,$$

l'autocovariance γ_X par l'autocovariance empirique (cf. section 1.2) :

$$\forall h \in \{0, \ldots, n-1\}, \quad \widehat{\gamma}(h) = \frac{1}{n} \sum_{t=0}^{n-h-1} (X_{t+h} - \overline{X}_n)(X_t - \overline{X}_n) \ ,$$

et l'autocorrélation ρ_X par l'autocorrélation empirique :

$$\forall h \in \{0, \ldots, n-1\}, \quad \widehat{\rho}(h) = \frac{\widehat{\gamma}(h)}{\widehat{\gamma}(0)} \ .$$

Les bonnes propriétés de convergence de ces estimateurs seront énoncées au paragraphe 7.1. On trouvera à la Figure 5.1 un exemple d'estimation de l'autocorrélation empirique d'un $AR(2)$, comparé à l'autocorrélation théorique.

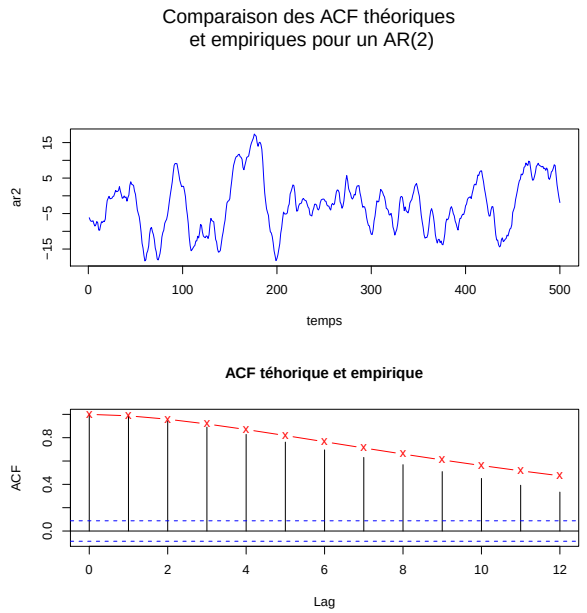

FIGURE 5.1 – Simulation d'un AR(2) et tracé de l'acf théorique et empirique

Chapitre 6

Prédiction linéaire

Un objectif fréquemment rencontré dans l'analyse des séries chronologiques est celui de prédire les valeurs futures d'une série. Dans ce chapitre, nous allons résoudre ce problème pour une série chronologique aléatoire de moyenne et covariances connues en nous restreignant à la *prédiction linéaire*.

Exemple 6.1 *On peut reprendre l'exemple au début de cette partie, qui correspondait à une variante non stationnaire de l'AR*(1) *:*

$$\begin{cases} Y_0 = 0 \\ Y_n = \phi Y_{n-1} + Z_n \quad \forall n \geqslant 1 \end{cases}$$

où $(Z_n)_{n\geqslant 1}$ *est une suite i.i.d de variables aléatoires gaussiennes centrées de variance* $\sigma^2 > 0$ *et* $|\phi| < 1$*. Nous avions vu qu'il était naturel, ayant observé* $Y_0, \ldots, Y_{n-1}$*, de prédire* Y_n *par* ϕY_{n-1}*. De la même manière, si on veut prédire* Y_{n+h-1}*, on a envie de proposer* $\phi^h Y_{n-1}$*. Nous allons voir plus bas que ce sont effectivement de bons choix.*

6.1 Rappels sur la projection orthogonale dans $\mathbb{R}^d$

Soit V un sous-espace vectoriel de $\mathbb{R}^d$. On suppose que $\mathbb{R}^d$ est muni du produit scalaire usuel :

$$\langle x, y\rangle = \sum_{i=1}^{d} x_i y_i \ ,$$

d'où on tire la norme euclidienne :

$$\|x\| = \sqrt{\langle x, x\rangle} = \sqrt{\sum_{i=1}^{d} x_i^2} \ .$$

Soit y un élément de $\mathbb{R}^d$. Il existe un unique élément de V qui soit le plus proche possible de y au sens de la norme euclidienne. Cet élément est appelé projection orthogonale de y sur V, noté $\Pi_V(y)$ et il est caractérisé par les propriétés suivantes (parmi tous les éléments de $\mathbb{R}^d$) :

$$\begin{cases} \Pi_V(y) \in V \\ \forall h \in V, \quad \langle y, h\rangle = \langle \Pi_V(y), h\rangle \end{cases} .$$

6.2 Prédiction linéaire

On suppose données $(X_1, \ldots, X_n, Y)$ des variables aléatoires de carré intégrable, définies sur un même espace de probabilité $(\Omega, \mathcal{F})$. On va chercher à *prédire* Y *de manière linéaire*

en fonction de $(X_1, \ldots, X_n)$ en mesurant la qualité de la prédiction par la norme L^2 (celle de l'*écart quadratique moyen*). Pour être plus précis, notons $\mathrm{Vect}(X_1, \ldots, X_n)$ l'ensemble des combinaisons linéaires de $X_1, \ldots, X_n$:

$$\mathrm{Vect}(X_1, \ldots, X_n) = \{\lambda_1 X_1 + \ldots + \lambda_n X_n \text{ t.q. } \lambda_1, \ldots, \lambda_n \in \mathbb{R}\} .$$

On veut donc trouver Y^* dans $\mathrm{Vect}(X_1, \ldots, X_n)$ tel que :

$$\mathbb{E}[(Y - Y^*)^2] = \min_{Y' \in \mathrm{Vect}(X_1, \ldots, X_n)} \mathbb{E}[(Y - Y')^2] .$$

Le but poursuivi est analogue à celui de la section 6.1, mais dans un espace vectoriel différent de $\mathbb{R}^d$. En effet, notons $E = L^2(\Omega, \mathcal{F})$ l'ensemble des variables aléatoires de carrés intégrables, définies sur $(\Omega, \mathcal{F})$. E est alors un espace vectoriel, et la formule suivante définit un produit scalaire sur E :

$$\forall U_1, U_2 \in E, \quad \langle U_1, U_2 \rangle = \mathbb{E}[U_1 U_2] ,$$

la norme associée étant :

$$\forall U_1 \in E, \quad \|U_1\| = \sqrt{\mathbb{E}[U_1^2]} .$$

On voit alors que le but recherché est, en notant $V = \mathrm{Vect}(X_1, \ldots, X_n)$, qui est un sous-espace vectoriel de E, de trouver l'élément de V le plus proche de Y au sens de la norme définie ci-dessus. C'est donc un problème de projection orthogonale au sens du produit scalaire ci-dessus.

Si on suppose connus Γ_X, la matrice de covariance du vecteur aléatoire $(X_1, \ldots, X_n)$ et les valeurs de $\mathbb{E}(X_j Y)$, pour j allant de 1 à n, alors la proposition suivante nous donne la solution de la prédiction linéaire de Y par $X_1, \ldots, X_n$.

Proposition 6.2 (Prédiction linéaire) *Il exite une unique variable aléatoire Y^* solution de :*

$$Y^* \in \mathrm{Vect}(X_1, \ldots, X_n) \quad \text{et} \quad \mathbb{E}[(Y - Y^*)^2] = \min_{Y' \in \mathrm{Vect}(X_1, \ldots, X_n)} \mathbb{E}[(Y - Y')^2] . \tag{6.1}$$

On l'appelle le meilleur prédicteur linéaire de Y sachant $X_1, \ldots, X_n$ et on le note $P(Y|X_1, \ldots, X_n)$. De plus, (6.1) est équivalente à :

$$Y^* \in \mathrm{Vect}(X_1, \ldots, X_n) \quad \text{et} \quad \forall j = 1, \ldots, n, \quad \mathbb{E}(X_j Y) = \mathbb{E}(X_j Y^*) .$$

Enfin, pour tout $(\alpha_1, \ldots, \alpha_n)$ dans $\mathbb{R}^n$,

$$P(Y|X_1, \ldots, X_n) = \alpha_1 X_1 + \ldots, \alpha_n X_n \Longleftrightarrow \Gamma_X \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} \mathbb{E}(X_1 Y) \\ \vdots \\ \mathbb{E}(X_n Y) \end{pmatrix} .$$

De plus, si $Y^* = P(Y|X_1, \ldots, X_n) = \alpha_1^* X_1 + \ldots + \alpha_n^* X_n$,

$$\mathbb{E}[(Y - Y^*)^2] = \mathbb{E}[Y^2] - (\alpha^*)^{\mathrm{T}} \Gamma_X \alpha^* .$$

Démonstration : On accepte l'existence. On définit la fonction suivante :

$$\Phi : \left\{ \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R}_+ \\ \alpha & \mapsto & \mathbb{E}[(\sum_{i=1}^n \alpha_i X_i - Y)^2] \end{array} \right. .$$

Cette fonction est différentiable en α. Cherchons ses points critiques, c'est à dire les valeurs α qui annulent toutes les dérivées partielles. On rappelle que si Φ admet un minimum, ce minimum est nécessairement un point critique. Pour $j \in \{1, \ldots, n\}$, on a :

$$\frac{\partial \Phi}{\partial \alpha_j} = 2\mathbb{E}[(\sum_{i=1}^n \alpha_i X_i - Y) X_j] .$$

Ainsi, α est un point critique de Φ si et seulement si :

$$\forall j \in \{1, \dots, n\}, \quad \mathbb{E}[\sum_{i=1}^{n} \alpha_i X_i X_j] = \mathbb{E}[Y X_j] \ ,$$

ce qui est équivalent à :

$$\Gamma_X \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} \mathbb{E}(X_1 Y) \\ \vdots \\ \mathbb{E}(X_n Y) \end{pmatrix} \ ,$$

ou encore à :

$$\forall h \in \mathrm{Vect}(X_1, \dots, X_n), \quad \mathbb{E}[h \sum_{i=1}^{n} \alpha_i X_i] = \mathbb{E}[hY] \ . \tag{6.2}$$

Notons maintenant Ψ la fonction que l'on veut minimiser, à savoir :

$$\Psi : \left\{ \begin{array}{rcl} \mathrm{Vect}(X_1, \dots, X_n) & \to & \mathbb{R}_+ \\ Y' & \mapsto & \mathbb{E}[(Y' - Y)^2] \end{array} \right. .$$

Soit α^* vérifiant (6.2), et $h \in \mathrm{Vect}(X_1, \dots, X_n)$. On a alors, en développant les carrés et en simplifiant :

$$\begin{aligned} \Psi(\sum_{i=1}^{n} \alpha_i^* X_i + h) - \Psi(\sum_{i=1}^{n} \alpha_i^* X_i) &= \mathbb{E}[h^2 + 2h(\sum_{i=1}^{n} \alpha_i^* X_i - Y)] \ , \\ &= \mathbb{E}[h^2] \ , \end{aligned}$$

grâce à (6.2). On en déduit que $\Psi(\sum_{i=1}^{n} \alpha_i^* X_i)$ est un minimum pour Ψ, et de plus que ce minimum est atteint en l'unique point $\sum_{i=1}^{n} \alpha_i^* X_i$. Il ne reste alors qu'à démontrer la dernière relation. En notant $Y^* = \sum_{i=1}^{n} \alpha_i^* X_i$, et en remarquant que $\mathbb{E}[YY^*] = \mathbb{E}[(Y^*)^2]$ d'après (6.2),

$$\begin{aligned} \mathbb{E}[(Y - Y^*)^2] &= \mathbb{E}[Y^2] - 2\mathbb{E}[YY^*] + \mathbb{E}[(Y^*)^2] \ , \\ &= \mathbb{E}[Y^2] - \mathbb{E}[(Y^*)^2] \ , \\ &= \mathbb{E}[Y^2] - \mathbb{E}[(\sum_{i=1}^{n} \alpha_i^* X_i)^2] \ , \\ &= \mathbb{E}[Y^2] - \mathbb{E}[\sum_{1 \leqslant i,j \leqslant n} \alpha_i^* \alpha_j^* X_i X_j] \ , \\ &= \mathbb{E}[Y^2] - \sum_{1 \leqslant i,j \leqslant n} \alpha_i^* \alpha_j^* \mathbb{E}[X_i X_j] \ , \\ &= \mathbb{E}[Y^2] - (\alpha^*)^{\mathrm{T}} \Gamma_X \alpha^* \ . \end{aligned}$$

□

R: solve() pour résoudre un système linéaire.

Exemple 6.3 *Voir l'exercice de TD 3.1.*

Enfin, la proposition suivante donne une relation très utile entre la prédiction de Y sachant $X_1, \dots, X_n$ et la prédiction de Y sachant $X_1, \dots, X_{n-1}$.

Proposition 6.4 *On définit I_n comme l'erreur de prédiction de X_n sachant $X_1, \dots, X_{n-1}$:*

$$I_n = X_n - P(X_n | X_1, \dots, X_{n-1}) \ .$$

Alors :

$$P(Y | X_1, \dots, X_n) = P(Y | X_1, \dots, X_{n-1}) + P(Y | I_n) \ .$$

Démonstration : Comme I_n est othogonal (au sens L^2) à $\mathrm{Vect}(X_1, \ldots, X_{n-1})$, c'est un résultat classique de projection orthogonale. Montrons-le tout de même à la main. Il suffit de montrer que :

$$\forall h \in \mathrm{Vect}(X_1, \ldots, X_n), \mathbb{E}[h(P(Y|X_1, \ldots, X_{n-1}) + P(Y|I_n))] = \mathbb{E}[hY] \ . \tag{6.3}$$

Remarquons que tout h de $\mathrm{Vect}(X_1, \ldots, X_n)$ peut s'écrire sous la forme :

$$h = \sum_{i=1}^{n-1} \alpha_i X_i + \alpha_n I_n \ .$$

De plus, I_n est orthogonal à $\mathrm{Vect}(X_1, \ldots, X_n)$. Ainsi,

$$\begin{aligned}
\mathbb{E}[Yh] &= \mathbb{E}[Y \sum_{i=1}^{n-1} \alpha_i X_i] + \mathbb{E}[Y \alpha_n I_n] \ , \\
&= \mathbb{E}[P(Y|X_1, \ldots, X_{n-1}) \sum_{i=1}^{n-1} \alpha_i X_i] + \mathbb{E}[P(Y|I_n) \alpha_n I_n] \ , \\
&= \mathbb{E}[(P(Y|X_1, \ldots, X_{n-1}) + P(Y|I_n))(\sum_{i=1}^{n-1} \alpha_i X_i + \alpha_n I_n)] \ , \\
&= \mathbb{E}[(P(Y|X_1, \ldots, X_{n-1}) + P(Y|I_n))h] \ ,
\end{aligned}$$

ce qui montre (6.3). □

6.3 Application aux processus SLC

Soit X un processus SLC de covariance γ. Un objectif important dans l'étude des séries chronologiques est de *prédire* X_{n+h-1}, avec $h \geqslant 1$ lorsqu'on a observé $X_0, \ldots, X_{n-1}$. C'est ce qu'on appelle la *prédiction à h pas.* La Proposition 6.2 donne alors les résultats suivants.

6.3.1 Prédiction à 1 pas

On pose

$$\begin{aligned}
X_n^* &= P(X_n|X_0, \ldots, X_{n-1}) \ , \\
&= \phi_{n,1} X_{n-1} + \ldots + \phi_{n,n} X_0 \ ,
\end{aligned}$$

et $v_n = \mathsf{Var}(X_n - X_n^*)$. En notant :

$$\Phi_n = \begin{pmatrix} \phi_{n,1} \\ \vdots \\ \phi_{n,n} \end{pmatrix} \quad \gamma_n = \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(n) \end{pmatrix} \ ,$$

et Γ_n la matrice de covariance du vecteur $(X_0, \ldots, X_{n-1})$ ($\Gamma_n(i,j) = \gamma(i-j)$), on obtient le système suivant, à résoudre pour calculer Φ_n et v_n :

$$\begin{cases} \Gamma_n \Phi_n = \gamma_n \\ v_n = \gamma(0) - \gamma_n^{\mathrm{T}} \Phi_n \end{cases} \ . \tag{6.4}$$

6.3.2 Prédiction à h pas

On pose :

$$P(X_{n+h-1}|X_0,\ldots,X_{n-1}) = \phi_{n,1}^{(h)}X_{n-1} + \ldots + \phi_{n,n}^{(h)}X_0 \ ,$$

et $v_n^{(h)} = \mathsf{Var}(X_{n+h-1} - P(X_{n+h-1}|X_0,\ldots,X_{n-1}))$.

$$\Phi_n^{(h)} = \begin{pmatrix} \phi_{n,1}^{(h)} \\ \vdots \\ \phi_{n,n}^{(h)} \end{pmatrix} \quad \gamma_n^{(h)} = \begin{pmatrix} \gamma(h) \\ \vdots \\ \gamma(n+h-1) \end{pmatrix} .$$

On obtient alors le système suivant, à résoudre pour calculer $\Phi_n^{(h)}$ et $v_n^{(h)}$:

$$\begin{cases} \Gamma_n \Phi_n^{(h)} = \gamma_n^{(h)} \\ v_n^{(h)} = \gamma(0) - (\gamma_n^{(h)})^{\mathrm{T}} \Phi_n^{(h)} \end{cases} .$$

Exemple 6.5 *Supposons que X soit un $AR(1)$ défini par :*

$$X_t = \phi X_{t-1} + Z_t \ ,$$

avec $|\phi| < 1$ et Z un bruit blanc de variance σ^2. On a alors, pour tout $k \geqslant 0$:

$$\gamma(k) = \frac{\sigma^2}{1-\phi^2}\phi^k \ .$$

Donc

$$\Gamma_n = \frac{\sigma^2}{1-\phi^2} \begin{pmatrix} 1 & \phi & \ldots & \phi^{n-1} \\ \phi & 1 & \ldots & \phi^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \ldots & 1 \end{pmatrix}$$

et

$$\gamma_n^{(h)} = \frac{\sigma^2}{1-\phi^2} \begin{pmatrix} \phi^h \\ \vdots \\ \phi^{n+h-1} \end{pmatrix} .$$

Il faut alors résoudre le système (6.4). Il est clair que la solution suivante marche :

$$\Phi_n^{(h)} = \begin{pmatrix} \phi^h \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

ce qui donne :

$$P(X_{n+h-1}|X_0,\ldots,X_{n-1}) = \phi^h X_{n-1} \ ,$$

et $v_n^{(h)} = \frac{\sigma^2(1-\phi^{2h})}{1-\phi^2}$.

Pour finir, voici une proposition bien utile : pour les processus stationnaires dont la covariance tend vers 0 à l'infini, les matrices Γ_n sont toujours inversibles.

Proposition 6.6 *Soit X un processus stationnaire (au sens large) tel que :*

– $\gamma_X(0) > 0$,

– *et* $\gamma_X(h) \xrightarrow[h\to+\infty]{} 0$.

Alors, pour tout $n \geqslant 1$, *la matrice* Γ_n *définie par :*

$$\forall 1 \leqslant i,j \leqslant n, \quad \Gamma_n(i,j) = \gamma_X(i-j) \ ,$$

est inversible.

Démonstration : Admise (cf. Proposition 5.1.1 page 167 de [3]). □

6.4 Projection sur tout le passé pour un ARMA causal et inversible

Il est parfois commode d'avoir une approximation de la prédiction, plus facile à calculer "de tête" que celle obtenue dans la section 6.3.

Définition 6.7 *Soit* X *un ARMA causal et inversible. Soit* $H_n^X = \overline{\mathrm{Vect}}((X_t)_{t\leqslant n})$ *le passé linéaire du processus* X. *On note* $P(X_n|H_{n-1}^X)$ *la projection, au sens* L^2, *de* X_n *sur* H_{n-1}^X. *On l'appelle prédiction de* X_n *sachant* H_{n-1}^X *et elle est cractérisée par :*

$$\begin{cases} P(X_n|H_{n-1}^X) \in H_{n-1}^X \\ \forall t \leqslant n-1, \quad \mathbb{E}(X_t X_n) = \mathbb{E}(X_t P(X_n|H_{n-1}^X)) \end{cases}$$

La proposition suivante donne la valeur de la prédiction à un pas pour un $ARMA$ en fonction de tout son passé.

Proposition 6.8 *Soit* X *un ARMA causal et inversible, défini par :*

$$\forall t \in \mathbb{Z}, \quad \phi(B)X_t = \theta(B)Z_t \ .$$

Alors, pour tout $n \in \mathbb{Z}$, $H_n^X = H_n^Z$. *De plus, si :*

$$\forall t \in \mathbb{Z}, \quad Z_t = \sum_{j\geqslant 0} \lambda_j X_{t-j} \ ,$$

alors,

$$P(X_n|H_{n-1}^X) = -\sum_{j\geqslant 1} \lambda_j X_{n-j} = X_n - Z_n \ .$$

Démonstration : Comme X est inversible, il existe une suite de réels λ_k tels que :

$$\forall t \in \mathbb{Z}, \quad Z_t = \sum_{j\geqslant 0} \lambda_j X_{t-j} \ ,$$

avec $\lambda_0 = 1$. On voit ainsi que $Z_t \in H_n^X$ pour tout $t \leqslant n$. Donc $H_n^Z \subset H_n^X$. D'un autre côté, comme X est causal, il existe une suite ψ_k telle que :

$$\forall t \in \mathbb{Z}, \quad X_t = \sum_{k\geqslant 0} \psi_k Z_{t-k} \ ,$$

et donc $X_t \in H_n^Z$ pour tout $t \leqslant n$. Donc $H_n^X \subset H_n^Z$. Posons :

$$Y_n = -\sum_{j\geqslant 1} \lambda_j X_{n-j} \ .$$

On a alors $Y_n \in H_{n-1}^X$, mais aussi, $Y_n = X_n - Z_n$. Soit $k \leqslant n-1$, on a alors :

$$\mathbb{E}(Y_n X_k) = \mathbb{E}(X_n X_k) - \mathbb{E}(Z_n X_k) \ ,$$

il reste donc à montrer que $\mathbb{E}(Z_n X_k) = 0$. Mais comme Z est un bruit blanc, $\mathbb{E}(Z_n W) = 0$ pour toute v.a. W appartenant à H_{n-1}^Z. Or $X_k \in H_{n-1}^Z$ pour tout $k \leqslant n-1$, d'après ce qu'on vient de montrer (la causalité suffit pour cela). □

Bien sûr, en pratique, on n'observe jamais tout le passé d'un $ARMA$, et si l'on veut utiliser la Proposition 6.8, on est amené à tronquer la somme infinie $\sum_{j\geqslant 1} \pi_j X_{n-j}$, ce qui donne un résultat correct lorsque n est grand. Si on veut un calcul exact lorsqu'on ne dispose que d'un nombre fini d'observations, il faut utiliser les formules de la section 6.3.1. Si on doit faire ce calcul pour plusieurs valeurs de n, l'inversion de Γ_X étant coûteuse, on peut utiliser un algorithme récursif comme le filtre de Kalman, cf. section 8.

6.5 Autocorrélation partielle

Définition 6.9 (Corrélation partielle) *Si X, Y sont deux variables aléatoires réelles, et U un vecteur aléatoire, on définit la* corrélation partielle de X et Y sachant U, *notée $\rho(X, Y|U)$, comme la corrélation entre les résidus de la prédiction de X sachant U et de la prédiction de Y sachant U :*

$$\rho(X, Y|U) = \frac{\mathsf{Cov}(X - P(X|U), Y - P(Y|U))}{\sqrt{\mathsf{Var}(X - P(X|U))\mathsf{Var}(Y - P(Y|U))}} \ .$$

Définition 6.10 (Fonction d'autocorrélation partielle) *La* fonction d'autocorrélation partielle *d'un processus SLC X, notée α_X, est définie par :*

$$\forall h \in \mathbb{Z}, \quad \alpha_X(h) = \rho(X_t, X_{t+h}|X_{t+1}, \ldots, X_{t+h-1}) \ ,$$

qui ne dépend pas de t, car X est stationnaire.

Proposition 6.11 *Soit X un processus SLC tel que :*
- $\gamma_X(0) > 0$,
- *et* $\gamma_X(h) \xrightarrow[h\to+\infty]{} 0$.

Avec les notations de la section 6.3.1, on a :

$$\alpha_X(h) = \phi_{hh} \ .$$

Démonstration : Tout d'abord, comme $\mathbb{E}[P(Y|U)(X - P(X|U))] = 0$, on a pour tous X, Y et U comme dans la Définition 6.9,

$$\rho(X, Y|U) = \frac{\mathsf{Cov}(X - P(X|U), Y)}{\sqrt{\mathsf{Var}(X - P(X|U))\mathsf{Var}(Y - P(Y|U))}} \ .$$

Maintenant, soit X un processus SLC. Comme X est centré, on a :

$$\alpha_X(h) = \frac{\mathbb{E}[(X_{t+h} - P(X_{t+h}|X_{t+h-1}, \ldots, X_{t+1}))X_t]}{\sqrt{\mathsf{Var}[X_{t+h} - P(X_{t+h}|X_{t+h-1}, \ldots, X_{t+1})]\mathsf{Var}[X_t - P(X_t|X_{t+h-1}, \ldots, X_{t+1})]}} \ .$$

En utilisant la Proposition 6.4 :

$$P(X_{t+h}|X_{t+h-1}, \ldots, X_t) = P(X_{t+h}|X_{t+h-1}, \ldots, X_{t+1}) + P(X_{t+h}|I) \ ,$$

où :

$$I = X_t - P(X_t|X_{t+h-1}, \ldots, X_{t+1}) \ .$$

Il existe un coefficient β tel que :

$$P(X_{t+h}|I) = \beta I = \beta X_t - \beta P(X_t|X_{t+h-1}, \ldots, X_{t+1}) \ .$$

Donc :

$$P(X_{t+h}|X_{t+h-1}, \ldots, X_t) = \beta X_t + \text{une combinaison linéaire en } X_{t+h-1}, \ldots, X_{t+1}$$

Or, par définition,

$$P(X_{t+h}|X_{t+h-1}, \ldots, X_t) = \phi_{h,1} X_{t+h-1} + \ldots + \phi_{h,h} X_t \ .$$

D'après la Proposition 6.6, les coefficients $\phi_{h,i}$ sont uniquement déterminés, donc on en déduit que $\beta = \phi_{h,h}$. On a donc :

$$\begin{aligned}
& \mathbb{E}[(X_{t+h} - P(X_{t+h}|X_{t+h-1}, \ldots, X_{t+1}))X_t] \\
= \ & \mathbb{E}[(X_{t+h} - P(X_{t+h}|X_{t+h-1}, \ldots, X_t))X_t] + \mathbb{E}[P(X_{t+h}|I)X_t] \ , \\
= \ & 0 + \mathbb{E}[\phi_{h,h}(X_t - P(X_t|X_{t+h-1}, \ldots, X_{t+1}))X_t] \ , \\
= \ & \phi_{h,h} \mathsf{Var}[X_t - P(X_t|X_{t+h-1}, \ldots, X_{t+1})] \ .
\end{aligned}$$

Enfin, comme X est stationnaire,

$$\mathsf{Var}[X_t - P(X_t|X_{t+h-1}, \ldots, X_{t+1})] = \mathsf{Var}[X_{t+h} - P(X_{t+h}|X_{t+h-1}, \ldots, X_{t+1})] \ ,$$

et il suffit de recoller les bouts. □

6.6 Caractérisation des AR et des MA

Voici un résultat qui sera crucial quand viendra l'heure d'estimer l'ordre d'un $ARMA$ (cf. section 7.4).

Proposition 6.12 *Soit X un processus $ARMA$.*

1. *X est un $AR(p)$ si et seulement si $\alpha_X(h) = 0$ pour tout $h > p$.*
2. *X est un $MA(q)$ si et seulement si $\rho_X(h) = 0$ pour tout $h > q$.*

Démonstration :

1. Il est clair que si X est un $AR(p)$, alors $\alpha_X(h) = 0$ pour tout $h > p$. Traitons la réciproque, supposons que $\alpha_X(h) = 0$ pour tout $h > p$. Pour $h = p + 1$, on en déduit :

$$P(X_t|X_{t-1}, \ldots, X_{t-p-1}) \in \mathrm{Vect}(X_{t-1}, \ldots, X_{t-p}) \ .$$

Donc :

$$P(X_t|X_{t-1}, \ldots, X_{t-p-1}) = P(X_t|X_{t-1}, \ldots, X_{t-p}) \ .$$

De même, pour tout $h \geqslant p$, on obtient :

$$P(X_t|X_{t-1}, \ldots, X_{t-h-1}) \in \mathrm{Vect}(X_{t-1}, \ldots, X_{t-h}) \ ,$$

et donc :

$$P(X_t|X_{t-1}, \ldots, X_{t-h-1}) = P(X_t|X_{t-1}, \ldots, X_{t-h}) \ .$$

Ce qui donne par récurrence, pour tout $h \geqslant p$

$$P(X_t|X_{t-1}, \ldots, X_{t-h}) = P(X_t|X_{t-1}, \ldots, X_{t-p}) \ .$$

On pose alors :

$$Z_t := X_t - P(X_t|X_{t-1}, \ldots, X_{t-p}) \ .$$

On aura fini si on montre que Z_t est un bruit blanc. Mais on a, pour $k \geqslant 1$:

$$\begin{aligned}
\mathbb{E}(Z_t Z_{t-k}) &= \mathbb{E}[(X_t - P(X_t|X_{t-1}, \ldots, X_{t-p}))(X_{t-k} - P(X_{t-k}|X_{t-k-1}, \ldots, X_{t-k-p}))] \ , \\
&= \mathbb{E}[(X_t - P(X_t|X_{t-1}, \ldots, X_{t-k-p}))(X_{t-k} - P(X_{t-k}|X_{t-k-1}, \ldots, X_{t-k-p}))] \ , \\
&= 0 \ ,
\end{aligned}$$

puisque $X_{t-k} \in \mathrm{Vect}(X_{t-1}, \ldots, X_{t-k-p})$ et $\mathrm{Vect}(X_{t-k-1}, \ldots, X_{t-k-p}) \subset \mathrm{Vect}(X_{t-1}, \ldots, X_{t-k-p})$.

2. Il est clair que si X est un $MA(q)$, alors $\rho_X(h) = 0$ pour tout $h > q$. Supposons maintenant que $\rho_X(h) = 0$ pour tout $h > q$. On peut écrire :

$$X_t = \sum_{k \geqslant 0} \psi_k Z_{t-k} \ ,$$

où Z est un bruit blanc de variance $\sigma^2 > 0$ et ψ est absolument sommable et vérifie $\psi_0 = 1$. Par conséquent, pour tout $k \geqslant 0$,

$$\psi_k = \frac{1}{\sigma^2} \mathbb{E}[X_t Z_{t-k}] \ .$$

On sait aussi que $H_n^X = H_n^Z$ pour tout $n \in \mathbb{Z}$. Par conséquent, $Z_{t-k} \in H_{t-k}^X$. Comme, par hypothèse,

$$\mathbb{E}(X_t X_{t-k}) = 0$$

pour tout $k > q$, on en déduit que :

$$\mathbb{E}(X_t Z_{t-k}) = 0$$

pour tout $k > q$. Donc, $\psi_k = 0$ pour tout $k > q$, et ainsi, X est un $MA(q)$.

□

Chapitre 7

Statistique sur les ARMA

Le but de ce chapitre est d'énoncer la marche à suivre ainsi que les résultats théoriques concernant l'estimation des paramètres d'un processus $ARMA$. Plus précisément, on observe n valeurs consécutives $(X_1, \ldots, X_n)$ d'un $ARMA(p, q)$ causal de moyenne non nécessairement nulle m :

$$\phi(B)(X - m)_t = \theta(B)Z_t \ ,$$

avec Z un bruit blanc i.i.d de variance $\sigma^2 > 0$. En général, m, p et q sont inconnus, ainsi que $\beta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$ et σ^2. On veut pouvoir :

- estimer p et q,
- estimer m, β et σ^2,
- prédire $X_n, \ldots, X_{n+h-1}$, assortis d'intervalles de confiance.

Aucun des résultats théoriques de ce chapitre ne sera démontré. Une référence possible est [3], section 10.8, page 375.

7.1 Moyenne, autocovariance et autocorrélation empiriques

Nous donnons d'abord des résultats de convergence pour les estimateurs empiriques de la moyenne et de l'autocorrélation. Ces résultats permettent d'obtenir des intervalles de confiance pour la moyenne et l'autocorrélation théoriques d'un $ARMA$.

7.1.1 Moyenne empirique

On note $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ la moyenne empirique.

Proposition 7.1

$$\sqrt{n}(\overline{X}_n - m) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, s^2) \ ,$$

où :

$$s^2 = \sum_{h\in\mathbb{Z}} |\gamma_X(h)| \ .$$

7.1.2 Autocovariance et autocorrélation empiriques

On estime l'autocovariance γ_X par :

$$\forall h \in \{0, \ldots, n-1\}, \quad \widehat{\gamma}(h) = \frac{1}{n}\sum_{i=1}^{n-h}(X_{i+h} - \overline{X}_n)(X_i - \overline{X}_n) \ ,$$

Noter que la normalisation (par n et non $n-h$) n'est pas la même qu'au paragraphe 5.8. Ceci assure que la matrice $\widehat{\Gamma}_h$ définie au paragraphe 7.2.1 soit inversible. On estime l'autocorrélation ρ_X par :

$$\forall h \in \{0, \dots, n-1\}, \quad \widehat{\rho}(h) = \frac{\widehat{\gamma}(h)}{\widehat{\gamma}(0)} .$$

Proposition 7.2 *Soit ρ la fonction d'autocorrélation de X. Soit h fixé dans $\mathbb{N}^*$ et :*

$$\widehat{\rho}_h = \begin{pmatrix} \widehat{\rho}(1) \\ \vdots \\ \widehat{\rho}(h) \end{pmatrix} \quad \rho_h = \begin{pmatrix} \rho(1) \\ \vdots \\ \rho(h) \end{pmatrix} .$$

Alors,

$$\sqrt{(}n)(\widehat{\rho}_h - \rho_h) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, \Sigma) ,$$

où :

$$\forall i, j \in \{1, \dots, h\}, \quad \Sigma_{i,j} = \sum_{l=1}^{+\infty} [\rho(l+i) + \rho(l-i) - 2\rho(l)\rho(i)].[\rho(l+j) + \rho(l-j) - 2\rho(l)\rho(j)] .$$

7.2 Estimation des paramètres d'un ARMA d'ordre connu

7.2.1 Méthode des moments pour un AR

On pose :

$$\widehat{\Gamma}_h = \begin{pmatrix} \widehat{\gamma}(0) & \widehat{\gamma}(1) & \dots & \widehat{\gamma}(h-1) \\ \widehat{\gamma}(1) & \widehat{\gamma}(0) & \dots & \widehat{\gamma}(h-2) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\gamma}(h-1) & \widehat{\gamma}(h-2) & \dots & \widehat{\gamma}(0) \end{pmatrix}$$

$$\widehat{\gamma}_h = \begin{pmatrix} \widehat{\gamma}(1) \\ \vdots \\ \widehat{\gamma}(h) \end{pmatrix} .$$

Si $\widehat{\gamma}(0) \neq 0$, $\widehat{\Gamma}_h$ est inversible et on note $\widehat{\Phi}_h$ le vecteur :

$$\widehat{\Phi}_h = \begin{pmatrix} \widehat{\phi}_{h,1} \\ \vdots \\ \widehat{\phi}_{h,h} \end{pmatrix} ,$$

unique solution du système :

$$\widehat{\Gamma}_h \widehat{\Phi}_h = \widehat{\gamma}_h .$$

7.2.1.1 PACF empirique

Définition 7.3 *L'auocorrélation partielle empirique $\widehat{\alpha}$ est définie par :*

$$\forall h \in \{1, \dots, n-1\}, \quad \widehat{\alpha}(h) = \widehat{\phi}_{hh} .$$

7.2.1.2 Le cas d'un AR(p) causal

On estime $\phi = (\phi_1, \ldots, \phi_p)$ par $\widehat{\phi} = (\widehat{\phi}_{p1}, \ldots, \widehat{\phi}_{pp})$ et σ^2 par :

$$\widehat{\sigma}^2 = \widehat{\gamma}(0) - (\widehat{\gamma}_p)^T \widehat{\Phi} \ .$$

Proposition 7.4

$$\widehat{\sigma}^2 \xrightarrow[n\to+\infty]{\mathbb{P}} \sigma^2 \ ,$$

et :

$$\sqrt{n}(\widehat{\phi} - \phi) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2 \Gamma_p^{-1}) \ ,$$

7.2.2 Estimation d'un ARMA gaussien par maximum de vraisemblance

On rappelle que $\beta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$. On suppose qu'on observe n valeurs consécutives $(X_1, \ldots, X_n)$ d'un $ARMA(p,q)$ causal de moyenne nulle :

$$\phi(B)X_t = \theta(B)Z_t \ ,$$

avec Z un bruit blanc gaussien i.i.d de variance $\sigma^2 > 0$. $X = (X_1, \ldots, X_n)$ est alors un vecteur gaussien de moyenne nulle. On peut montrer que l'estimateur du maximum de vraisemblance est asymptotiquement gaussien :

Proposition 7.5

$$\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_{p+q}(0, V(\beta)) \ ,$$

où :

$$\sigma^2 V(\beta)^{-1} = \begin{pmatrix} \mathsf{Cov}_\beta(\overrightarrow{U}) & \mathsf{Cov}_\beta(\overrightarrow{U}, \overrightarrow{V}) \\ Cov_\beta(\overrightarrow{V}, \overrightarrow{U}) & \mathsf{Cov}_\beta(\overrightarrow{V}) \end{pmatrix}$$

avec :

$$\overrightarrow{U} = \begin{pmatrix} U_1 \\ \vdots \\ U_p \end{pmatrix} \quad \overrightarrow{V} = \begin{pmatrix} V_1 \\ \vdots \\ V_q \end{pmatrix} \ ,$$

et :

$$\begin{cases} U_t = \phi_1 U_{t-1} + \ldots + \phi_p U_{t-p} + Z_t \\ V_t = -\theta_1 V_{t-1} - \ldots - \theta_q V_{t-q} + Z_t \end{cases}$$

Enfin,

$$\sqrt{n}\left(\frac{\widehat{\sigma}^2}{\sigma^2} - 1\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1) \ .$$

7.2.3 Pseudo-vraisemblance gaussienne

Même si l'$ARMA$ n'est pas gaussien, on peut utiliser le résultat précédent pour estimer (β, σ^2) par les mêmes $(\widehat{\beta}, \widehat{\sigma}^2)$. On suppose donc que X est un $ARMA(p,q)$ causal :

$$\phi(B)X_t = \theta(B)Z_t \ ,$$

avec Z un bruit blanc i.i.d de variance $\sigma^2 > 0$. Alors, les conclusions de la Proposition 7.5 sont toujours valables.

7.3 Validation

La validation du modèle s'effectuera en examinant les rédidus standardisés :

$$\forall i \in \{1, \dots, n\}, \quad \widehat{Z}_i = \frac{I_i(\widehat{\beta})}{\sqrt{\widehat{\sigma}^2 v_i(\widehat{\beta})}} .$$

1. Si le bruit blanc Z_t est un bruit blanc i.i.d, ces résidus doivent être approximativement i.i.d. Si le bruit blanc Z_t est un bruit blanc i.i.d gaussien, ces résidus doivent être approximativement i.i.d $\mathcal{N}(0,1)$.
2. On examine la fonction d'autocorrélation empirique $\widehat{\rho}_Z$ des résidus standardisés sur une vingtaine ou une trentaine de valeurs. Si l'hypothèse sur l'ordre (p,q) est bonne, et que le bruit blanc Z_t est un bruit blanc i.i.d, $(\widehat{\rho}_Z(1), \dots, \widehat{\rho}_Z(h))$ forment approximativement un vecteur de gaussiennes i.i.d de loi $\mathcal{N}(0, \frac{1}{n})$.
3. On effectue un test de Ljung-Box. La statistique de Ljung-Box est :

$$Q_{LB} = n(n+2) \sum_{k=1}^{K} \frac{\widehat{\rho}_Z(k)}{n-k} ,$$

   où K est à choisir. Si l'hypothèse sur l'ordre (p,q) est bonne, et que le bruit blanc Z_t est un bruit blanc i.i.d, Q_{LB} doit suivre approximativement une loi du chi-deux à $K-p-q$ degrés de liberté. On rejettera notre modèle au niveau asymptotique α si :

$$Q_{LB} > q_{1-\alpha, K-p-q} ,$$

   où $q_{1-\alpha,K-p-q}$ est le quantile d'ordre $1-\alpha$ de la loi du chi-deux à $K-p-q$ degrés de liberté. En pratique, on prend en général K valant entre 10 et 30. Lorsqu'on veut se prémunir contre une possible saisonnalité conservée dans les résidus, on prend souvent K valant une ou deux fois la valeur de la période.

Enfin, *on considèrera notre modèle comme validé* s'il passe ces trois épreuves. Remarquons pour finir qu'il y a des cas où on peut craindre un défaut particulier de notre modèle qui serait de ne pas inclure un certain degré de liberté supplémentaire (par exemple, on ajuste un $AR(4)$, mais on se demande si on n'aurait pas plutôt affaire à un $AR(6)$ avec $\phi_5 = 0$). Dans ce genre de cas, il est bien sûr conseillé d'ajuster le modèle le plus grand, puis de tester l'appartenance au sous-modèle, par exemple en testant la nullité de certains coefficients (i.e ajuster un $AR(6)$ avec $\phi_5 = 0$ dans notre exemple, puis tester $\phi_6 = 0$).

R: residuals() permet de récupérer les résidus d'un ajustement avec arima().
qqnorm() et qqlines() pour inspecter la normalité des résidus.
Box.test(type="Ljung-Box") pour le test de Ljung-Box, avec les options "fitdf" pour donner le nombre de paramètres estimés (en général $p+q$) et "lag" pour préciser la valeur de K.

7.4 Détermination de l'ordre d'un ARMA

7.4.1 Détermination de l'ordre d'un AR

Un $AR(p)$ est caractérisé par $\alpha_X(h) = 0$ pour tout $h > p$ (cf. Proposition 6.12). D'après l'asymptotique de la pacf empirique (cf. Proposition 7.4), on peut déduire que si X est un $AR(p)$, alors pour tout $h > p$, $\widehat{\alpha}(h)$ doit se comporter comme une gaussienne $\mathcal{N}(0, \frac{1}{n})$. On estimera p par :

$$\widehat{p} = \min \left\{ p \geqslant 0 \text{ t.q. } \forall h \geqslant p+1,\ \widehat{\alpha}(h) \in \left[-\frac{1.96}{\sqrt{n}}, \frac{1.96}{\sqrt{n}} \right] \right\} .$$

En fait, le "$\forall h \geqslant p+1$" est un peu trop demandé, on tolèrera quelques faibles dépassements. On n'oublie pas de valider le modèle (section 7.3). Si le modèle n'est pas validé, on peut essayer d'augmenter la valeur de $\widehat{p}$.

R: pacf()

7.4.2 Détermination de l'ordre d'un MA

Un $MA(q)$ est caractérisé par $\rho_X(h) = 0$ pour tout $h > q$ (cf. Proposition 6.12). D'après l'asymptotique de l'autocorrélation empirique (cf. Proposition 7.2), on en déduit que si X est un $MA(q)$, alors pour tout $h > q$, $\widehat{\rho}(h)$ doit se comporter comme une gaussienne :

$$\mathcal{N}(0, \frac{1}{n}(1 + 2\rho^2(1) + \ldots + 2\rho^2(q))) \ .$$

On estimera donc q par :

$$\widehat{q} = \min \left\{ q \geqslant 0 \text{ t.q. } \forall h \geqslant q+1,\ \widehat{\rho}(h) \in \left[-\frac{1.96\sqrt{\widehat{\sigma}_q^2}}{\sqrt{n}}, \frac{1.96\sqrt{\widehat{\sigma}_q^2}}{\sqrt{n}} \right] \right\} \ ,$$

avec :

$$\widehat{\sigma}_q^2 = 1 + 2\widehat{\rho}^2(1) + \ldots + 2\widehat{\rho}^2(q)$$

En fait, le "$\forall h \geqslant q+1$" est un peu trop demandé, on tolèrera quelques faibles dépassements. On n'oublie pas de valider le modèle (section 7.3). Si le modèle n'est pas validé, on peut essayer d'augmenter la valeur de $\widehat{q}$.

R: acf() avec l'option `ci.type="ma"`.

7.4.3 Détermination de l'ordre d'un ARMA

On recherche l'ordre $(\widehat{p}, \widehat{q})$ de telle sorte que $\widehat{p}+\widehat{q}$ soit le plus petit possible, tout en conservant un modèle validé. Pour cela, on adoptera la démarche suivante :

1. On fait l'hypothèse qu'on a un AR pur, et on identifie $\widehat{p}_1$ comme dans la section 7.4.1.
2. On fait l'hypothèse qu'on a un MA pur, et on identifie $\widehat{q}_1$ comme dans la section 7.4.2.
3. On essaie de réduire le nombre de paramètres. Pour cela, on pose $\widehat{M} = \max\{\widehat{p}_1, \widehat{q}_1\}$, et on cherche (recherche exhaustive ou à tâtons) un modèle $(\widehat{p}, \widehat{q})$ parmi les (p, q) tels que $p + q \leqslant \widehat{M}$. On s'efforcera de trouver $(\widehat{p}, \widehat{q})$ tel que $\widehat{p} + \widehat{q}$ soit le plus petit possible, tout en conservant un modèle validé. *Si nécessaire*, on peut utiliser un critère comme le critère AIC pour distinguer plusieurs modèles à peu près équivalents (il faut alors garder celui qui a l'AIC minimal).

7.5 Prédiction de valeurs futures

Une fois qu'on dispose d'un modèle $ARMA(p, q)$ validé, on peut l'utiliser pour prédire les valeurs futures : on agira comme pour prédire un $ARMA$ de paramètre $\widehat{\beta}, \widehat{\sigma}^2$, avec les variances des erreurs de prédiction associées. Si on suppose le bruit gaussien, on donnera des intervalles de confiance gaussiens. Sinon, on pourra utiliser une estimation empirique des quantiles des erreurs de prédiction.

R: predict() ou predict.Arima() sur une sortie d'arima()

Troisième partie

Modèles de séries chronologiques non stationnaires

Maintenant que nous disposons d'un modèle simple de bruit corrélé, les $ARMA$ du chapitre 5, nous allons l'utiliser comme base pour construire des modèles plus complexes, notamment des modèles de séries chronologiques non stationnaires.

Nous commencerons dans le chapitre 8 par présenter une représentation très utile, valable pour certains modèles : le modèle d'état. C'est en fait une représentation par chaîne de Markov linéaire qui permet de calculer récursivement (grâce au *filtre de Kalman*) des quantités essentielles comme les valeurs successives des prédictions linéaires. Les $ARMA$, mais aussi certaines de leurs généralisations, les modèles à fonction de transfert, les $ARIMA$ et $SARIMA$, ont une représentation en modèle d'état.

L'objectif du reste du chapitre sera donc de présenter quelques modèles plus complexes de séries chronologiques, largement utilisés dans les applications. Ils sont regroupés en deux catégories : ceux qui mettent en évidence des variables explicatives (régression à résidu $ARMA$ et modèle à fonction de transfert) dans le chapitre 9, et les autres ($ARIMA$, $SARIMA$) dans le chapitre 10.

Chapitre 8

Modèles d'état et filtre de Kalman

8.1 Modèle d'état

Définition 8.1 *On dit qu'un processus vectoriel* $\overrightarrow{Y} = (\overrightarrow{Y}_t)_{t\geqslant 0}$ *à valeurs dans* $\mathbb{R}^k$ *admet un* modèle d'état *(linéaire et homogène en temps) lorsqu'il existe des matrices* F*,* G*,* Q *et* R *(les paramètres du modèle d'état) telles que les équations suivantes soient vérifiées :*

$$\left\{\begin{array}{lll} \forall t \geqslant 0, & \overrightarrow{X}_{t+1} & = & F\overrightarrow{X}_t + \overrightarrow{V}_t \quad \textit{Equation d'état} \\ \forall t \geqslant 0, & \overrightarrow{Y}_t & = & G\overrightarrow{X}_t + \overrightarrow{W}_t \quad \textit{Equation d'observation} \end{array}\right. \tag{8.1}$$

où $(X_0, (\overrightarrow{V}_t)_{t\geqslant 0}, (\overrightarrow{W}_t)_{t\geqslant 0})$ *est une collection de vecteurs aléatoires décorrélés, de seconds moments finis et :*

$(\overrightarrow{X}_t)_{t\geqslant 0}$	*est un processus appelé* processus des états
$(\overrightarrow{V}_t)_{t\geqslant 0}$	*est un processus appelé* bruit d'état*, de moyenne nulle et de matrice de covariance* Q *de taille* (r, r)
F	*est une matrice de taille* (r, r)
$(\overrightarrow{Y}_t)_{t\geqslant 0}$	*est un processus appelé* processus des observations
$(\overrightarrow{W}_t)_{t\geqslant 0}$	*est un processus appelé* bruit d'observation*, de moyenne nulle et de matrice de covariance* R *de taille*
G	*est une matrice de taille* (k, r)

Remarque 8.2 *On peut étendre les résultats de ce chapitre à des cas où les matrices* F*,* G*,* Q *et* R *dépendent du temps, et où il existe une matrice* S_t *non nécessairement nulle telle que :*

$$\mathbb{E}[\overrightarrow{V}_t \overrightarrow{W}_t^T] = S_t \ .$$

Exemple 8.3 Modèle d'état pour les $ARMA$

Les $ARMA$ *rentrent dans le cadre des modèles d'état. En effet, soit* Y *un* $ARMA(p, q)$ *causal (et de moyenne nulle) avec :*

$$\forall t \in \mathbb{Z}, \quad \phi(B)Y_t = \theta(B)Z_t \ ,$$

où Z *est un bruit blanc de variance* σ^2*. On définit* U *l'*$AR(p)$ *causal :*

$$\forall t \in \mathbb{Z}, \quad \phi(B)U_t = Z_t \ ,$$

On pose $r = \max\{p, q+1\}$*,* $\theta_0 = 1$*,* $\phi_k = 0$ *si* $k > p$*,* $\theta_k = 0$ *si* $k > q$*, et :*

$$\overrightarrow{X}_t = \begin{pmatrix} U_t \\ \vdots \\ U_{t-r+1} \end{pmatrix}$$

$$\overrightarrow{V}_t = \begin{pmatrix} Z_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \textit{de covariance } Q = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

$$F = \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 & \dots & \phi_r \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}$$

$$G = (\theta_0, \dots, \theta_r)$$

Alors, l'ARMA Y vérifie le modèle d'état 8.1.

8.2 Le filtre de Kalman

On suppose que $\overrightarrow{Y}$ suit le modèle d'état 8.1. Si l'on connait les valeurs de $P_0(\overrightarrow{X}_1) = \mathbb{E}(\overrightarrow{X}_1)$ et $C_1 = \mathsf{Cov}(\overrightarrow{X}_1)$, le filtre de Kalman permet de calculer récursivement les valeurs de $P(\overrightarrow{Y}_t|1, \overrightarrow{Y}_1, \dots, \overrightarrow{Y}_{t-1})$ pour $t = 0, \dots, n$ sur la base des observations $(\overrightarrow{Y}_1, \dots, \overrightarrow{Y}_n)$ ainsi que les covariances des erreurs de prédiction.

1. **Notations.** Pour un vecteur $\overrightarrow{A}$ de dimension l, on note $P_t(\overrightarrow{A})$ le vecteur :

$$P_t(\overrightarrow{A}) = \begin{pmatrix} P(A_1|1, \overrightarrow{Y}_1, \dots, \overrightarrow{Y}_t) \\ \vdots \\ P(A_l|1, \overrightarrow{Y}_1, \dots, \overrightarrow{Y}_t) \end{pmatrix}.$$

   Remarquons que $P_0(\overrightarrow{A}) = P(\overrightarrow{A}|1) = \mathbb{E}(\overrightarrow{A})$. On note aussi :

$$\begin{cases} \overrightarrow{I}_t &= \overrightarrow{Y}_t - P_{t-1}(\overrightarrow{Y}_t) \\ S_t &= \mathsf{Cov}(\overrightarrow{I}_t) \\ C_t &= \mathsf{Cov}[\overrightarrow{X}_t - P_{t-1}(\overrightarrow{X}_t)] \\ D_t &= \mathsf{Cov}[\overrightarrow{X}_t - P_t(\overrightarrow{X}_t)] \end{cases}$$

2. **Initialisation.**
   (a) $P_0(\overrightarrow{X}_1) = \mathbb{E}(\overrightarrow{X}_1)$
   (b) $C_1 = \mathsf{Cov}(\overrightarrow{X}_1)$
3. **Récurrence.** Pour t allant de 1 à n :
   (a) $P_{t-1}(\overrightarrow{Y}_t) = GP_{t-1}(\overrightarrow{X}_t)$
   (b) $\begin{cases} \overrightarrow{I}_t &= \overrightarrow{Y}_t - P_{t-1}(\overrightarrow{Y}_t) \\ S_t &= GC_tG^T + R \end{cases}$
   (c) $K_t = C_tG^TS_t^{-1}$
   (d) $\begin{cases} P_t(\overrightarrow{X}_t) &= P_{t-1}(\overrightarrow{X}_t) + K_t\overrightarrow{I}_t \\ D_t &= C_t - K_tS_tK_t^T \end{cases}$
   (e) $\begin{cases} P_t(\overrightarrow{X}_{t+1}) &= FP_t(\overrightarrow{X}_t) \\ C_{t+1} &= FD_tF^T + Q \end{cases}$

Démonstration : On note $\overrightarrow{Y_0}$ la variable aléatoire constante égale à 1.

(a) Rappelons que W_t est de moyenne nulle et décorrélé de $\overrightarrow{Y}_1, \ldots, \overrightarrow{Y}_{t-1}$, donc :

$$P_{t-1}(\overrightarrow{Y}_t) = GP_{t-1}(\overrightarrow{X}_t) + P_t(\overrightarrow{W_t}) = GP_{t-1}(\overrightarrow{X}_t) \ .$$

(b) On a :

$$\begin{aligned} S_t &= \mathbb{E}(\overrightarrow{I}_t \overrightarrow{I}_t^T) \ , \\ &= G\mathbb{E}[\overrightarrow{X}_t \overrightarrow{X}_t^T]G^T - G\mathbb{E}[P_{t-1}(\mathbf{X_{t-1}})P_{t-1}(\mathbf{X_{t-1}})^T]G^T + R \ , \\ &= GC_tG^T + R \ . \end{aligned}$$

(d) Pour estimer $\overrightarrow{X}_t$ à l'instant t, i.e en ayant observé $\overrightarrow{Y}_1, \ldots, \overrightarrow{Y}_t$, on décompose la projection à l'aide de la Proposition 6.4 :

$$P(\overrightarrow{X}_t|\overrightarrow{Y}_0, \ldots, \overrightarrow{Y}_t) = P(\overrightarrow{X}_t|\overrightarrow{Y}_0, \ldots, \overrightarrow{Y}_{t-1}, \overrightarrow{I}_t) = P(\overrightarrow{X}_t|\overrightarrow{Y}_0, \ldots, \overrightarrow{Y}_{t-1}) + P(\overrightarrow{X}_t|\overrightarrow{I}_t) \ .$$

C'est à dire $P_t(\overrightarrow{X}_t) = P_{t-1}(\overrightarrow{X}_t) + P(\overrightarrow{X}_t|\overrightarrow{I}_t)$. En utilisant la Proposition 6.2 :

$$P(\overrightarrow{X}_t|\overrightarrow{I}_t) = \left(S_t^{-1}\mathbb{E}(\overrightarrow{I}_t\overrightarrow{X}_t^T)\right)^T \overrightarrow{I}_t = \mathbb{E}(\overrightarrow{X}_t\overrightarrow{I}_t^T)S_t^{-1}\overrightarrow{I}_t \ ,$$

où $S_t = \mathbb{E}(\overrightarrow{I}_t\overrightarrow{I}_t^T)$ est la matrice de covariance des innovations à l'instant t. On pose $K_t = \mathbb{E}(\overrightarrow{X}_t\overrightarrow{I}_t^T)S_t^{-1}$. Il faut évaluer récursivement K_t. En écrivant $\overrightarrow{I}_t = G(\overrightarrow{X}_t - P_{t-1}(\overrightarrow{X}_t)) + W_t$, et en notant que W_t est décorrélé de $\overrightarrow{X}_t$,

$$\mathbb{E}(\overrightarrow{X}_t\overrightarrow{I}_t^T) = \mathbb{E}[\overrightarrow{X}_t(\overrightarrow{X}_t - P_{t-1}(\overrightarrow{X}_t))^T]G^T = C_tG^T \ ,$$

où $C_t = \mathbb{E}[(\overrightarrow{X}_t - P_{t-1}(\overrightarrow{X}_t))(\overrightarrow{X}_t - P_{t-1}(\overrightarrow{X}_t))^T]$. On a donc :

$$K_t = C_tG^TS_t^{-1} \ ,$$

et :

$$P_t(\overrightarrow{X}_t) = P_{t-1}(\overrightarrow{X}_t) + K_t\overrightarrow{I}_t \ .$$

Enfin,

$$\begin{aligned} D_t = C_t - \mathbb{E}[(P_t(\overrightarrow{X}_t) - P_{t-1}(\overrightarrow{X}_t))(P_t(\overrightarrow{X}_t) - P_{t-1}(\overrightarrow{X}_t))^T] &= C_t - \mathbb{E}[(K_t\overrightarrow{I}_t)(K_t\overrightarrow{I}_t)^T] \ , \\ &= C_t - K_tS_tK_t^T \ . \end{aligned}$$

(e) Comme V_t est de moyenne nulle et décorrélé de $\overrightarrow{Y}_1, \ldots, \overrightarrow{Y}_t$,

$$P_t(\overrightarrow{X}_t) = P_t(F\overrightarrow{X}_t + V_t) = FP_t(\overrightarrow{X}_t) + P_t(V_t) = FP_t(\overrightarrow{X}_t) \ .$$

On a aussi :

$$\begin{aligned} C_{t+1} &= \mathbb{E}[\overrightarrow{X}_{t+1}\overrightarrow{X}_{t+1}^T] - \mathbb{E}[P_t(\overrightarrow{X}_t)P_t(\overrightarrow{X}_t)^T] \ , \\ &= F\mathbb{E}[\overrightarrow{X}_t\overrightarrow{X}_t^T]F^T + Q - F\mathbb{E}[P_t(\mathbf{X_t})P_t(\mathbf{X_t})^T]F^T \ , \\ &= FD_tF^T + Q \ . \end{aligned}$$

□

Remarque 8.4 *Si une valeur est manquante, il faut adapter les équations du filtre de Kalman comme suit :*

$$\overrightarrow{I}_t = 0$$

et

$$K_t = 0$$

Remarque 8.5 Initialisation pour un $ARMA$
On reprend le modèle d'état décrit à l'exemple 8.3. On a alors $P_0(\overrightarrow{X_1}) = \overrightarrow{0}$ *et* C_1 *est la matrice de covariance de* $(U_1, \ldots, U_r)$. *Ainsi :*

$$C_1 = \begin{pmatrix} \gamma(0) & \gamma(1) & \ldots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \ldots & \gamma(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \ldots & \gamma(0) \end{pmatrix}$$

où γ *est la fonction de covariance de l'AR(p)* U.

Remarquons que pour calculer C_1 *en fonction des paramètres de* U, *on peut utiliser les équations aux différences. Pour un AR(p), celles-ci s'écrivent facilement. On montre que le vecteur :*

$$\gamma_r = \begin{pmatrix} \gamma(0) \\ \vdots \\ \gamma(r) \end{pmatrix}$$

est l'unique solution de l'équation :

$$A\gamma_r = g \ ,$$

où :

$$g = \begin{pmatrix} \sigma_Z^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad A = A_1 + A_2 \ ,$$

$$A_1 = \begin{pmatrix} 0 & -\phi_1 & -\phi_2 & \ldots & -\phi_{r-1} & -\phi_r \\ 0 & -\phi_2 & -\phi_3 & \ldots & -\phi_r & 0 \\ \vdots & \vdots & \vdots & \iddots & \iddots & \vdots \\ 0 & -\phi_{r-1} & \phi_r & 0 & \ldots & 0 \\ 0 & -\phi_r & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 \end{pmatrix}$$

$$A_2 = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ -\phi_1 & 1 & 0 & \ldots & 0 & 0 \\ -\phi_2 & -\phi_1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ -\phi_{r-1} & -\phi_{r-2} & \ldots & -\phi_1 & 1 & 0 \\ -\phi_r & -\phi_{r-1} & -\phi_{r-2} & \ldots & -\phi_1 & 1 \end{pmatrix}$$

Remarque 8.6 Initialisation inconnue
Lorsqu'on ne connait pas précisément la valeur de $P_0(X_1)$ *(et donc en général de* C_1*), on peut donner une estimation approximative de celle-ci, et traduire cette incertitude dans* C_1, C_1 *devant être d'autant plus "grande" que l'incertitude est grande.*

8.3 Filtre de Kalman et vraisemblance gaussienne

Proposition 8.7 *Soit $(U_1, \ldots, U_n)$ un vecteur de variables aléatoires centrées de carré intégrable. Notons Γ_n la matrice de covariance du vecteur $(U_1, \ldots, U_n)$:*

$$\forall i, j \in \{1, \ldots, n\}, \quad [\Gamma_n]_{i,j} = \mathsf{Cov}(U) := \mathbb{E}(U_i U_j) \ ,$$

et U_i^* la prédiction linéaire de U_i sachant $U_1, \ldots, U_{i-1}$. On suppose que Γ_n est inversible. Alors,

$$(U - \mathbb{E}(U))^T \Gamma_n^{-1} (U - \mathbb{E}(U)) = \gamma(U) = \sum_{i=1}^{n} \frac{(U_i - U_i^*)^2}{v_{i-1}} \ ,$$

où $v_{i-1} = \mathbb{E}[(U_i - U_i^*)^2]$. De plus,

$$det(\Gamma_n) = \prod_{i=1}^{n} v_{i-1} \ .$$

Démonstration : Remarquons d'abord que les innovations $(I_i)_{i=1,\ldots,n}$, où $I_i = U_i - U_i^*$ sont décorrélées entre elles. Ainsi, leur matrice de covariance est la matrice diagonale D telle que $D_{ii} = v_{i-1}$. Par ailleurs, U_i^* est une fonction linéaire de $U_1, \ldots, U_{i-1}$. Donc il existe une matrice C triangulaire avec des éléments diagonaux égaux à 1 et telle que $CI = U$ (en notant I le vecteur des innovations). Par conséquent :

$$\Gamma_n = \mathsf{Cov}(U) = \mathsf{Cov}(C^{-1} I) = CDC^T \ ,$$

et ainsi,

$$U^T \Gamma_n^{-1} U = U^T (C^T)^{-1} D^{-1} C^{-1} U = I^T D^{-1} I = \sum_{i=1}^{n} \frac{(U_i - U_i^*)^2}{v_{i-1}} \ .$$

Et bien sûr, $det(\Gamma_n) = det(C)^2 det(D) = \prod_{i=1}^{n} v_{i-1}$. □

L'intérêt de la Proposition 8.7 réside surtout dans le fait que lorsque $(U_1, \ldots, U_n)$ est un vecteur gaussien centré, la vraisemblance est exactement une fonction de $U\Gamma_n^{-1}U^T$ et $det(\Gamma_n)$. En effet, la vraisemblance vaut :

$$\frac{1}{(\sqrt{2\pi det(\Gamma_n)})^n} e^{-\frac{1}{2} U \Gamma_n^{-1} U} = \frac{1}{(\sqrt{2\pi \prod_{i=1}^{n} v_{i-1}})^n} e^{-\frac{1}{2} \sum_{i=1}^{n} \frac{(U_i - U_i^*)^2}{v_{i-1}}} \ .$$

Le filtre de Kalman, combiné à la Proposition 8.7 permet de calculer rapidement (avec un ordinateur) la vraisemblance gaussienne d'un vecteur satisfaisant un modèle d'état. Précisons cela dans le cas d'un processus $ARMA$. On suppose qu'on observe n valeurs consécutives $(X_1, \ldots, X_n)$ d'un $ARMA(p, q)$ causal de moyenne nulle :

$$\phi(B) X_t = \theta(B) Z_t \ ,$$

avec Z un bruit blanc gaussien i.i.d de variance $\sigma^2 > 0$. On note $\beta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$. $X = (X_1, \ldots, X_n)$ est alors un vecteur gaussien de moyenne nulle. On note $X_i^*(\beta)$ la prédiction linéaire de X_i sachant $X_1, \ldots, X_{i-1}$, dans le modèle où on suppose $\sigma^2 = 1$. Notation :

$$X_i^*(\beta) = P_{(\beta,1)}(X_i | X_1, \ldots, X_{i-1}) \ .$$

Puis, on note :

$$\begin{cases} I_i(\beta) & = & X_i - X_i^*(\beta) \\ v_i(\beta) & = & \mathsf{Var}_{\beta,1}(I_i(\beta)) \end{cases}$$

Proposition 8.8 *L'estimateur du maximum de vraisemblance* $(\widehat{\beta}, \widehat{\sigma}^2)$ *de* (β, σ^2) *est donné par :*

$$\widehat{\beta} = \underset{\beta \in \mathbb{R}^{p+q}}{\operatorname{argmin}} \left[\frac{1}{n} \sum_{i=1}^{n} \log v_i(\beta) + \log \left(\sum_{i=1}^{n} \frac{(I_i(\beta))^2}{v_i(\beta)} \right) \right] ,$$

et

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \frac{(I_i(\widehat{\beta}))^2}{v_i(\widehat{\beta})} .$$

Chapitre 9

Variables explicatives et modèle à fonction de transfert

Dans ce chapitre, on considère des séries chronologiques qui s'écrivent sous la forme :

$$\forall t \in \mathbb{Z},\ Y_t = \sum_{i=1}^{k} \beta_i u^{(i)}(t) + \varepsilon_t \ ,$$

où ε est un bruit $ARMA$ de moyenne nulle, β un paramètre inconnu et les $u^{(i)}$ sont des variables explicatives. On considèrera deux cas :

- lorsque les $u^{(i)}$ sont des fonctions déterministes du temps, on aboutit à ce qu'on appelle la *régression à résidu ARMA*,
- lorsque $u^{(i)}(t) = X_{t-i}$, où X est lui-même un $ARMA$ observé, on aboutit à ce qu'on appelle un *modèle à fonction de transfert*.

9.1 Régression à résidu ARMA

Etant données k séries chronologiques déterministes $u^{(1)}, \ldots, u^{(k)}$, on considère le modèle suivant :

$$\forall t \in \mathbb{Z},\ Y_t = \sum_{i=1}^{k} \beta_i u^{(i)}(t) + \varepsilon_t \ ,$$

où ε est un bruit $ARMA(p,q)$ de moyenne nulle et $\beta \in \mathbb{R}^k$ est inconnu. En d'autres termes, $Y - \sum_{i=1}^{k} \beta_i u^{(i)}$ est un $ARMA$ de moyenne nulle. On peut donc estimer (par maximum de vraisemblance gaussien) simultanément β et les paramètres de ε dès qu'on connait l'ordre (p,q) de ε. On adoptera la méthode suivante si on ne connait pas cet ordre (ce qui est le cas général) :

1. on estime β par moindres carrés ordinaires,
2. sur $\widehat{\varepsilon} := Y - \sum_{i=1}^{k} \widehat{\beta}_i u^{(i)}(t)$, on estime l'ordre (p,q) de ε,
3. on revient à une estimation simultanée de β et des paramètres de ε par maximum de vraisemblance gaussien.

R: arima() avec régresseur externe : option `"xreg="`

Exemple 9.1 Saisonnalité *Il est courant que des phénomènes périodiques se produisent (influence des saisons, des jours de travail ou de repos, des heures de transport etc.) et aient une énorme influence sur les séries chronologiques (penser au débit d'un fleuve alimenté par la fonte d'un glacier au printemps). Dans ce cas, on pourra inclure une non-stationnarité sous la forme*

d'une fonction périodique de période p, ce qu'on appelle aussi une saisonnalité de période p. *Cela rentre dans le cadre ci-dessus, en prenant :*

$$u^{(i)}(t) = \begin{cases} 1 & \textit{si } p \textit{ divise } t-i \\ 0 & \textit{sinon} \end{cases}$$

Exemple 9.2 Tendance polynômiale *Il existe souvent une non stationnarité sous la forme d'une* tendance à long terme, *c'est à dire que la moyenne à moyen terme de la série évolue lentement sur l'ensemble de la série. Cela peut parfois être modélisé par un polynôme de degré d en la variable t. Cela rentre également dans le cadre ci-dessus, en prenant :*

$$u^{(i)}(t) = t^i \ .$$

On peut bien sûr combiner une tendance polynômiale avec une saisonnalité, et faire varier les formes de tendances lentes en prenant d'autres fonctions que des polynômes.

9.2 Modèle à fonction de transfert

Imaginons que l'on observe un couple d'$ARMA$ (X_t, Y_t), aux instants $t = 1, \ldots, n$ et qu'il existe un lien de causalité entre X et Y. Peut-on tirer partie de ce lien pour prédire de façon plus fine Y ? L'idée des modèles à fonction de transfert est précisément celle-ci, en tenant compte d'un lien de causalité linéaire entre X et Y. Plus précisément, on considère Z et W deux bruits blancs décorrélés, un $ARMA$ X défini par :

$$X_t = \frac{\theta^X(B)}{\phi^X(B)} Z_t \ , \tag{9.1}$$

où θ^X et ϕ^X sont des polynômes, un $ARMA$ N défini par :

$$N_t = \frac{\theta^N(B)}{\phi^N(B)} W_t \ , \tag{9.2}$$

où θ^N et ϕ^N sont des polynômes, et un processus Y défini par :

$$Y_t = \frac{\nu(B)}{\omega(B)} X_t + N_t \ , \tag{9.3}$$

où $\nu(B)$ et $\omega(B)$ sont des polynômes. On suppose comme d'habitude que les filtres $\frac{\theta^X(B)}{\phi^X(B)}$, $\frac{\theta^N(B)}{\phi^N(B)}$ et $\frac{\nu(B)}{\omega(B)}$ sont causaux et que :

$$\theta_0^X = \phi_0^X = \theta_0^N = \phi_0^N = \omega_0 = 1 \ .$$

Par contre, on ne suppose pas nécessairement que $\nu_0 = 1$. Comme ν_0 peut être nul, on écrit parfois $\nu(B)$ sous la forme $B^b \widetilde{\nu}(B)$ avec $b \geqslant 0$ et $\widetilde{\nu}_0 \neq 0$. Le paramètre b est alors appelé *le délai* du filtre $\frac{\nu(B)}{\omega(B)}$. *Pour finir, on suppose qu'on observe (X_t, Y_t), aux instants $t = 1, \ldots, n$.*

Remarque 9.3 *Les $ARMA$ N et X étant décorrélés, il est possible de montrer que Y est également un processus $ARMA$. Le modèle présenté est à moyenne nulle, on peut le modifier facilement pour introduire des moyennes non nécessairement nulles.*

9.2.1 Modèle d'état

On remarque que l'équation du modèle de transfert (9.3) peut s'écrire :

$$Y_t = \nu(B)\theta^X(B)\left(\frac{1}{\omega(B)\phi^X(B)}\right)Z_t + \theta^N(B)\left(\frac{1}{\phi^N(B)}\right)W_t\ .$$

On définit deux AR par :

$$U^X = \left(\frac{1}{\omega(B)\phi^X(B)}\right)Z\ ,$$

et

$$U^N = \left(\frac{1}{\phi^N(B)}\right)W\ .$$

On a alors :

$$Y_t = \nu(B)\theta^X(B)U_t^X + \theta^N(B)U_t^N\ , \tag{9.4}$$

et

$$X_t = \omega(B)\theta^X(B)U_t^X\ . \tag{9.5}$$

On peut donc mettre en oeuvre la même stratégie que pour un $ARMA$ afin de représenter le modèle à fonction de transfert par un modèle d'état :

- l'équation d'état témoignera de l'évolution indépendante des deux AR U^X et U^N,
- l'équation d'observation s'obtient ensuite facilement à partir des équations (9.5) et (9.4).

Plus précisément, si d_ω est le degré de $\omega(B)$ et p^X le degré de ϕ^X, on écrit :

$$\widetilde{\omega}(B) := \omega(B)\phi^X(B) = 1 - \widetilde{\omega}_1 B - \ldots - \widetilde{\omega}_{d_\omega+p^X}B^{d_\omega+p^X}\ ,$$

et on pose $\widetilde{\omega}_i = 0$ si $i > d_\omega + p$. De même, on définit d_ν le degré de ν et q^X le degré de θ^X, et on écrit :

$$\widetilde{\nu}(B) := \nu(B)\theta^X(B) = 1 + \widetilde{\nu}_1 B + \ldots + \widetilde{\nu}_{d_\nu+q^X}B^{d_\nu+q^X}\ ,$$

On pose alors $r^X = \max\{d_\omega + p^X, d_\omega + q^X + 1, d_\nu + q^X\}$,

$$F^X = \begin{pmatrix} \widetilde{\omega}_1 & \widetilde{\omega}_2 & \widetilde{\omega}_3 & \ldots & \widetilde{\omega}_{r^X} \\ 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots, 0 & 1 & 0 & \end{pmatrix}, \quad Q^X = \begin{pmatrix} \sigma_Z^2 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{pmatrix},$$

$$G^X = (\theta_0^X, \ldots, \theta_{r^X-1}^X), \quad \text{et} \quad G^{YX} = (\widetilde{\nu}_0, \ldots, \widetilde{\nu}_{r^X-1})\ .$$

Pour le modèle d'état de N, on pose $r^N = \max\{p^N, q^N + 1\}$, où p^N est le degré de ϕ^N et q^N celui de θ^N, puis :

$$F^N = \begin{pmatrix} \phi_1^N & \phi_2^N & \phi_3^N & \ldots & \phi_{r^N}^N \\ 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots, 0 & 1 & 0 & \end{pmatrix}, \quad Q^N = \begin{pmatrix} \sigma_W^2 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{pmatrix},$$

et

$$G^N = (\theta_0^N, \ldots, \theta_{r^N-1}^N)\ .$$

Enfin, on définit les matrices par blocs :

$$F = \begin{pmatrix} F^X & 0 \\ 0 & F^N \end{pmatrix}, \quad Q = \begin{pmatrix} Q^X & 0 \\ 0 & Q^N \end{pmatrix}, \quad \text{et} \quad G = \begin{pmatrix} G^X & 0 \\ G^{YX} & G^N \end{pmatrix}\ .$$

Alors, le modèle d'état défini par F, Q et G est équivalent au modèle de transfert défini par les équations (9.1), (9.2) et (9.3).

9.2.2 Identification

On décrit ci-dessous la méthode dite de Box et Jenkins pour identifier les différents paramètres du modèle à fonction de transfert lorsque Z est un bruit blanc i.i.d. Commençons par une remarque théorique. En multipliant l'équation (9.3) par $\frac{\phi^X(B)}{\theta^X(B)}$, on obtient :

$$\frac{\phi^X(B)}{\theta^X(B)}Y_t = \frac{\nu(B)}{\omega(B)}Z_t + \frac{\phi^X(B)}{\theta^X(B)}N_t \ .$$

Posons $Y'_t = \frac{\phi^X(B)}{\theta^X(B)}Y_t$, et :

$$\frac{\nu(B)}{\omega(B)} = \sum_{k\geqslant 0}\psi_k B^k \ .$$

On a alors, pour $k \geqslant 0$:

$$\mathbb{E}[Y'_t Z_{t-k}] = \psi_k \sigma_Z^2 \ .$$

Si on définit la fonction de corrélation théorique de Y' et Z par :

$$\rho_{YZ}(k) = \frac{\mathsf{Cov}[Y'_t, Z_{t-k}]}{\sqrt{\mathsf{Var}(Y'_t)\mathsf{Var}(Z_t)}} \ ,$$

on obtient donc :

$$\psi_k = \rho_{Y'Z}(k)\frac{\sigma_{Y'}}{\sigma_Z} \ .$$

Lorsqu'on observe uniquement X et Y, ceci suggère la possibilité d'estimer les valeurs de ψ_k en commençant par estimer les valeurs de Z. Plus précisément, on utilisera la procédure suivante.

1. On identifie l'ordre (p^X, q^X) et les paramètres de l'$ARMA$ X, et on forme les innovations $\widehat{Z}_t$, $t = 1, \ldots, n$.
2. On calcule $\widehat{Y}'_t = \frac{\widehat{\phi}^X(B)}{\widehat{\theta}^X(B)}Y_t$, $t = 1, \ldots, n$ et la fonction de corrélation croisée empirique entre $\widehat{Y}'$ et $\widehat{Z}$. Si le modèle à fonction de transfert est valide, $\widehat{\rho}_{Y'Z}(k)$ doit suivre asymptotiquement une loi $\mathcal{N}(0, \frac{1}{n})$ lorsque $\rho_{Y'Z}(k)$ est nulle, donc notamment pour $k < 0$.
3. Pour les valeurs k auxquelles on peut considérer que $\rho_{Y'Z}(k)$ est significativement différente de zéro, on estime ψ_k par :
$$\widehat{\psi}_k := \widehat{\rho}_{Y'Z}(k)\frac{\widehat{\sigma}_{Y'}}{\widehat{\sigma}_Z} \ .$$
4. Il faut alors, au vu des valeurs des $\widehat{\psi}_k$, proposer une hypothèse pour les degrés de $\nu(B)$ et $\omega(B)$, et le paramètre de délai b. C'est en général l'étape la plus délicate.
5. On estime le bruit N en posant :
$$\widehat{N}_t = Y_t - \sum_{k\geqslant 0}\widehat{\psi}_k X_{t-k} \ .$$
6. On ajuste un modèle $ARMA$ sur $\widehat{N}$.
7. On a alors une hypothèse sur tous les ordres des $ARMA$ apparaissant dans le modèle à fonction de transfert. Lorsque $\omega(B) = 1$, on a même une estimation de tous les coefficients. Dans tous les cas, on peut alors estimer ou réestimer toutes les valeurs des coefficients par maximum de vraisemblance gaussien, en utilisant la représentation en modèle d'état de la section 9.2.1.
8. Toutes les estimations d'$ARMA$ doivent être validées par les techniques de la section 7.3.

Chapitre 10

ARIMA et SARIMA

Dans ce chapitre, une approche différente de la non-stationnarité est proposée. On s'intéresse à des séries chronologiques aléatoires qui se ramènent, après un nombre suffisant de *différentiations* à des $ARMA$.

10.1 ARIMA

10.1.1 Définition

On rappelle que B désigne l'opérateur retard. On définit l'opréateur de différentiation par $\nabla = (1 - B)$. Autrement dit :

$$\nabla X_t = X_t - X_{t-1} \ .$$

Définition 10.1 *Soit $d \geqslant 0$ un nombre entier naturel. On dit qu'une série chronologique aléatoire $(X_t)_{t \geqslant 0}$ est un $ARIMA(p, d, q)$ s'il existe un $ARMA(p, q)$ Y de moyenne nulle tel que :*

$$\forall t \geqslant 0, \quad Y_t = \nabla^d X_t \ .$$

On dit aussi que X est un processus $ARMA(p, q)$ de moyenne nulle intégré d fois.

X satisfait donc une équation de la forme :

$$\forall t \geqslant 0, \ \Phi(B)\nabla^d X_t = \Theta(B) Z_t \tag{10.1}$$

où Φ et Θ sont des polynômes de degrés respectifs p et q, et Z est un bruit blanc. Remarquons que si $\nabla^d X_t$ est un $ARMA(p, q)$ *de moyenne non nulle* m, alors

$$\forall t \geqslant 0, \quad X_t = m\frac{t^d}{d!} + \widetilde{X}_t \ ,$$

où $\widetilde{X}_t$ est un $ARIMA(p, d, q)$. Les processus $ARIMA(p, d, q)$ ne sont pas stationnaires dès que $d \geqslant 1$.

La Figure 10.1 représente une réalisation d'un processus $ARIMA(1, 1, 0)$ avec $\phi = 0.9$ et sa fonction de corrélation empirique. Une telle fonction de corrélation empirique (à décroissance lente) est un bon indicateur du fait que l'on a affaire à un processus $ARIMA$ (ou qu'il existe une tendance dans la série).

La Figure 10.2 montre la corrélation empirique et la corrélation partielle empirique du processus obtenu en différentiant une fois le processus d'origine. La simple observation de ces quatre figures suggère fortement que le processus considéré est un $ARIMA(1, 1, 0)$.

R: Pour estimer un $ARIMA(p, d, q)$ par maximum de vraisemblance gaussien : arima() en précisant `order=(p,d,q)`.

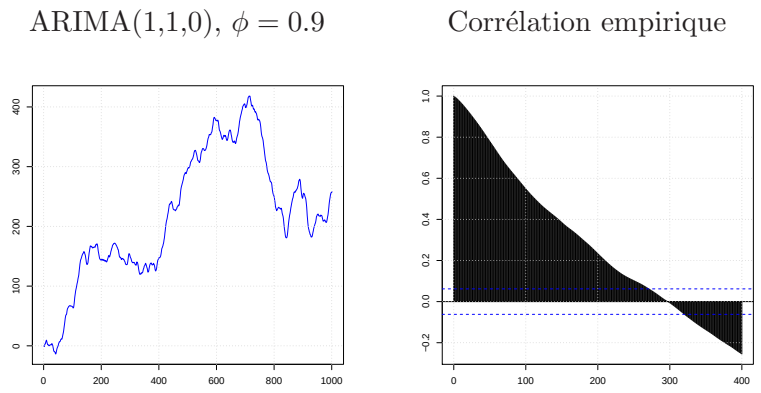

FIGURE 10.1 – Un $ARIMA(1,1,0)$ avec $\phi = 0.9$ et sa fonction de corrélation empirique

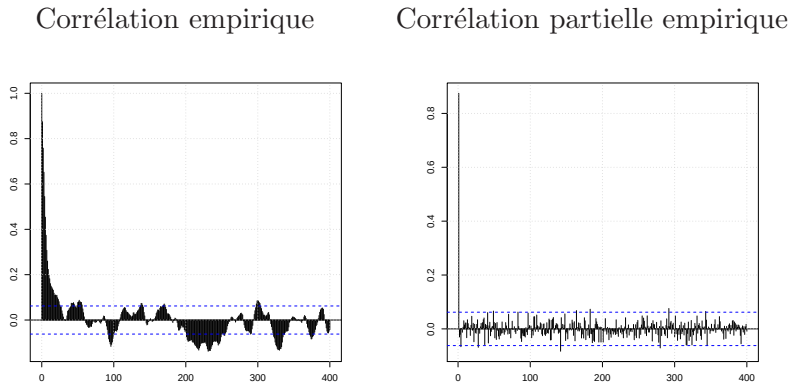

FIGURE 10.2 – Corrélations empiriques du processus de la Figure 10.1 différentié une fois

10.1.2 Identification

Soit X une série dont on suppose qu'elle suit un modèle $ARIMA(p,d,q)$ avec p, d et q inconnus.

1. On différencie X jusqu'à observer une série stationnaire $Y = \nabla^d X$.
2. On identifie (p,q) et on ajuste un $ARMA(p,q)$ sur la série X.
3. On revient à l'identification de l'$ARIMA(p,d,q)$ par maximum de vraisemblance gaussien sur la série originale X.

En général, après quelques différentiations, la série semble stationnaire, mais on ne sait pas assurer qu'elle est de moyenne nulle. On a alors un modèle où

$$\forall t \geqslant 0, \quad X_t = m\frac{t^d}{d!} + \widetilde{X}_t \ ,$$

avec $\widetilde{X}_t$ est un $ARIMA(p,d,q)$. On peut alors modifier l'identification comme suit :

1. On différencie X jusqu'à observer une série stationnaire $Y = \nabla^d X$.
2. On identifie (p,q) et on ajuste un $ARMA(p,q)$ de moyenne non nécessairement nulle sur la série X.
3. On teste la nullité de la moyenne.
4. Si on rejette la nullité de la moyenne, on ajuste sur la série originale X un modèle $ARIMA(p,d,q)$ avec régresseur $(1:n)^d$ par maximum de vraisemblance gaussien. Si on accepte l'hypothèse de nullité de la moyenne, on ajuste un modèle $ARIMA(p,d,q)$.

10.2 SARIMA

10.2.1 Définition

Les modèles $SARIMA$ (S pour Seasonal, c'est à dire saisonnier) permettent de rendre compte des variations saisonnières dans la série considérée, ces variations pouvant elles-mêmes présenter un caractère aléatoire.

La Figure 10.3 représente des parties de réalisations de divers niveaux de processus $SARIMA$, générés à partir d'une même séquence de bruit blanc. On trouve successivement (de gauche à droite puis de haut en bas) :

- un processus $ARMA$ présentant une composante saisonnière de période 20,
- le même processus intégré une fois, à la période de l'échantillon,
- le processus initial intégré une fois, à la période de la saison,
- le processus a été intégré une fois à la période de l'échantillon, et une fois à la période de la saison.

Soit $s \geqslant 1$ un entier. On définit l'opérateur de différentiation à l'échelle de la période s par :

$$\nabla_s = (1 - B^s) \ .$$

Définition 10.2 *Soient $d \geqslant 0$, $D \geqslant 0$ et $s \geqslant 1$ des entiers. On dit qu'une série chronologique aléatoire $(X_t)_{t\geqslant 0}$ est un $SARIMA(p,d,q) \times (P,D,Q)_s$ s'il existe des polynômes Φ, $\Phi^{(s)}$, Θ, $\Theta^{(s)}$ et un $ARMA$ Y de moyenne nulle tels que :*

$$\forall t \geqslant 0, \quad Y_t = \nabla^d \nabla_s^D X_t \ ,$$

$$\Phi(B)\Phi^{(s)}(B^s)Y_t = \Theta(B)\Theta^{(s)}(B^s)Z_t \ ,$$

où Z est un bruit blanc, Φ est de degré p, $\Phi^{(s)}$ de degré P, Θ de degré q et $\Theta^{(s)}$ de degré Q.

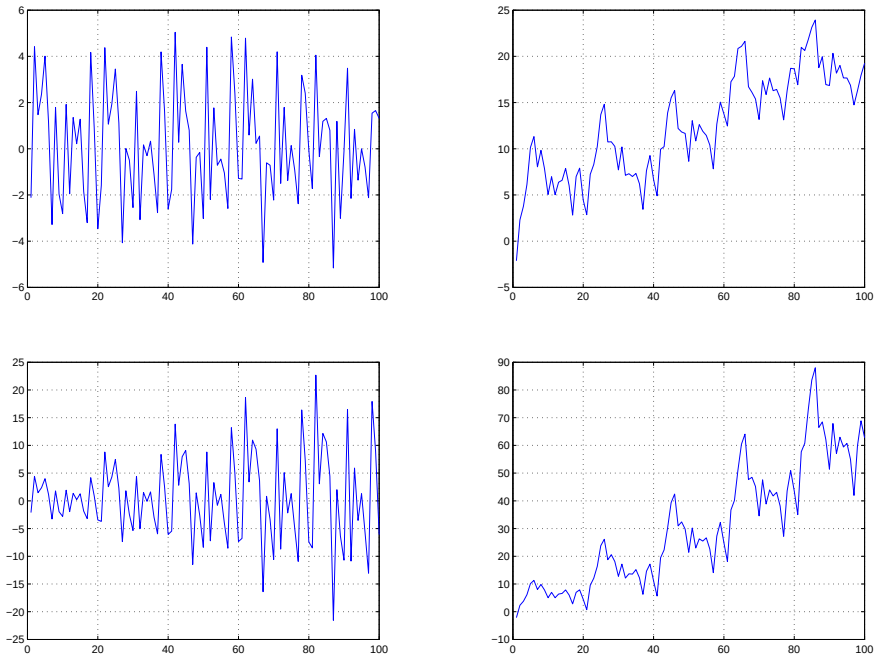

FIGURE 10.3 – De l'$ARMA$ au $SARIMA$

Il est intéressant de voir l'$ARMA$ Y sous la forme suivante :

$$Y_t = \frac{\Theta^{(s)}(B^s)}{\Phi^{(s)}(B^s)} N_t \ , \tag{10.2}$$

où N_t est l'$ARMA(p,q)$:

$$N_t = \frac{\Theta(B)}{\Phi(B)} Z_t \ . \tag{10.3}$$

Exemple 10.3 *Soit N un $ARMA(p,q)$ et Y un processus défini par :*

$$Y_t = 0.9 Y_{t-12} + N_t \ .$$

Un tel processus est une extension assez naturelle de la notion de saisonnalité : si le bruit N est assez faible, Y n'est pas loin d'être périodique, au sens où Y_t est fortement corrélé à Y_{t-12}, mais cette relation est bruitée par un bruit autocorrélé N. Y est un $SARIMA(p,0,q) \times (1,0,0)_{12}$.

10.2.2 Identification

Dans un processus $SARIMA$, il y a interaction entre deux modèles : celui de l'équation (10.3) décrit les évolutions à la période de l'échantillon, c'est-à-dire entre deux saisons successives, et celui de l'équation (10.2) décrit les évolutions à la période de la saison. Cette interaction rend complexe l'interprétation de la covariance empirique dans le cas général. On indique ici quelques principes élémentaires pour une identification simple, dans l'hypothèse où il y a un "découplage" entre les deux modèles. Plus précisément, notons ρ la fonction de corrélation de l'$ARMA$ $\frac{\Theta(B)}{\Phi(B)} Z_t$ donné par l'équation (10.3), et $\rho^{(s)}$ la fonction de corrélation de l'$ARMA$ $\frac{\Theta^{(s)}(B^s)}{\Phi^{(s)}(B^s)} Z_t$. On a alors :

$$\forall h \in \mathbb{Z}, \rho_Y(h) = C \rho * \rho^{(s)}(h) := C \sum_{k \in \mathbb{Z}} \rho(k) \rho^{(s)}(h-k) \ ,$$

où C est une constante. Remarquons que $\rho^{(s)}(h)$ est nulle lorsque h n'est pas un multiple de s, et que $\rho^{(s)}(ks) = \widetilde{\rho}^{(s)}(k)$ où $\widetilde{\rho}^{(s)}$ est la fonction d'autocorrélation de l'$ARMA$ $\frac{\Theta^{(s)}(B)}{\Phi^{(s)}(B)} Z_t$. *On suppose que $\rho(h)$ est quasiment nulle lorsque $|h| \geqslant s/2$* : c'est l'hypothèse de découplage. Dans ce cas, il est facile de voir que :

$$\forall |h| \leqslant s/2, \quad \rho_Y(h) \simeq \rho(h) \ ,$$

et :

$$\forall k \in \mathbb{Z}, \quad \rho_Y(ks) \simeq \rho^{(s)}(ks) \ .$$

Cela suggère d'utiliser la procédure suivante (en supposant s connu) :

1. trouver les ordres d et D tels que le processus $Y_t = (1-B)^d(1-B^s)^D X_t$ soit stationnaire,
2. utiliser la corrélation empirique et la corrélation partielle empirique pour les valeurs multiples de s pour en déduire des valeurs préliminaires des ordres P et Q et les coefficients correspondants,
3. utiliser la corrélation empirique et la corrélation partielle empirique pour les valeurs jusqu'à $s/2$ pour en déduire les ordres p et q et les coefficients correspondants,
4. les techniques d'ajustement et de validation décrites dans le chapitre 7 sont applicables.

Remarque 10.4 *Il est souvent nécessaire d'employer une transformation de Box-Cox avant tout traitement.*

Remarque 10.5 *Lorsqu'il n'y a pas ce découplage entre les deux modèles, une méthode plus complexe peut éventuellement être envisagée :*

1. *Trouver les ordres d et D tels que le processus $Y_t = (1-B)^d(1-B^s)^D X_t$ soit stationnaire.*
2. *Faire une hypothèse sur les valeurs de p, q, P, Q.*
3. *Développer la formule du processus $ARMA$ supposé obtenu après différentiation :*

$$\Phi(B)\Phi^{(s)}(B^s)Y_t = \Theta(B)\Theta^{(s)}(B^s)Z_t$$

   devient :

$$\Phi_g(B)Y_t = \Theta_g(B)Z_t$$

   où l'indice 'g' signifie 'global'.
4. *Estimer globalement les valeurs des paramètres.*
5. *Revenir sur l'hypothèse de départ si nécessaire.*

Les coefficients de Φ, $\Phi^{(s)}$, Θ, $\Theta^{(s)}$ seront calculés à partir des coefficients de Φ_g et Θ_g obtenus.

Remarque 10.6 *En général, comme pour les $ARIMA$, la série semble stationnaire après quelques différentiations (dans la pratique, d et D valent généralement 0 ou 1). Mais on ne sait pas assurer qu'elle est de moyenne nulle. On a alors un modèle où*

$$\forall t \geqslant 0, \quad X_t = m\frac{t^{d+D}}{(d+D)!} + \widetilde{X}_t \ ,$$

avec $\widetilde{X}_t$ est un $SARIMA(p,d,q) \times (P,D,Q)_s$. On peut alors modifier l'identification comme dans le cas des $ARIMA$, en testant la nullité de la série différentiée, et en ajoutant si nécessaire un régresseur $(1:n)^{d+D}$ pour l'estimation du $SARIMA$ sur la série originale.

R: Pour estimer un $SARIMA(p,d,q)\times(P,D,Q)_s$ par maximum de vraisemblance gaussien : arima() en précisant `order=(p,d,q)`, et `seasonal=list(period=s,order=(P,D,Q))`.

10.3 Pour aller plus loin

Nous avons vu plusieurs types de modèles : $ARMA$, modèle linéaire à résidu $ARMA$, modèle à fonction de transfert, $ARIMA$ et $SARIMA$. Ils ont tous des représentations assez simples en modèles d'état, mais la liste ne s'arrête pas là : on pourrait parler de $PARMA$ (pour "periodic $ARMA$"), d'$ARMA$ vectoriel, de modèle structurel, etc. On peut également inventer un modèle d'état ad hoc pour chaque problème donné, et c'est dans cette flexibilité que réside la force des modèles d'état.

Quatrième partie

Analyse spectrale

Cette partie adopte un point de vue proche du traitement du signal, ou plutôt de l'analyse des signaux car on parlera peu de traitement. Pourquoi analyser un signal ? D'abord pour le "comprendre", c'est à dire en avoir une vision simple et efficace. Simple, cela veut dire être capable de le résumer avec le moins de mots possibles. Efficace par rapport à quel but ? Ça dépend. Par exemple, on peut vouloir le débruiter, restaurer un signal perdu, prédire les valeurs futures, le compresser pour gagner de la place, estimer des paramètres importants etc. Le but de ce chapitre est avant tout de présenter l'outil mathématique fondamental permettant de "comprendre" un peu mieux un signal, ou une série chronologique. Cet outil est la transformée de Fourier. Il est fondamental pour des raisons historiques, et aussi parce qu'il est très bien adapté à l'étude des séries stationnaires.

Le point crucial à retenir est que les signaux (ou séries chronologiques) de longueur n que l'on observe sont, d'un point de vue mathématique, des vecteurs dans un espace de grande dimension, $\mathbb{R}^n$, et que pour "comprendre" ces signaux, les apprivoiser, les examiner, il faut pouvoir "tourner autour", comme on tournerai autour d'une sculpture pour en saisir tous les aspects. Tourner autour d'une sculpture, c'est changer de repère dans l'espace $\mathbb{R}^3$. Tourner autour d'un signal, c'est aussi changer de repère (ou simplement de base), mais dans $\mathbb{R}^n$. Il se trouve qu'il existe une base particulièrement adaptée aux séries chronologiques stationnaires, c'est la *base de Fourier*. Lorsqu'on regarde un $ARMA$ dans cette base, ses coefficients sont décorrélés ! De plus, les coefficients d'un bruit $ARMA$ sont assez bien répartis en valeur absolue dans cette base, alors que les coefficients d'un signal régulier sont plutôt concentrés. Ceci a une importance fondamentale pour la détection de signaux noyés dans du bruit, ou plus généralement dans une phase purement descriptive d'une série chronologique, pour comprendre ce qu'elle contient. Enfin, comme ce qu'elle contient peut ne pas être stationnaire, on sera amené à examiner le signal à travers différentes fenêtres, ce qui donnera la transformée de Fourier à fenêtre glissante (ou transformée de Fourier à court terme).

Chapitre 11

Signaux numériques et changement de base

11.1 Quelques notions d'algèbre linéaire

11.1.1 Signaux réels ou complexes, changement de base

Mathématiquement un signal numérique de longueur n est une suite de n valeurs complexes $(x_1, \ldots, x_n)$. On dit qu'il ext complexe si ses valeurs sont complexes, réel si ses valeurs sont réelles. L'ensemble des signaux numériques complexes de longueur n correspond donc à l'espace $\mathbb{C}^n$, tandis que l'ensemble des signaux numériques réels de longueur n correspond à l'espace $\mathbb{R}^n$.

Une *base* de $\mathbb{C}^n$ est un ensemble de n éléments de $\mathbb{C}^n$ tels que que tout élément de $\mathbb{C}^n$ puisse s'écrire comme combinaison linéaire des éléments de la base. L'exemple le plus naturel est sans doute la *base canonique* de $\mathbb{C}^n$, $\mathcal{B}_{can}$:

$$\mathcal{B}_{can} = \{(1, 0, \ldots, 0)^T, (0, 1, 0, \ldots, 0)^T, \ldots, (0, \ldots, 0, 1)^T\} .$$

Autrement dit, $\mathcal{B}_{can}$ est l'ensemble des vecteurs $\overrightarrow{e_1}, \ldots, \overrightarrow{e_n}$, où :

$$\forall k \in \{1, \ldots, n\},\ \forall j \in \{1, \ldots, n\} \quad \overrightarrow{e_k}(j) = \mathbb{1}_{k=j} .$$

Lorsqu'on considère un élément $x = (x_1, \ldots, x_n)$ de $\mathbb{C}^n$, les coefficients de x dans la base canonique sont justement $x_1, \ldots, x_n$. Autrement dit :

$$\forall x \in \mathbb{C}^n, \quad x = \sum_{k=1}^{n} x_k \overrightarrow{e_k} .$$

Soient $\mathcal{B}$ et $\mathcal{C}$ deux bases de $\mathbb{C}^n$. Comment, connaissant les coordonnées d'un vecteur x dans la base $\mathcal{B}$, calculer les coordonnées de x dans la base $\mathcal{C}$? Soit x un vecteur, $\mathcal{B} = (\overrightarrow{b_1}, \ldots, \overrightarrow{b_n})$ et $\mathcal{C} = (\overrightarrow{c_1}, \ldots, \overrightarrow{c_n})$ deux bases de $\mathbb{C}^n$. On note $(\lambda_1, \ldots, \lambda_n)$ les coordonnées de x dans la base $\mathcal{B}$, et pour tout l, on note $(p_{1l}, \ldots, p_{nl})$ les coordonnées de $\overrightarrow{b_l}$ dans la base $\mathcal{C}$. On a donc :

$$x = \sum_{l=1}^{n} \lambda_l \overrightarrow{b_l} \quad \forall l \in \{1, \ldots, n\},\ \overrightarrow{b_l} = \sum_{k=1}^{n} p_{kl} \overrightarrow{c_k} .$$

Par conséquent,

$$x = \sum_{l=1}^{n} \lambda_l \sum_{k=1}^{n} p_{kl} \overrightarrow{c_k} = \sum_{k=1}^{n} \left(\sum_{l=1}^{n} p_{kl} \lambda_l \right) \overrightarrow{c_k} .$$

Ce qui signifie que la l^{e} coordonnée de x dans la base $\mathcal{C}$ est $(\sum_{l=1}^{n} p_{kl} \lambda_l)$. Notons λ le vecteur colonne des coordonnées de x dans la base $\mathcal{B}$, μ le vecteur colonne des coordonnées de x dans la base $\mathcal{C}$ et $P_{\mathcal{C}}^{\mathcal{B}}$ la matrice $(p_{kl})_{1 \leqslant k \leqslant n 1 \leqslant l \leqslant n}$. Ce que l'on vient de montrer s'écrit matriciellement :

$$\mu = P_{\mathcal{C}}^{\mathcal{B}} \lambda .$$

Remarque 11.1 *Souvent, la nouvelle base est donnée par ses coordonnées dans l'ancienne base, et on dispose donc de* $P_{\mathcal{B}}^{\mathcal{C}}$ *plutôt que* $P_{\mathcal{C}}^{\mathcal{B}}$*. La matrice* $P_{\mathcal{B}}^{\mathcal{C}}$ *s'appelle traditionnellement* matrice de changement de base de $\mathcal{B}$ à $\mathcal{C}$ *(ou* matrice de passage de $\mathcal{B}$ à $\mathcal{C}$*). On a alors :*

$$P_{\mathcal{C}}^{\mathcal{B}} = (P_{\mathcal{B}}^{\mathcal{C}})^{-1} \ .$$

11.1.2 Bases orthonormées

L'espace $\mathbb{C}^n$ est muni du produit scalaire :

$$\forall x, y \in \mathbb{C}^n, \quad \langle x, y \rangle = \sum_{k=1}^{n} \overline{x}_k y_k \ ,$$

et de la norme correspondante :

$$\forall x \in \mathbb{C}^n, \quad \|x\| = \sqrt{\sum_{k=1}^{n} |x_k|^2} \ .$$

Ces définitions ont un sens également sur $\mathbb{R}^n$. Deux vecteurs sont dits orthogonaux si leur produit scalaire est nul. On dit qu'une famille de vecteurs est orthogonale si les vecteurs qui la constituent sont 2 à 2 orthogonaux. On dit qu'un vecteur est normé si sa norme vaut 1, et on dit qu'une famille de vecteurs est est orthonormée si c'est une famille orthogonale constituée de vecteurs normés. Un exemple de base orthonormée est la base canonique.

Remarque 11.2 *Si* $\mathcal{B}$ *et* $\mathcal{C}$ *sont des bases orthonormées, on peut voir que* $P_{\mathcal{C}}^{\mathcal{B}}$ *est une matrice orthogonale, c'est à dire que* $(P_{\mathcal{C}}^{\mathcal{B}})^* P_{\mathcal{C}}^{\mathcal{B}} = I_n$*, où* A^* *désigne la matrice* transconjuguée*, c'est à dire* $\overline{A}^T$*. On a donc :*

$$P_{\mathcal{C}}^{\mathcal{B}} = (P_{\mathcal{B}}^{\mathcal{C}})^* \ .$$

Proposition 11.3 *Une famille orthonormée de* n *vecteurs de* $\mathbb{C}^n$ *(respectivement* $\mathbb{R}^n$*) est une base de* $\mathbb{C}^n$ *(respectivement* $\mathbb{R}^n$*). De plus, les coordonnées d'un vecteur* x *dans une base orthonormée* $(\overrightarrow{e}_0, \ldots, \overrightarrow{e}_{n-1})$ *sont* $(\langle \overrightarrow{e}_0, x \rangle, \ldots, \langle \overrightarrow{e}_{n-1}, x \rangle)$*.*

11.2 Effet d'un changement de base orthonormale sur un bruit blanc

Soit $W = (W_0, \ldots, W_{n-1})$ un bruit blanc faible de variance σ^2. Soit $\mathcal{B}$ une base orthonormale de $\mathbb{R}^n$ et A la matrice de changement de base de $\mathcal{B}$ à la base canonique. On a donc $AA^* = I_n$. Les coordonnées de W dans la base $\mathcal{B}$ forment le vecteur AW, qui a pour matrice de covariance $A\sigma^2 I_n A^* = \sigma^2 I_n$. Ainsi, AW est encore un bruit blanc faible, de même variance que W.

De plus, si W est un bruit blanc fort gaussien, on voit que AW est encore un bruit blanc fort gaussien, de même variance.

Chapitre 12

Bases de Fourier finies

12.1 Transformée de Fourier Discrète (TFD)

12.1.1 Rappels sur les complexes

Nous aurons besoin de la notion d'*exponentielle complexe*. Lorsque α est un nombre réel, on note $\exp(i\alpha)$ ou $e^{i\alpha}$ le nombre complexe :

$$e^{i\alpha} = \cos(\alpha) + i\sin(\alpha) \ .$$

Le point remarquable est que cette exponentielle se comporte à bien des égards comme l'exponentielle réelle, principalement à cause de l'identité suivante :

$$e^{i\alpha}e^{i\beta} = e^{i(\alpha+\beta)} \ .$$

On peut alors montrer que pour tous réels α et β,

$$(e^{i\alpha})^{\beta} = e^{i\alpha\beta} \ .$$

Par ailleurs, il est facile de voir que :

$$e^{i2\pi} = 1, \quad e^{i\pi/2} = i, \quad e^{-i\pi/2} = -i \quad \text{et} \quad e^{i\pi} = -1 \ ,$$

et également :

$$\overline{e^{i\alpha}} = e^{-i\alpha} \ .$$

De plus, si $\alpha \in \mathbb{R}$, $e^{i\alpha}$ vaut 1 si et seulement si α est de la forme $2k\pi$ avec $k \in \mathbb{Z}$. Une conséquence qui sera utilisée dans la section suivante est que pour tout $n \in \mathbb{N}^*$ et tout entier j de $\{-n+1, \ldots, n-1\}$,

$$\sum_{t=0}^{n-1} e^{\frac{2ijt\pi}{n}} = \begin{cases} n & \text{si } j = 0 \\ 0 & \text{si } j \neq 0 \end{cases} \tag{12.1}$$

Enfin, tout nombre complexe non nul z peut s'écrire de manière unique sous la forme $z = |z|e^{i\theta}$ avec $\theta \in [-\pi, \pi[$. θ s'appelle l'argument de z.

R: Arg(z) renvoie l'argument de z dans $[-\pi, \pi]$.

12.1.2 Définition de la TFD

On définit une nouvelle base de $\mathbb{C}^n$:

$$\forall j \in \{0, \ldots, n-1\}, \quad \overrightarrow{e}_j = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 \\ e^{i\omega_j} \\ e^{2i\omega_j} \\ \vdots \\ e^{(n-1)i\omega_j} \end{pmatrix} \ ,$$

où $\omega_j = \frac{2j\pi}{n}$.

Proposition 12.1 $(\overrightarrow{e}_0, \ldots, \overrightarrow{e}_{n-1})$ *est une base orthonormée de* $\mathbb{C}^n$ *qui est appelée* la base de Fourier discrète *(ou parfois base de Fourier finie). On appelle transformée de Fourier discrète (TFD) d'un vecteur* x *de* $\mathbb{C}^n$ *le vecteur* $\mathcal{F}_n x$ *de ses coordonnées dans la base de Fourier :*

$$\forall j \in \{0, \ldots, n-1\}, \quad \mathcal{F}_n x(j) = \langle \overrightarrow{e_j}, x \rangle = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} e^{-\frac{2ij\pi t}{n}} x_t \ .$$

Quand il n'y a pas de risque de confusion avec un estimateur, on notera $\widehat{x}$ *à la place de* $\mathcal{F}_n x$*. On a donc la formule d'inversion :*

$$\forall t \in \{0, \ldots, n-1\}, \quad x_t = \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} e^{+\frac{2ij\pi t}{n}} \widehat{x}(j) \ .$$

Démonstration : Il suffit de montrer que $(\overrightarrow{e}_0, \ldots, \overrightarrow{e}_{n-1})$ est une famille orthonormée, ce qui découle de l'équation (12.1). □

Une conséquence importante de la Proposition 12.1 est la formule suivante, dite *formule de Parseval* :

$$\forall x \in \mathbb{C}^n, \quad \sum_{t=0}^{n-1} |x_t|^2 = \sum_{j=0}^{n-1} |\widehat{x}_j|^2 \ .$$

Remarque 12.2 *La TFD est souvent définie sans la normalisation par* $\sqrt{n}$*.*

R: fft() donne la TFD sans la normalisation par $\sqrt{n}$, en commençant par la fréquence nulle.

fft(..., inverse=TRUE) donne la TFD inverse, sans la normalisation par $\sqrt{n}$, en commençant par la fréquence nulle.

Quand on examine la transformée de Fourier d'un signal, on examine en priorité son *périodogramme*, c'est à dire le carré du module de la transformée de Fourier. C'est plus simple et en général aussi efficace que d'observer partie réelle et partie imaginaire, par exemple. De plus, le module de la transformée de Fourier contient des informations qui ne sont, en général, pas utile au premier abord.

Définition 12.3 *Le* périodogramme *d'un signal* x *est le carré du module de sa transformée de Fourier discrète.*

Remarque 12.4 *Toute la base étant périodique de période* n*, l'espace naturel des temps (et des fréquences) est* $\mathbb{Z}/n\mathbb{Z}$*. Cela revient à périodiser les signaux et leurs transformées de Fourier en posant que* $f(k+n) = f(k)$ *pour tout* $k \in \mathbb{Z}$ *et tout signal* f*.*

Faisons une remarque importante pour la représentation du périodogramme (ou de la transformée de Fourier). La définition des $\overrightarrow{e}_j$ a un sens pour tout j dans $\mathbb{Z}$, et on a alors : $\overrightarrow{e}_{n-j} = \overrightarrow{e}_{-j}$ pour tout j. Par conséquent, la famille $(\overrightarrow{e_j},\ j = -\lfloor \frac{n-1}{2} \rfloor, \ldots, \lfloor \frac{n}{2} \rfloor)$ est également une base orthonormée, et c'est la base de Fourier réordonnée de façon à avoir $\overrightarrow{e}_0$ située au milieu (à peu près). On représentera le périodogramme dans cet ordre à cause de la symétrie naturelle que possède cette base : $\overrightarrow{e}_j$ et $\overrightarrow{e}_{-j}$ ont même fréquence réelle $\frac{j}{n}$. Les abscisses naturelles seront donc les féquences $\frac{1}{n}\lfloor \frac{n-1}{2} \rfloor, \ldots, \frac{1}{n}\lfloor \frac{n}{2} \rfloor$.

12.1.3 Propriétés de la TFD

La transformée de Fourier discrète possède quelques propriétés remarquables.

1. **Fréquence nulle.** $\widehat{x}(0) = \sqrt{n}\left(\frac{1}{n}\sum_{t=0}^{n-1} x_t\right)$.
2. **Fonctions réelles.** si $x \in \mathbb{R}^n$, alors $\widehat{x}(j) = \overline{\widehat{x}(n-j)}$ pour tout j de $\{0, \ldots, n-1\}$.
3. **Fonctions périodiques.** Si $q = \frac{n}{k}$ divise n et que $x \in \mathbb{C}^n$ est périodique de période q, alors :
$$\forall j \in \{0, \ldots, n-1\}, \quad j \notin \{0, q, 2q, \ldots, (k-1)q\} \Longrightarrow \widehat{x}(j) = 0 \ .$$
4. **Retard.** On définit B, l'opérateur retard (sur $\mathbb{Z}/n\mathbb{Z}$) comme suit :
$$\forall t \in \{0, \ldots, n-1\}, \quad (Bx)_t = x_{t-1} \ .$$
Alors,
$$\forall j \in \{0, \ldots, n-1\}, \quad \widehat{Bx}(j) = \widehat{x}(j)e^{-\frac{2ij\pi}{n}} \ ,$$
et par conséquent, pour tout polynôme Φ,
$$\forall j \in \{0, \ldots, n-1\}, \quad (\widehat{\Phi(B)x})(j) = \widehat{x}(j)\Phi(e^{-\frac{2ij\pi}{n}}) \ .$$
Un exemple est celui de la dérivée discrète. On note $\nabla = Id - B$ l'opérateur de différentation discrète :
$$\forall t \in \{0, \ldots, n-1\}, \quad (\nabla x) = x_t - x_{t-1} \ .$$
Alors :
$$\forall j \in \{0, \ldots, n-1\}, \quad \widehat{\nabla x}(j) = \widehat{x}(j)(1 - e^{-\frac{2ij\pi}{n}}) \ .$$
5. **Convolution.** On définit la convolution $x * y$ de deux vecteurs x et y de $\mathbb{C}^n$ par :
$$(x * y)_t = \frac{1}{\sqrt{n}} \sum_{s=0}^{n-1} x(s)y(t-s) \ .$$
Alors :
$$\forall j \in \{0, \ldots, n-1\}, \quad \widehat{x * y}(j) = \widehat{x}(j)\widehat{y}(j) \ .$$

Exemple 12.5 *Soit x la série donnée par $x_t = t$, pour $t = 0, \ldots, n-1$. Alors :*
$$\forall j \in \{1, \ldots, n-1\}, \quad \widehat{x}(j) = \frac{\sqrt{n}}{(e^{-\frac{2ij\pi}{n}} - 1)} \ .$$

Exemple 12.6 *Supposons que $n = kq$, avec k et q dans $\mathbb{N}^*$ et $x_t = e^{it\frac{2\pi}{q}}$, pour $t = 0, \ldots, n-1$. Alors :*
$$\widehat{x}(j) = \begin{cases} 0 & \text{si} \quad j \neq k \\ \sqrt{n} & \text{si} \quad j = k \end{cases}$$

12.2 Effet de la TFD sur les ARMA

Définition 12.7 *Soit X un processus ARMA défini par :*
$$\phi(B)X_t = \theta(B)Z_t \ ,$$
où Z est un bruit blanc de variance σ^2. La densité spectrale *de X est la fonction définie sur $[-\pi, \pi]$ par :*
$$P_X(\omega) = \sigma^2 \frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2} \ .$$

Remarque 12.8 *Cette définition n'est pas la plus répandue : en général, la densité spectrale est plutôt définie avec un facteur* $1/2\pi$ *par :*

$$P_X(\omega) = \frac{\sigma^2}{2\pi} \frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2} ,$$

mais la formule adoptée dans la Définition 12.7 simplifie un peu les écritures par la suite.

Une des justifications de l'intérêt de la transformée de Fourier discrète pour l'étude des $ARMA$ est le fait que celle-ci se comporte essentiellement comme une suite de variables aléatoires indépendantes lorsque n est grand.

Théorème 12.9 *Soit* X *un processus* $ARMA$ *inversible, piloté par un bruit blanc* Z *i.i.d de variance* σ^2*. Soit* P_X *la densité spectrale de* X*. On note, pour tout* $\omega \in [0, \pi[$*, le périodogramme* I_n *suivant :*

$$I_n(\omega) = \left| \mathcal{F}_n X \left(\lfloor \frac{n\omega}{2\pi} \rfloor \right) \right|^2 .$$

Alors, pour tous $0 < \omega_1 < \ldots < \omega_r < \pi$ *fixés, le vecteur aléatoire* $(I_n(\omega_1), \ldots, I_n(\omega_r))$ *converge en loi vers un vecteur de* r *variables aléatoires indépendantes, de lois exponentielles de moyennes* $(P_X(\omega_1), \ldots, P_X(\omega_r))$*.*

Démonstration : On donne juste une esquisse de démonstration. Tout d'abord, au lieu de considérer un $ARMA$ sur $\mathbb{Z}$, prenons X un $ARMA$ sur $\mathbb{Z}/n\mathbb{Z}$, c'est à dire un signal aléatoire périodique de période n tel que :

$$\forall t \in \mathbb{Z},\ \phi(B)X_t = \theta(B)Z_t ,$$

où Z est unbruit blanc sur $\mathbb{Z}/n\mathbb{Z}$, ou de manière équivalente, un signal sur $\mathbb{Z}$ périodique de période n tel que $(Z_0, \ldots, Z_{n-1})$ soit un bruit blanc. En utilisant les propriétés de la transformée de Fourier vis-à-vis de l'opérateur retard, on obtient :

$$\forall j,\ \widehat{X}(j) = \frac{\theta(e^{-\frac{2ij\pi}{n}})}{\phi(e^{-\frac{2ij\pi}{n}})} \widehat{Z}(j) .$$

On voit donc tout de suite que $\widehat{X}(0), \ldots, \widehat{X}(n-1)$ sont décorrélées, et que :

$$\mathsf{Var}(\widehat{X}(j)) = P_X(\frac{2j\pi}{n}) .$$

Si Z est gaussien, $\widehat{Z}$ aussi, et le résultat est alors évident. Sinon, il faut utiliser un théorème central limite du type de Lindeberg (vectoriel) pour utiliser le fait que $\widehat{Z}$ est asymptotiquement gaussien, et en déduire la même chose pour $\widehat{X}$. Comme la somme de deux carrés de gaussiennes indépendantes de variance $s/2$ est une loi exponentielle de moyenne s, $|\widehat{X}(j)|^2$ est adymptotiquement de loi exponentielle de moyenne $P_X(\frac{2j\pi}{n})$. □

Une illustration de ce phénomène est donnée à la Figure 12.1.

12.3 Phénomène de Gibbs et effets de bord

La base de Fourier étant composée de fonctions périodiques plutôt régulières (en fait d'autant plus régulières que la fréquence est basse) elle approche difficilement les fonctions présentant des irrégularités, notamment des discontinuités. Plus précisément, pour approcher correctement un signal régulier un petit nombre de coefficients d'assez basse fréquence suffira, alors que pour

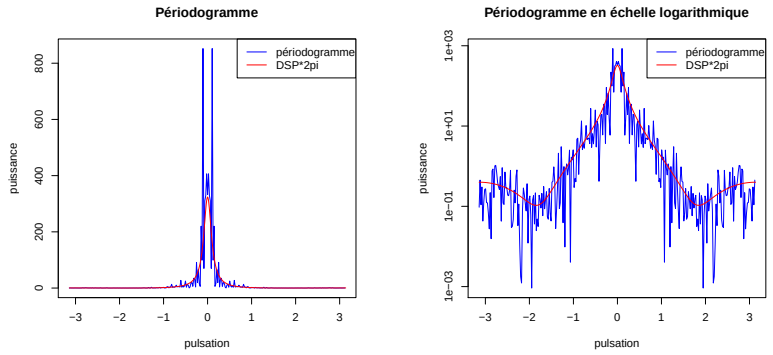

FIGURE 12.1 – Comparaison entre le périodogramme d'un ARMA(1,2) et sa densité spectrale de puissance

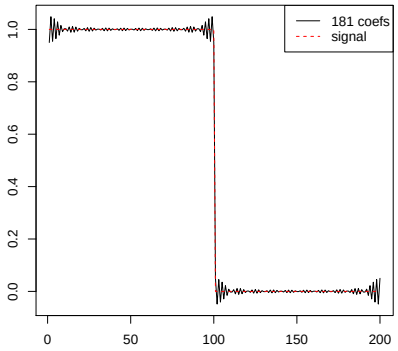

FIGURE 12.2 – Le phénomène de Gibbs

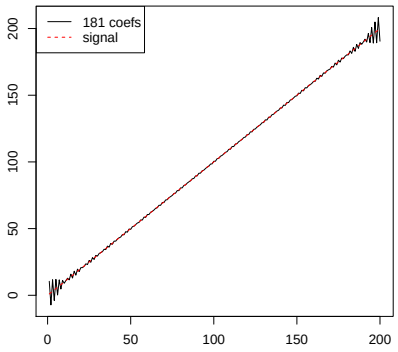

FIGURE 12.3 – Un autre phénomène de Gibbs

approcher correctement un signal irrégulier, il sera nécessaire d'aller très loin dans les hautes fréquences. Un exemple est donné par le "phénomène de Gibbs", illustré à la Figure 12.2.

Ce type de phénomène prend une saveur particulière quand on n'aperçoit pas la discontinuité. Examinons le signal de la Figure 12.3 et sa reconstruction avec 90% des coefficents.

Il semble y avoir un phénomène de Gibbs alors que le signal semble extrêmement régulier ! En fait, il faut se remémorer que l'ensemble naturel des temps est $\mathbb{Z}/n\mathbb{Z}$. Une autre façon de voir est de dire que la régularité qui importe est celle du signal périodisé $\widetilde{x}$ obtenu à partir de x par :

$$\forall t \in \{1, \ldots, n\},\ \forall n \in \mathbb{Z}, \quad \widetilde{x}_{t+n} = x_t \ .$$

Le signal périodisé obtenu à partir du signal de la Figure 12.3 présente en effet des discontinuité, cf. Figure 12.4. Le phénomène de Gibbs est dû ici à un "effet de bord".

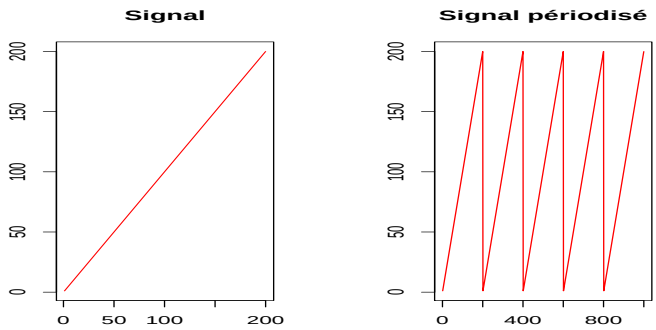

FIGURE 12.4 – Périodisation du signal de la Figure 12.3

12.4 Transformée en cosinus discret

En analyse des signaux, les discontinuités "au bord", comme à la Figure 12.3 sont plus fréquentes que les discontinuités à l'intérieur de la fenêtre de temps du type de la Figure 12.2. Une façon de contourner ce problème *pour les signaux réels* est de considérer des transformées de Fourier légèrement différentes. L'une d'elles est la *transformée en cosinus discret (DCT)*, appellation regroupant en fait huit variantes. On considèrera simplement une des plus populaires, la DCT-II.

On définit une nouvelle base de $\mathbb{R}^n$:

$$\overrightarrow{f_0}(k) = \frac{1}{\sqrt{n}}, \quad \forall k \in \{0, \ldots, n-1\} \ ,$$

$$\overrightarrow{f_j}(k) = \sqrt{\frac{2}{n}} \cos\left(\frac{j\pi}{n}\left(k + \frac{1}{2}\right)\right) \ , \quad \forall j \in \{1, \ldots, n-1\},\ \forall k \in \{0, \ldots, n-1\} \ .$$

Proposition 12.10 *$(\overrightarrow{f_0}, \ldots, \overrightarrow{f_{n-1}})$ est une base orthonormée de $\mathbb{R}^n$. On appelle transformée en cosinus discret (DCT) d'un vecteur x de $\mathbb{R}^n$ le vecteur $DCT[x]$ de ses coordonnées dans la base $(\overrightarrow{f_0}, \ldots, \overrightarrow{f_{n-1}})$:*

$$\forall j \in \{0, \ldots, n-1\}, \quad DCT[x](j) = \langle \overrightarrow{f_j}, x \rangle \ .$$

On a alors :

$$DCT[x](0) = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} x_k \ ,$$

$$DCT[x](j) = \sqrt{\frac{2}{n}} \sum_{k=0}^{n-1} x_k \cos\left(\frac{j\pi}{n}\left(k + \frac{1}{2}\right)\right), \quad \forall j \in \{1, \ldots, n-1\}$$

Pour voir en quoi cette transformation apporte quelque chose au problème des effets de bords, il est possible de l'interpréter comme la TFD d'une transformation paire du signal x, ce qui élimine la discontinuité au bord pour le signal périodisé.

La Figure 12.5 compare le spectre version TFD et le spectre version DCT pour la fonction linéaire de la Figure 12.3.

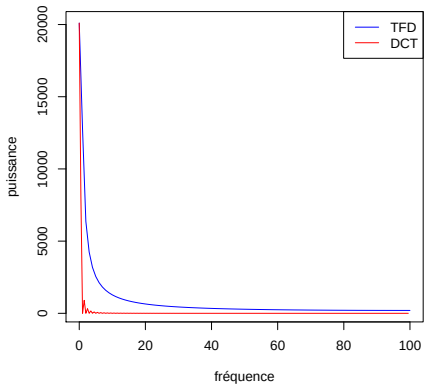

FIGURE 12.5 – Comparaison de la répartition des spectres TFD et DCT pour la fonction linéaire de la Figure 12.3

Chapitre 13

Transformée de Fourier à court terme

La transformée de Fourier est particulièrement adaptée aux signaux stationnaires. Comment peut-on utiliser les idées de la TFD lorsqu'on a affaire à un signal non stationnaire ? Pour mieux comprendre la problématique, citons un extrait de l'article de l'Encyclopedia Universalis sur les ondelettes, écrit par Grossmann et Torrésani.

> « L'analyse de Fourier nous enseigne qu'un signal quelconque peut s'écrire comme une somme de telles sinusoïdes, de fréquences et d'amplitudes variables. Un signal est entièrement caractérisé par l'ensemble des amplitudes des sinusoïdes, qui forme ce que l'on appelle sa transformée de Fourier. La transformée de Fourier est porteuse de précieuses informations sur le signal analysé (elle contient en fait toutes les informations disponibles). On sait par exemple que si elle n'a que de faibles valeurs pour des valeurs élevées de la variable de fréquence, ceci signifie que le signal varie lentement. Inversement, si elle prend des valeurs importantes pour les hautes fréquences, le signal contient une quantité non-négligeables de hautes fréquences, et donc varie rapidement, au moins dans certaines zones. Et c'est précisément là que nous touchons du doigt l'une des limitations importantes de l'analyse de Fourier usuelle. La transformée de Fourier du signal est incapable de localiser les portions du signal dans lesquelles les variations sont rapides, ni celles où elles sont lentes. »

Il y a plusieurs manières de remédier à ce problème : les plus connues sont sans doute *la transformée de Fourier à court terme* (ou transformée de Fourier à fenêtre glissante) et *les ondelettes*. Nous ne parlerons que de la transformée de Fourier à court terme. L'idée est simple : puisque l'on perd, dans la transformée de Fourier, les informations sur la localisation, nous allons localiser la transformée de Fourier, en l'appliquant sur une fenêtre glissante, de manière similaire à ce qu'on a fait pour l'analyse de la stationnarité par l'autocovariance empirique à court terme, cf. section 1.3.

13.1 Transformée de Fourier à court terme

Soit g un vecteur non nul de $\mathbb{C}^n$ qui jouera le rôle de "fenêtre". Typiquement, g aura une allure de cloche centrée autour de 0. A la fenêtre g on associe la transformée de Fourier à court terme correspondante, qui à tout vecteur $f \in \mathbb{C}^n$, associe une matrice de taille $n \times n$:

$$T_g f(k,l) = \sum_{j=0}^{n-1} f(j)\overline{g}(j-k)e^{-2i\pi\frac{jl}{n}} \ .$$

Une autre façon de l'écrire est :

$$T_g f(k,l) = \langle g^{(l,k)}, f \rangle \ ,$$

où les $g^{(l,k)}$ sont les *atomes de Gabor*, ou gaborettes :

$$\forall j \in \{0, \ldots, n-1\}, \quad g_j^{(l,k)} = g(j-l)e^{2i\pi \frac{jk}{n}} \ .$$

La Figure 13.1 représente les parties réelles de quelques gaborettes lorsque g est une fenêtre gaussienne.

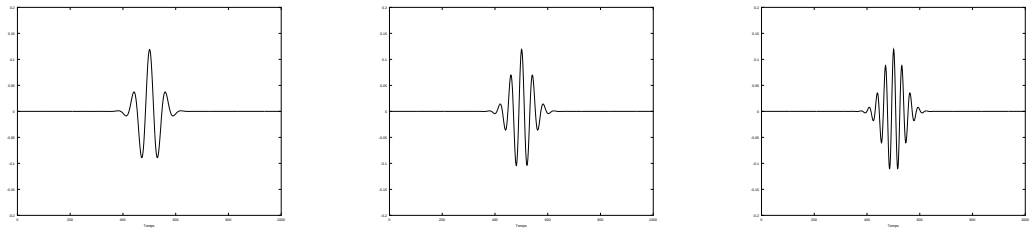

FIGURE 13.1 – Quelques parties réelles de gaborettes avec fenêtre gaussienne

Il est facile de montrer une formule de type Parseval. Pour tout vecteur f de $\mathbb{C}^n$,

$$\|f\|_2^2 = \frac{1}{n\|g\|_2^2} \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} |T_g f(k,l)|^2 \ . \tag{13.1}$$

On a également une formule de synthèse :

$$f = \frac{1}{n\|g\|_2^2} \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} T_g f(k,l) g^{(l,k)} = \frac{1}{n\|g\|_2^2} \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} \langle g^{(l,k)}, f \rangle g^{(l,k)} \ . \tag{13.2}$$

Remarquons que le nombre de gaborettes est n^2, la transformée de Fourier à court terme est donc très redondante, et il n'est donc pas question que les gaborettes forment une base de l'espace $\mathbb{C}^n$.

Remarque 13.1 *La fonction g est en général choisie très proche de 0 aux "extrêmités", c'est à dire près de $n/2$, ce qui atténue l'effet de discontinuité au bord vu dans la section 12.3.*

Par ailleurs, le choix de la "taille" de la fenêtre est bien sûr important. Si la fenêtre est "large", on aura du mal à identifier précisément l'instant où survient un évènement et on aura donc ce qu'on appelle une mauvaise résolution temporelle. Par contre, la fenêtre étant large, cela permet d'avoir le temps de voir osciller un certain nombre de fonctions de la base de Fourier, ce qui permettra de les distinguer plus précisément, on aura donc une bonne résolution fréquentielle. Au contraire, si la fenêtre est petite, on aura une bonne résolution temporelle, et une mauvaise résolution fréquentielle.

Définition 13.2 *Un* spectrogramme *d'un signal x est le carré du module d'une transformée de Fourier à court terme de x.*

Un exemple de spectrogramme (sous-échantillonné en temps) est donné à la Figure 13.2.

13.2 Transformée de Gabor

Dans les applications pratiques de traitement du signal, le fait que la transformée de Fourier à court terme contienne autant de coefficients est un problème important pour la vitesse de calcul car la longueur des signaux est souvent très importante. Par exemple, 3 secondes de

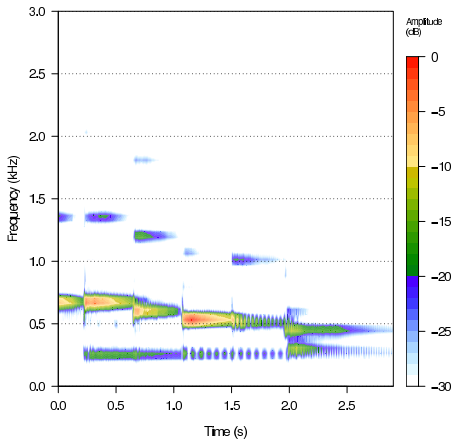

FIGURE 13.2 – Spectrogramme de quelques secondes des variations Goldberg

musique échantillonnée à 44,1 kHz donne un vecteur de taille $3 * (44,1).10^3 = 132300$. Mais puisque pour un signal de longueur n, la transformée de Fourier à fenêtre glissante renvoie n^2 coefficients, on se doute qu'il doit y avoir un moyen de réduire ce nombre de coefficients tout en gardant toute l'information sur le signal. C'est ce que cherche à accomplir la transformée de Gabor, qui correspond à un sous-échantillonage de la transformée de Fourier à court terme. La présentation qui est faite ci-dessous est tirée de [7].

Soient a et b deux entiers naturels divisant n :

$$n = aL = bK \ .$$

On définit la *transformée de Gabor discrète* (de pas d'échantillonnage en temps a et de pas d'échantillonnage en fréquence b) comme la matrice c de taille $K \times L$ donnée par :

$$c(k,l) = \langle g_{k,l}, f \rangle \ ,$$

où $g_{k,l} = g^{la,kb}$. Une question importante est de savoir inverser la transformée de Gabor discrète, si c'est possible, et de savoir si on conserve une formule de type Parseval, comme (13.1). Ceci mène naturellement à la notion de *repère*.

13.2.1 Notion de repère

Définition 13.3 *Une famille de vecteurs* $(x_0, \ldots, x_{J-1})$ *de* $\mathbb{C}^n$ *est* un repère *de* $\mathbb{C}^n$ *s'il existe des constantes* $0 < A \leqslant B$ *telles que, pour tout vecteur* f *de* $\mathbb{C}^n$*,*

$$A\|f\|^2 \leqslant \sum_{j=0}^{J-1} |\langle x_j, f \rangle|^2 \leqslant B\|f\|^2 \ .$$

Si $A = B$*, on dit que le repère est* tendu*. Si* $J > n$*, on dit que le repère est* redondant*.* A *et* B *sont appelées les bornes du repère* $(x_0, \ldots, x_{J-1})$*.*

Remarque 13.4 *Il ne faut pas confondre cette notion de repère avec celle vue au collège, qui correspond à la donnée d'un couple (origine, base) dans un espace affine.*

On peut voir facilement qu'une famille finie de vecteurs est un repère si et seulement si elle engendre $\mathbb{C}^n$. Lorsqu'un repère est redondant, il forme une famille génératrice non libre, et n'est donc pas une base.

Soit $(x_0, \ldots, x_{J-1})$ un repère de $\mathbb{C}^n$. Définissons l'*opérateur de repère* S de $\mathbb{C}^n$ dans $\mathbb{C}^n$ par :

$$(Sf)(l) = \sum_{j=0}^{J-1} \langle x_j, f \rangle x_j(l) \ .$$

La définition d'un repère assure que cet opérateur linéaire (qu'on peut voir comme une matrice de taille $n \times n$) est inversible, c'est à dire qu'il y a un moyen de retrouver f à partir de Sf. Pour calculer de manière simple cet inverse, la notion de *repère dual* est importante. On dit que $(x_0, \ldots, x_{J-1})$ est un repère dual de $(x_0, \ldots, x_{J-1})$ si pour tout f de $\mathbb{C}^n$,

$$f(l) = \sum_{j=0}^{J-1} \langle x_j, f \rangle y_j(l) = \sum_{j=0}^{J-1} \langle y_j, f \rangle x_j(l) \ .$$

On obtient ainsi une formule de synthèse semblable à (13.2)

Remarque 13.5 *Le rapport A/B est important, car il contrôle la convergence des algorithmes de synthèse : plus il est proche de 1, plus la convergence est rapide.*

Le *repère dual canonique* $(\widetilde{x}_0, \ldots, \widetilde{x}_{J-1})$ est obtenu à partir de $(x_0, \ldots, x_{J-1})$ par :

$$\forall j \in \{0, \ldots, J-1\}, \quad \widetilde{x}_j = S^{-1} x_j \ .$$

Plaçons nous maintenant dans le cas des *repères de Gabor*. On suppose donc que la famille de vecteurs $(g_{k,l},\ 0 \leqslant k \leqslant K-1,\ 0 \leqslant l \leqslant L-1)$ forme un repère. Dans ce cas, le *repère dual canonique* est encore un repère de Gabor, pour une fenêtre appelée *fenêtre duale canonique* $\widetilde{g}$ valant $\widetilde{g} = S^{-1} g$.

`Concernant l'implémentation, il existe la boîte à outil LTFAT (Linear Time Frequency Analysis Toolbox) développée pour Matlab/Octave par Peter Søendergaard, Bruno Torrésani et Peter Balazs, cf. [7].`

13.3 Pour aller plus loin

Les différentes variations exposées ci-dessus sur le thème de la transformée de Fourier ne sont pas, loin s'en faut, les seules façons d'analyser un signal. Les mathématiciens, physiciens et ingénieurs ont développé bien d'autres outils : bases de splines, bases d'ondelettes, paquets d'ondelettes etc. Chaque application nécessite de chercher le bon outil pour analyser (ou transformer) le signal.

Bibliographie

[1] R. Biscay, M. Lavielle, A. Gonzalez, I. Clark, and P. Valdés. Maximum a posteriori estimation of change points in the EEG. *Int. J. of Bio-Medical Computing*, 38 :189–196, 1995.

[2] Peter J. Brockwell and Richard A. Davis. *Introduction to time series and forecasting. 2nd ed.* Springer Texts in Statistics. New York, NY : Springer., 2002.

[3] Peter J. Brockwell and Richard A. Davis. *Time series : Theory and methods. 2nd printing of the 1991 2nd ed.* Springer Series in Statistics. New York, NY : Springer. xvi, 578 p., 2009.

[4] Xavier Guyon. Cours "séries temporelles". http://samos.univ-paris1.fr/archives/ftp/cours/guyon/NotesdeCours-ST-Cotonou.pdf, Octobre 2008.

[5] Keith W. Hipel and A. Ian McLeod. *Time series modelling of water resources and environmental systems.* Number 45 in Developments in water science. Elsevier, 1994. http://www.stats.uwo.ca/faculty/aim/1994Book/.

[6] Michel Prenat. Séries chronologiques. http://www.math.u-psud.fr/~rossigno/Enseignement/ens_files/cours_sc.pdf, 2010.

[7] Peter L. Søndergaard, Bruno Torrésani, and Peter Balazs. The Linear Time Frequency Analysis Toolbox. *International Journal of Wavelets, Multiresolution Analysis and Information Processing*, submitted, 2011.

[8] Bruno Torrésani. Mathématiques pour le traitement du signal. http://www.latp.univ-mrs.fr/~torresan/MMTS/cours09.pdf, 2009.

Index

additif, voir modèle additif
arima, 65
ARMA, 26
    modèle d'état, 55
    simulation, 32
autocorrélation
    croisée, 9
    empirique, 9
    fonction d', 22
    partielle, 41
autocovariance
    empirique, 9
    à court terme, 10
    fonction d', 22

base, 75
    canonique, 75
    changement de, 75
    de Fourier, 77
    orthonormale, 75
Box-Cox, voir transformation de Box-Cox
bruit blanc
    faible, 22
    fort, 21, 22

causalité, 29
corrélation
    empirique, 8
    fonction de, voir autocorrélation
covariance
    empirique, 8
    fonction de, 22

DCT, 82
densité spectrale, 79
désaisonnalisation, 13
différences (équation aux), 31
différentiation, 16

effet de bord, 80
exponentielle complexe, 77

filtre linéaire, 25

inversibilité, 29

Kalman (filtre), 56

lissage, 13
loess, 13

médiane
    empirique, 7
modèle
    à fonction de transfert, 62
    additif, 15
    d'état, 55

périodogramme, 78
phénomène de Gibbs, 80
phénomène saisonnier, 13
prédiction linéaire, 35

régression
    à résidu ARMA, 61
    droite de, 8

saisonnalité, 13, 61
sarima, 67
série chronologique, 22
    aléatoire, 22
    description, 16
    stationnaire, 10, 22
SLC, 22
stationnarisation, 15
stationnarité, 22
    empirique, 10

tendance, 13
tendance polynômiale, 62
TFD, 77
transformation de Box-Cox, 15
transformée de Fourier
    à court terme, 85
    discrète, 77
transformée en cosinus discret, 82