

Séries Temporelles (SIMDI226)

Céline Lévy-Leduc, Eric Moulines, François Roueff



Télécom ParisTech

Notations et conventions

Les vecteurs de \mathbb{C}^d sont identifiés aux matrices $d \times 1$.

La norme euclidienne de $x \in \mathbb{C}^d$ est notée $|x|$.

La transposée de la matrice A est notée A^T .

La transposée conjuguée de la matrice A est notée A^H .

La variance de la variable aléatoire X est notée $\mathbb{V}(X)$.

La matrice d'autocovariance du vecteur aléatoire X est notée $\text{cov}(X)$.

La matrice de covariance entre 2 vecteurs aléatoires X et Y est notée $\text{cov}(X, Y)$.

La notation $(X_t) \sim \text{BB}(0, \sigma^2)$ désigne un bruit blanc faible de variance σ^2

La notation $(X_t) \sim \text{IID}(0, \sigma^2)$ désigne un bruit blanc fort de variance σ^2

Table des matières

Partie I Processus

1	Processus aléatoires et stationnarité	3
1.1	Introduction	4
1.2	Définition et construction de la loi d'un processus aléatoire	7
1.2.1	Processus aléatoire	7
1.2.2	Répartitions finies	7
1.2.3	Processus gaussiens réels	10
1.3	Stationnarité stricte d'un processus à temps discret	13
1.3.1	Définition	13
1.3.2	Transformations préservant la stationnarité	14
	Problèmes	17
2	Eléments d'Analyse Hilbertienne	19
2.1	Définitions	20
2.2	Famille orthogonale et orthonormal	25
2.3	Séries de Fourier	29
2.4	Projection et principe d'orthogonalité	33
2.5	Isométries et isomorphismes d'espaces de Hilbert	36
3	Processus stationnaires au second ordre	39
3.1	Processus du second ordre	40
3.2	Covariance d'un processus stationnaire au second ordre	41
3.2.1	Propriétés	41
3.2.2	Interprétation de la fonction d'autocovariance	44
3.3	Mesure spectrale d'un processus stationnaire	48
3.4	Représentation harmonique des processus	53
3.4.1	Champ aléatoire non corrélé et intégrale stochastique	53
3.4.2	Intégrale stochastique par rapport à un champ	54
3.4.3	Champ spectral et représentation spectrale	55
	Problèmes	58

Partie II Filtrage et prédiction

4	Filtrage des processus stationnaires au second ordre	63
4.1	Filtrages linéaires de processus au second ordre	64
4.2	Processus ARMA	68
4.2.1	Processus $MA(q)$	68
4.2.2	Processus $AR(p)$	69
4.2.3	Cas général	73
4.3	Filtrage des processus : théorie générale	81
	Problèmes	86
5	Prédiction des processus stationnaires au second ordre	89
5.1	Prédiction linéaire de processus stationnaires	90
5.2	Algorithme de Levinson-Durbin	95
5.3	Algorithme des innovations	99
	Problèmes	102
6	Décomposition de Wold	103
6.1	Décomposition de Wold	104
6.2	Preuve du théorème 6.1.6	109
7	Modèles d'états	111
7.1	Définition et exemples	112
7.2	Modèles d'états stationnaires	113
7.3	Le filtre de Kalman	115
7.3.1	Innovation	117
7.3.2	Equations de prédiction et de filtrage	118
7.3.3	Equations de lissage	124
7.4	Le modèle linéaire Gaussien	127
7.4.1	Lissage dans le modèle linéaire Gaussien	131
7.5	La méthode à deux filtres	133
7.5.1	Paramétrisation par la matrice d'information	133
7.5.2	Le modèle linéaire gaussien	134
7.5.3	Récursion rétrograde	137

Partie III Estimation statistique

8	Introduction aux statistiques asymptotiques	141
8.1	Notions de convergence	142
8.2	Suites tendues	143
8.3	Caractérisations de la convergence en loi	145
8.4	théorème de continuité	148
8.5	Symboles o et O stochastiques	149
8.6	Delta-méthode	152
8.7	Intégrabilité uniforme et Convergence des moments	153

9 Estimation de la moyenne	155
9.1 Consistence de la moyenne empirique	156
9.2 Distribution asymptotique : processus linéaire fort	157
9.2.1 Théorèmes limites pour les processus m -dépendants ..	158
9.2.2 Théorème limite pour des processus linéaires forts	161
9.3 Extension à des processus linéaires plus généraux	163
Problèmes	165
10 Estimation de la covariance	169
10.1 Estimation des coefficients d'autocovariance et d'autocorrélation	170
10.2 Preuves des Théorèmes 10.1.4 et 10.1.3	176
Index	185
Littérature	187

Partie I

Processus

Chapitre 1

Processus aléatoires et stationnarité

Dans ce chapitre, nous introduisons des concepts de base concernant l'analyse des séries temporelles. En particulier, nous définissons les notions de stationnarité et de fonction d'autocovariance.

1.1 Introduction

Une série temporelle (ou série chronologique) est un ensemble d'observations x_t , chacune étant enregistrée à un instant t . On rencontre des séries temporelles dans des domaines très variés tels que la médecine, les télécommunications ou l'économétrie.

Dans la suite, nous proposons de considérer les observations comme des réalisations d'un processus aléatoire $(X_t)_{t \in T}$ dont nous donnons la définition dans le paragraphe 1.2. Les quelques exemples qui suivent illustrent la diversité des situations dans lesquelles la modélisation stochastique (ou aléatoire) des séries temporelles joue un rôle important.

Exemple 1.1.1 (Battements cardiaques). La figure 1.1 représente l'évolution, sur une durée totale de 900 secondes, du rythme cardiaque d'un sujet au repos. Ce rythme est mesuré en nombre de battements par minute toutes les 0.5 secondes.

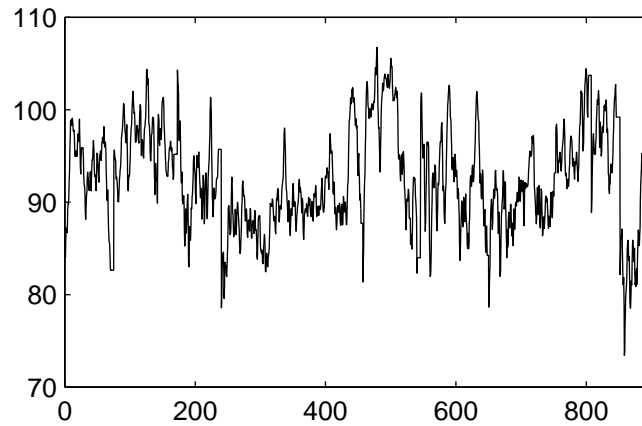


Fig. 1.1 *Battements cardiaques : évolution du nombre de battements par minute en fonction du temps mesuré en seconde.*

Exemple 1.1.2 (Trafic internet). La figure 1.2 représente les temps d'inter-arrivées de paquets TCP, mesurés en secondes, sur la passerelle du laboratoire

Lawrence Livermore. La trace représentée a été obtenue en enregistrant 2 heures de trafic. Pendant cette durée, environ 1.3 millions de paquets TCP, UDP, etc. ont été enregistrés, en utilisant la procédure *tcpdump* sur une station Sun. D'autres séries de ce type peuvent être obtenues sur *The Internet Traffic Archive*, <http://ita.ee.lbl.gov/>.

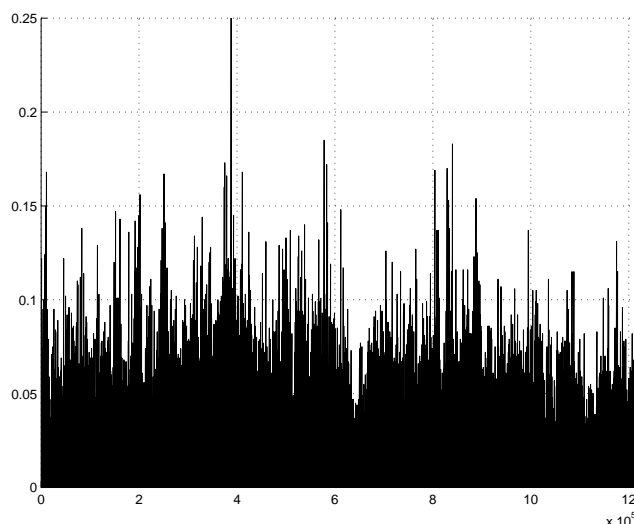


Fig. 1.2 Trace de trafic Internet : temps d'inter-arrivées de paquets TCP.

Exemple 1.1.3 (Parole). La figure 1.3 représente un segment de signal vocal échantillonné (la fréquence d'échantillonnage est de 8000 Hz). Ce segment de signal correspond à la réalisation du phonème *ch* (comme dans *chat*) qui est un son dit *fricatif*, c'est-à-dire produit par les turbulences du flot d'air au voisinage d'une constriction (ou resserrement) du conduit vocal.

Exemple 1.1.4 (Indice financier). La figure 1.4 représente les cours d'ouverture journaliers de l'indice Standard and Poor 500, du 2 Janvier 1990 au 25 Août 2000. l'indice S&P500 est calculé à partir de 500 actions choisies parmi les valeurs cotées au New York Stock Exchange (NYSE) et au NASDAQ en fonction de leur capitalisation, leur liquidité, leur représentativité dans différents secteurs d'activité. Cet indice est obtenu en pondérant le prix des actions par le nombre total d'actions, le poids de chaque valeur dans l'indice composite étant proportionnel à la capitalisation.

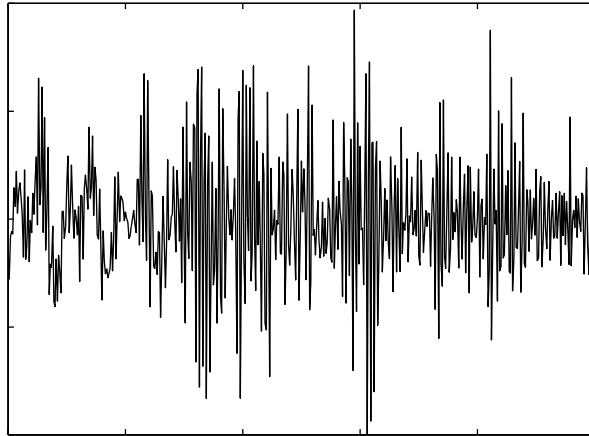


Fig. 1.3 *Signal de parole échantillonné à 8000 Hz : son non voisé ch.*

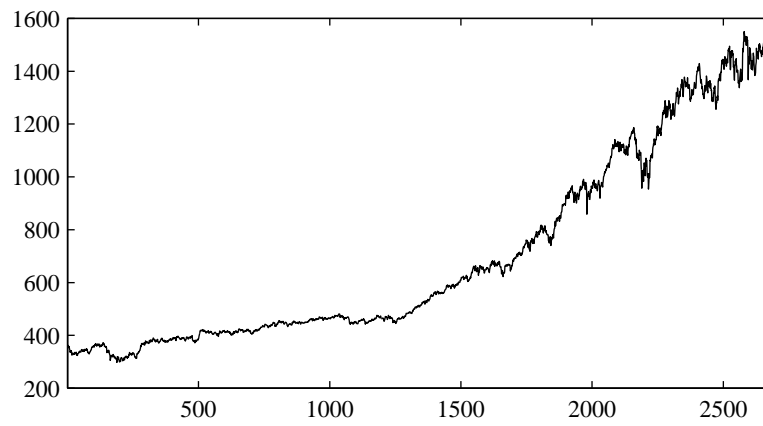


Fig. 1.4 *Cours quotidien d'ouverture de l'indice S&P500 : entre Janvier 1990 et Août 2000.*

1.2 Définition et construction de la loi d'un processus aléatoire

1.2.1 Processus aléatoire

Définition 1.2.1 (Processus aléatoire). Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, T un ensemble d'indices et (E, \mathcal{E}) un espace mesurable. On appelle processus aléatoire une famille $(X_t)_{t \in T}$ de v.a. à valeurs dans (E, \mathcal{E}) indexées par $t \in T$.

Le paramètre t représente par exemple le temps. Lorsque $T = \mathbb{Z}$ ou \mathbb{N} , nous dirons que le processus est à *temps discret* et, lorsque $T = \mathbb{R}$ ou \mathbb{R}_+ , que le processus est à *temps continu*. Dans la suite, nous nous intéresserons sauf exception aux processus à temps discret avec $T = \mathbb{Z}$. Quant à (E, \mathcal{E}) , nous considérerons le plus souvent $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (où $\mathcal{B}(\mathbb{R})$ est la tribu borélienne de \mathbb{R}) ou $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Dans le premier cas, on dira que le processus aléatoire est *scalaire*. Dans le second, nous dirons que le processus est *vectorel*.

Notons qu'un processus peut être vu comme une application $X : \Omega \times T \rightarrow E$, $(\omega, t) \mapsto X_t(\omega)$ telle que, à chaque instant $t \in T$, l'application $\omega \mapsto X_t(\omega)$ est une variable aléatoire de (E, \mathcal{E}) .

Définition 1.2.2 (Trajectoire). Pour chaque $\omega \in \Omega$, l'application $t \mapsto X_t(\omega)$ est une fonction de $T \rightarrow E$ qui s'appelle la *trajectoire* associée à l'épreuve ω .

1.2.2 Répartitions finies

Etant donnés 2 espaces mesurables (E_1, \mathcal{E}_1) et (E_2, \mathcal{E}_2) , on définit l'espace mesurable produit $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ où \times désigne le produit cartésien usuel des ensembles et \otimes l'opération correspondante sur les tribus : $\mathcal{E}_1 \otimes \mathcal{E}_2$ désigne la tribu engendrée par $\{A_1 \times A_2, A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2\}$, ce que l'on écrira

$$\mathcal{E}_1 \otimes \mathcal{E}_2 = \sigma\{A_1 \times A_2 : A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2\}.$$

Comme la classe d'ensembles $\{A_1 \times A_2 : A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2\}$ est stable par intersection finie, une probabilité sur $\mathcal{E}_1 \otimes \mathcal{E}_2$ est *caractérisée* par sa restriction à cette classe (voir [1, Corollaire 6.1]).

On définit de même un espace mesurable produit $(E_1 \times \cdots \times E_n, \mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_n)$ à partir d'un nombre fini n d'espaces mesurables (E_t, \mathcal{E}_t) , $t \in T$. Si T n'est pas de cardinal fini, cette définition se généralise en considérant la tribu engendrée par les *cylindres* sur le produit cartésien $\prod_{t \in T} E_t$ qui contient l'ensemble des familles $(x_t)_{t \in T}$ telles que $x_t \in E_t$ pour tout $t \in T$. Examinons le cas qui nous servira par la suite où $(E_t, \mathcal{E}_t) = (E, \mathcal{E})$ pour tout $t \in T$. On

note alors $E^T = \prod_{t \in T} E$ l'ensemble des trajectoires $(x_t)_{t \in T}$ telles que $x_t \in E$ pour tout t , que l'on munit de la tribu engendrée par les cylindres

$$\mathcal{E}^{\otimes T} = \sigma \left\{ \prod_{t \in I} A_t \times E^{T \setminus I} : I \in \mathcal{I}, \forall t \in I, A_t \in \mathcal{E} \right\} ,$$

où l'on note \mathcal{I} l'ensemble des parties finies de T .

Soit $X = (X_t)_{t \in T}$ un processus défini sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans (E, \mathcal{E}) et $I \in \mathcal{I}$. On note \mathbb{P}_I la loi du vecteur aléatoire $\{X_t, t \in I\}$, c'est-à-dire la mesure image de \mathbb{P} par ce vecteur : \mathbb{P}_I est la probabilité sur $(E^I, \mathcal{E}^{\otimes I})$ définie par

$$\mathbb{P}_I \left(\prod_{t \in I} A_t \right) = \mathbb{P}(X_t \in A_t, t \in I) , \quad (1.1)$$

où $A_t, t \in T$ sont des éléments quelconques de la tribu \mathcal{E} . La probabilité \mathbb{P}_I est une *probabilité fini-dimensionnelle* ou *répartition finie* du processus X .

Définition 1.2.3. On appelle *famille des répartitions finies* l'ensemble des répartitions finies $(\mathbb{P}_I, I \in \mathcal{I})$.

La spécification de la mesure \mathbb{P}_I permet de calculer la probabilité d'événements de la forme $\mathbb{P}(\cap_{t \in I} \{X_t \in A_t\})$ où $\{A_t, t \in I\}$ est une famille d'éléments de la tribu \mathcal{E} , ou de manière équivalente, de calculer l'espérance $\mathbb{E} \left[\prod_{t \in I} f_t(X_t) \right]$ où pour tout $t \in I$, f_t est une fonction borélienne positive. Soit $J \subset I$ deux parties finies ordonnées. Soit $\Pi_{I,J}$ la projection canonique de E^I sur E^J définie par

$$\Pi_{I,J}[x] = (x_t)_{t \in J} \quad \text{pour tout } x = (x_t)_{t \in I} \in E^I . \quad (1.2)$$

La projection canonique préserve uniquement les coordonnées du vecteur appartenant au sous ensemble d'indices J . Par la définition (1.1), on observe que \mathbb{P}_J est la mesure image de $\Pi_{I,J}$ définie sur l'espace de probabilité $(E^I, \mathcal{E}^{\otimes I}, \mathbb{P}_I)$:

$$\mathbb{P}_I \circ \Pi_{I,J}^{-1} = \mathbb{P}_J . \quad (1.3)$$

Cette relation formalise le résultat intuitif que la distribution fini-dimensionnelle d'un sous-ensemble $J \subset I$ se déduit de la distribution fini-dimensionnelle \mathbb{P}_I en "intégrant" par rapport aux variables X_t sur l'ensemble des t appartenant au complémentaire de J dans I . Cette propriété montre que la famille des répartitions finies d'un processus est fortement structurée. En particulier, les répartitions finies doivent, au moins, vérifier les conditions de compatibilité (1.3). Nous allons voir dans la suite que cette condition est en fait aussi *suffisante*.

Soit Π_I la projection canonique de E^T sur E^I ,

$$\Pi_I(x) = (x_t)_{t \in I} \quad \text{pour tout } x = (x_t)_{t \in T} \in E^T . \quad (1.4)$$

Si $I = \{s\}$ avec $s \in T$, on notera simplement

$$\Pi_s(x) = \Pi_{\{s\}}(x) = x_s \quad \text{pour tout } x = (x_t)_{t \in T} \in E^T. \quad (1.5)$$

Le théorème suivant montre comment on peut passer d'une famille de répartitions finies à une unique mesure de probabilité sur $(E^T, \mathcal{E}^{\otimes T})$, pourvu que la condition de compatibilité (1.3) soit satisfaite.

Théorème 1.2.4 (théorème de Kolmogorov). *Soit $(\nu_I)_{I \in \mathcal{I}}$ une famille de probabilités indexées par l'ensemble des parties finies ordonnées de T telle, que pour tout $I \in \mathcal{I}$, ν_I est une probabilité sur $(E^I, \mathcal{E}^{\otimes I})$. Supposons de plus que la famille $\{\nu_I, I \in \mathcal{I}\}$ vérifie les conditions de compatibilité (1.3) : pour tout $I, J \in \mathcal{I}$, tel que $I \subset J$, $\nu_I \circ \Pi_{I,J}^{-1} = \nu_J$. Alors, il existe une unique probabilité \mathbb{P} sur l'espace mesurable $(E^T, \mathcal{E}^{\otimes T})$ telle que, pour tout $I \in \mathcal{I}$, $\nu_I = \mathbb{P} \circ \Pi_I^{-1}$.*

Preuve. Remarquons que la classe des cylindres est une semi-algèbre au sens de [4, p. 297]. On définit \mathbb{P} sur cette classe par

$$\mathbb{P} \left(\prod_{t \in I} A_t \times E^{T \setminus I} \right) = \nu_I \left(\prod_{t \in I} A_t \right),$$

où I décrit \mathcal{I} et $A_t \in \mathcal{E}$ pour tout $t \in I$. La condition de compatibilité implique que \mathbb{P} vérifie les hypothèses de [4, Proposition 9]. Il s'en suit une extension unique à l'algèbre engendrée par les cylindres, c'est-à-dire à la plus petite classe d'ensembles de E^T stable par intersection finie et par passage au complémentaire contenant les cylindres de E^T . Par le théorème de Carathéodory, voir [4, Théorème 8], on obtient une unique extension de \mathbb{P} à la tribu $\mathcal{E}^{\otimes T}$. \square

Ceci nous permet de décrire les répartitions finies d'un processus donné à partir d'une seule probabilité sur $(E^T, \mathcal{E}^{\otimes T})$, la *loi* (ou *mesure image*) du processus, définie comme suit.

Définition 1.2.5 (Loi d'un processus). Soit $X = (X_t)_{t \in T}$ un processus défini sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans (E, \mathcal{E}) . La *mesure image* \mathbb{P}_X est l'unique probabilité définie sur $(E^T, \mathcal{E}^{\otimes T})$ par $\mathbb{P}_X \circ \Pi_I^{-1} = \mathbb{P}_I$ pour tout $I \in \mathcal{I}$, i.e.

$$\mathbb{P}_X \left(\prod_{t \in I} A_t \times E^{T \setminus I} \right) = \mathbb{P}(X_t \in A_t, t \in I)$$

pour tout $(A_t)_{t \in I} \in \mathcal{E}^I$.

L'existence et l'unicité de \mathbb{P}_X est une conséquence du théorème 1.2.4. Cette loi est donc *entièrement* déterminée par la donnée des répartitions finies.

La définition suivante permet de voir \mathbb{P}_X comme la probabilité d'une variable aléatoire à valeurs dans $(E^T, \mathcal{E}^{\otimes T})$. Cette variable aléatoire est obtenue comme la trajectoire du *processus canonique* défini comme suit.

Définition 1.2.6 (Processus canonique). Soit (E, \mathcal{E}) un espace mesurable et (E^T, \mathcal{E}^T) l'espace mesurable des trajectoires correspondants. La famille canonique sur (E^T, \mathcal{E}^T) est la famille des fonctions mesurables $(\xi_t)_{t \in T}$ définies sur (E^T, \mathcal{E}^T) à valeurs dans (E, \mathcal{E}) par $\xi_t(\omega) = \omega_t$ pour tout $\omega = (\omega_t)_{t \in T} \in E^T$.

Quand on munit (E^T, \mathcal{E}^T) de la *mesure image* \mathbb{P}_X , on appelle la famille canonique $(\xi_t)_{t \in T}$ définies sur $(E^T, \mathcal{E}^T, \mathbb{P}_X)$ le *processus canonique* associé à X .

On a supposé jusqu'à présent le processus $X = (X_t)_{t \in T}$ donné. Le théorème 1.2.4 peut aussi être utilisé pour le construire, sous la forme d'un processus canonique, comme le montre l'exemple suivant, puis le paragraphe 1.2.3 qui introduit une classe particulière de processus : la classe des processus gaussiens.

Exemple 1.2.7 (Suite de v.a. indépendantes). Soit $(\nu_t)_{t \in T}$ une suite de probabilités sur (E, \mathcal{E}) . Pour $I \in \mathcal{I}$, on pose

$$\nu_I = \bigotimes_{t \in I} \nu_t, \quad (1.6)$$

où \otimes désigne le produit tensoriel sur les probabilités (loi du vecteur à composantes indépendantes et de lois marginales données par les ν_t , $t \in I$). Il est clair que l'on définit ainsi une famille $(\nu_I)_{I \in \mathcal{I}}$ compatible, c'est-à-dire, vérifiant la condition donnée par l'équation (1.3). Donc, si $\Omega = E^T$, $X_t(\omega) = \omega_t$ et $\mathcal{F} = \sigma(X_t, t \in T)$, il existe une unique probabilité \mathbb{P} sur (Ω, \mathcal{F}) telle que $(X_t)_{t \in T}$ soit une suite de v.a. indépendantes telles que $X_t \sim \nu_t$ pour tout $t \in T$.

1.2.3 Processus gaussiens réels

Nous introduisons à présent une classe importante de processus aléatoires en modélisation stochastique : la classe des processus gaussiens. Rappelons tout d'abord la définition des variables aléatoires gaussiennes, univariées puis multivariées. Une description plus détaillée peut être trouvée dans [1, Chapitre 16].

Définition 1.2.8 (Variable aléatoire gaussienne réelle). On dit que X est une variable aléatoire réelle gaussienne si sa loi de probabilité a pour fonction caractéristique :

$$\phi_X(u) = \mathbb{E} [e^{iuX}] = \exp(i\mu u - \sigma^2 u^2 / 2)$$

où $\mu \in \mathbb{R}$ et $\sigma \in \mathbb{R}^+$.

On en déduit que $\mathbb{E}[X] = \mu$ et que $\mathbb{V}(X) = \sigma^2$. Si $\sigma \neq 0$, la loi possède une densité de probabilité qui a pour expression :

$$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \quad (1.7)$$

Si $\sigma = 0$, on a alors $X = \mu$ p.s. La définition suivante étend cette définition aux vecteurs aléatoires de dimension n .

Définition 1.2.9 (Vecteur gaussien réel). Un vecteur aléatoire réel de dimension n $[X_1, \dots, X_n]^T$ est un vecteur gaussien si toute combinaison linéaire de X_1, \dots, X_n est une variable aléatoire gaussienne réelle.

Notons μ le vecteur moyenne de $[X_1, \dots, X_n]^T$ et Γ sa matrice de covariance. Par définition d'un vecteur aléatoire gaussien, pour tout $u \in \mathbb{R}^n$, la variable aléatoire $Y = \sum_{k=1}^n u_k X_k = u^T X$ est une variable aléatoire réelle gaussienne. Par conséquent, sa loi est complètement déterminée par sa moyenne et sa variance qui ont pour expressions respectives :

$$\mathbb{E}[Y] = \sum_{k=1}^n u_k \mathbb{E}[X_k] = u^T \mu \quad \text{et} \quad \mathbb{V}(Y) = \sum_{j,k=1}^n u_j u_k \text{cov}(X_j, X_k) = u^T \Gamma u$$

On en déduit l'expression, en fonction de μ et de Γ , de la fonction caractéristique de la loi de probabilité d'un vecteur gaussien $[X(1), \dots, X(n)]^T$:

$$\phi_X(u) = \mathbb{E}[\exp(iu^T X)] = \mathbb{E}[\exp(iY)] = \exp\left(iu^T \mu - \frac{1}{2}u^T \Gamma u\right) \quad (1.8)$$

Réciproquement, si un vecteur aléatoire X de taille n a une fonction caractéristique de cette forme, on obtient immédiatement que X est un vecteur gaussien en calculant la fonction caractéristique de ses produits scalaires. Cette propriété permet d'obtenir la proposition suivante.

Proposition 1.2.10. *La loi d'un vecteur gaussien X de taille n est entièrement caractérisé par son vecteur moyenne μ et sa matrice d'autocovariance Γ . On notera*

$$X \sim \mathcal{N}_n(\mu, \Gamma).$$

Réciproquement pour tout vecteur $\mu \in \mathbb{R}^n$ et toute matrice symétrique positive Γ , il existe un vecteur aléatoire X tel que $X \sim \mathcal{N}_n(\mu, \Gamma)$.

Preuve. La première partie de l'énoncé découle directement de (1.8). Démontrons maintenant la réciproque. Tout d'abord le résultat est vrai pour $n = 1$ comme nous l'avons rappelé plus haut. On passe aisément au cas où Γ est diagonale. En effet, notons σ_i^2 , $i = 1, \dots, n$ ses éléments diagonaux et $\mu = [\mu_1, \dots, \mu_n]^T$. Alors il suffit de prendre X_1, \dots, X_n indépendants tels que $X_i \sim \mathcal{N}_n(\mu_i, \sigma_i^2)$ pour $i = 1, \dots, n$. On vérifie aisément que $X \sim \mathcal{N}_n(\mu, \Gamma)$ en calculant sa fonction caractéristique. Pour passer du cas des matrices diagonales à une

matrice Γ symétrique positive quelconque, on utilise le lemme suivant dont la preuve est laissée à titre d'exercice.

Lemme 1.2.11. *Soit $X \sim \mathcal{N}_n(\mu, \Gamma)$ avec $\mu \in \mathbb{R}^n$ et Γ matrice symétrique positive $n \times n$. Alors pour toute matrice A de taille $p \times n$, on a $AX \sim \mathcal{N}_n(A\mu, A\Gamma A^T)$.*

Pour conclure la preuve de la proposition 1.2.10, il suffit de remarquer que toute matrice symétrique positive Γ est diagonalisable en base orthonormée et s'écrit donc $\Gamma = U\Sigma U^T$ avec Σ matrice diagonale positive et U matrice orthogonale. Il suffit alors de prendre $Y \sim \mathcal{N}_n(U^T\mu, \Sigma)$ et de poser $X = UY$ et le lemme donne $X \sim \mathcal{N}_n(\mu, \Gamma)$ comme recherché. \square

On montre facilement la proposition suivante (voir [1, Corollaire 16.1]).

Proposition 1.2.12. *Soit $X \sim \mathcal{N}_n(\mu, \Gamma)$ avec $\mu \in \mathbb{R}^n$ et Γ matrice symétrique positive $n \times n$. Alors X a des composantes indépendantes si et seulement si Γ est une matrice diagonale.*

En utilisant le même procédé de preuve que pour la proposition 1.2.10, i.e. en considérant le cas Γ diagonale puis la diagonalisation de Γ pour passer au cas général, on obtient aussi le résultat suivant (voir [1, Corollaire 16.2]).

Proposition 1.2.13. *Soit $X \sim \mathcal{N}_n(\mu, \Gamma)$ avec $\mu \in \mathbb{R}^n$ et Γ matrice symétrique positive $n \times n$. Si Γ est de rang plein, alors la loi de probabilité de X possède une densité dans \mathbb{R}^n dont l'expression est :*

$$p_X(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(\Gamma)}} \exp\left(-\frac{1}{2}(x - \mu)^T \Gamma^{-1}(x - \mu)\right), \quad x \in \mathbb{R}^n.$$

Dans le cas où Γ est de rang $r < n$, c'est à dire où Γ possède $n - r$ valeurs propres nulles, X se trouve, avec probabilité 1, dans un sous espace affine de dimension r de \mathbb{R}^n . En effet, il existe alors $r - n$ vecteurs a_i formant une famille libre tels que $\text{cov}(a_i^T X) = 0$ et donc $a_i^T X = a_i^T \mu$ p.s. X n'admet donc évidemment pas de densité dans ce cas.

Nous étendons maintenant la notion de vecteur gaussien à celle de *processus gaussien*.

Définition 1.2.14 (Processus gaussien réel). On dit qu'un processus réel $X = (X_t)_{t \in T}$ est gaussien si, pour tout ensemble fini d'indices $I = \{t_1, t_2, \dots, t_n\}$, $[X_{t_1}, X_{t_2}, \dots, X_{t_n}]^T$ est un vecteur gaussien.

Ainsi un vecteur gaussien $[X_1, \dots, X_n]^T$ peut être lui-même vu comme un processus gaussien $\{X_t, t \in \{1, \dots, n\}\}$. Cette définition n'a donc un intérêt que dans le cas où T est de cardinal infini. D'après (1.8), la famille des répartitions finies est caractérisée par la donnée de la fonction moyenne $\mu : t \in T \mapsto \mu(t) \in \mathbb{R}$ et de la fonction de covariance $\gamma : (t, s) \in (T \times T) \mapsto \gamma(t, s) \in \mathbb{R}$. De plus, pour tout ensemble fini d'indices $I = \{t_1, t_2, \dots, t_n\}$, la matrice

Γ_I d'éléments $\Gamma_I(m, k) = \gamma(t_m, t_k)$, où $1 \leq m, k \leq n$, est une matrice de covariance d'un vecteur aléatoire de dimension n . Elle est donc symétrique positive. Réciproquement, donnons nous une fonction $\mu : t \in T \mapsto m(t) \in \mathbb{R}$ et une fonction $\gamma : (t, s) \in (T \times T) \mapsto \gamma(t, s) \in \mathbb{R}$ telle que, pour tout ensemble fini d'indices I , la matrice Γ_I est symétrique positive. On peut alors définir, pour tout ensemble fini d'indices $I = \{t_1, t_2, \dots, t_n\}$, une probabilité gaussienne ν_I sur \mathbb{R}^n par :

$$\nu_I \stackrel{\text{def}}{=} \mathcal{N}_n(\mu_I, \Gamma_I) \quad (1.9)$$

où $\mu_I = [\mu(t_1), \dots, \mu(t_n)]^T$. La famille $(\nu_I, I \in \mathcal{I})$, ainsi définie, vérifie les conditions de compatibilité et l'on a ainsi établi, d'après le théorème 1.2.4, le résultat suivant :

Théorème 1.2.15. *Soit T un ensemble d'indices quelconque, μ une fonction réelle définie sur T et γ une fonction réelle définie sur $T \times T$ dont toutes les restrictions Γ_I aux ensembles $I \times I$ avec $I \subseteq T$ fini forment des matrices symétriques positives. Il existe un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et un processus aléatoire $\{X_t, t \in T\}$ gaussien défini sur cet espace vérifiant*

$$\mu(t) = \mathbb{E}[X_t] \quad \text{et} \quad \gamma(s, t) = \mathbb{E}[(X_s - \mu(s))(X_t - \mu(t))] .$$

1.3 Stationnarité stricte d'un processus à temps discret

1.3.1 Définition

La notion de stationnarité joue un rôle central dans la théorie des processus aléatoires. On distingue ci-dessous deux versions de cette propriété, la *stationnarité stricte* qui fait référence à l'invariance des répartitions finies par translation de l'origine des temps, et une notion plus faible, la *stationnarité au second ordre*, qui impose l'invariance par translation des moments d'ordre un et deux uniquement, lorsque ceux-ci existent.

Définition 1.3.1 (Opérateurs de décalage et de retard). On suppose $T = \mathbb{Z}$ ou $T = \mathbb{N}$. On note S et l'on appelle *opérateur de décalage* (*Shift*) l'application $E^T \rightarrow E^T$ définie par

$$S(x) = (x_{t+1})_{t \in T} \quad \text{pour tout} \quad x = (x_t)_{t \in T} \in E^T .$$

Pour tout $\tau \in T$, on définit S^τ par

$$S^\tau(x) = (x_{t+\tau})_{t \in T} \quad \text{pour tout} \quad x = (x_t)_{t \in T} \in E^T .$$

Définition 1.3.2 (Stationnarité stricte). On pose $T = \mathbb{Z}$ ou $T = \mathbb{N}$. Un processus aléatoire $\{X_t, t \in T\}$ est stationnaire au sens strict si X et $S \circ X$ ont même loi, *i.e.* $\mathbb{P}_{S \circ X} = \mathbb{P}_X$.

Par caractérisation de la loi image par les répartitions finies, on a $\mathbb{P}_{S \circ X} = \mathbb{P}_X$ si et seulement si

$$\mathbb{P}_{S \circ X} \circ \Pi_I^{-1} = \mathbb{P}_X \circ \Pi_I^{-1}$$

pour toute partie finie $I \in \mathcal{I}$. Or $\mathbb{P}_{S \circ X} \circ \Pi_I^{-1} = \mathbb{P}_X \circ (\Pi_I \circ S)^{-1}$ et $\Pi_I \circ S = \Pi_{I+1}$, où $I+1 = \{t+1, t \in I\}$. On en conclut que $\{X_t, t \in T\}$ est *stationnaire au sens strict* si et seulement si, pour toute partie finie $I \in \mathcal{I}$,

$$\mathbb{P}_I = \mathbb{P}_{I+1}.$$

On remarque aussi que la stationnarité au sens strict implique que X et $S^\tau \circ X$ ont même loi pour tout $\tau \in T$ et donc aussi $\mathbb{P}_I = \mathbb{P}_{I+\tau}$, où $I+\tau = \{t+\tau, t \in I\}$.

Exemple 1.3.3 (Processus i.i.d.). Soit $(Z_t)_{t \in T}$ une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d) à valeurs dans \mathbb{R}^d . Alors $(Z_t)_{t \in T}$ est un processus stationnaire au sens strict, car, pour toute partie finie ordonnée $I = \{t_1, < t_2 < \dots < t_n\}$ et tous boréliens A_1, \dots, A_n de \mathbb{R}^d , nous avons :

$$\mathbb{P}(Z_{t_1} \in A_1, \dots, Z_{t_n} \in A_n) = \prod_{j=1}^n \mathbb{P}(Z_0 \in A_j),$$

qui ne dépend pas de t_1, \dots, t_n . Notons que d'après l'exemple 1.2.7, pour toute probabilité ν sur \mathbb{R}^d , on sait construire un processus (Z_t) i.i.d. de *loi marginale* ν , c'est-à-dire tel que $Z_t \sim \nu$ pour tout $t \in T$.

1.3.2 Transformations préservant la stationnarité

On pose $T = \mathbb{Z}$, $E = \mathbb{C}^d$ et $\mathcal{E} = \mathcal{B}(\mathbb{C}^d)$ pour un entier $d \geq 1$. Commençons par un exemple simple.

Exemple 1.3.4 (Transformation d'un processus i.i.d.). Soit Z un processus i.i.d. (voir exemple 1.3.3). Soient k un entier et g une fonction borélienne de \mathbb{R}^k dans \mathbb{R} . On peut vérifier que le processus aléatoire $(X_t)_{t \in \mathbb{Z}}$ défini par

$$X_t = g(Z(t), Z(t-1), \dots, Z(t-k+1))$$

est encore un processus aléatoire stationnaire au sens strict. Par contre, ce processus obtenu par transformation n'est plus i.i.d dans la mesure où, dès que $k \geq 1$, $X_t, X_{t+1}, \dots, X_{t+k-1}$ ont bien la même distribution marginale

mais sont, en général, dépendants car fonctions de variables aléatoires communes. Un tel processus est dit k -dépendant dans la mesure où, par contre, $\tau \geq k$ implique que $(X_s)_{s \leq t}$ et $(X_s)_{s > t+\tau}$ sont indépendants pour tout t . Les processus m -dépendants peuvent être utilisés pour approcher une grande classe de processus dépendants afin d'étudier le comportement asymptotique de statistiques usuelles telles que la moyenne empirique. Nous y reviendrons au chapitre 9.

On remarque que dans cet exemple, pour déduire la stationnarité de X , il n'est pas nécessaire d'utiliser que Z est i.i.d. mais seulement qu'il est stationnaire. En fait, pour vérifier la stationnarité, il est souvent pratique de raisonner directement sur les lois des trajectoires en utilisant la notion de filtrage.

Définition 1.3.5. Soit ϕ une application mesurable de $(E^T, \mathcal{E}^{\otimes T})$ dans $(F^T, \mathcal{F}^{\otimes T})$ et $X = (X_t)_{t \in T}$ un processus à valeurs dans (E, \mathcal{E}) . On appelle *filtré* du processus X par la transformation ϕ le processus $Y = (Y_t)_{t \in T}$ à valeurs dans (F, \mathcal{F}) défini par $Y = \phi \circ X$, c'est-à-dire $Y_t = \Pi_t(\phi(X))$ pour tout $t \in T$, où Π_t est défini par (1.5). Si ϕ est une application linéaire, on parlera de *filtrage linéaire*.

L'exemple 1.3.4 est un exemple de filtrage (en général non-linéaire, à moins que g soit une forme linéaire). La transformation associée à cet exemple est l'application $\phi : \mathbb{R}^{\mathbb{Z}} \rightarrow \mathbb{R}^{\mathbb{Z}}$ définie par

$$\phi((x_t)_{t \in \mathbb{Z}}) = (g(x_t, x_{t-1}, \dots, x_{t-k+1}))_{t \in \mathbb{Z}}.$$

Exemple 1.3.6 (Décalage). Un exemple fondamental de filtrage linéaire de processus est obtenu en prenant $\phi = S$ où S est l'opérateur de décalage de la définition 1.3.1. Dans ce cas $Y_t = X_{t+1}$ pour tout $t \in \mathbb{Z}$.

Exemple 1.3.7 (Filtre à réponse impulsionnelle finie (RIF)). Soient $n \geq 1$ et $t_1 < \dots < t_n$ des éléments de \mathbb{Z} et $\alpha_1, \dots, \alpha_n \in E$. Alors $\sum_i \alpha_i S^{-t_i}$ définit un filtrage linéaire pour n'importe quel processus $X = (X_t)_{t \in \mathbb{Z}}$ pour lequel la sortie est donnée par

$$Y_t = \sum_{i=1}^n \alpha_i X_{t-t_i}, \quad t \in \mathbb{Z}.$$

Exemple 1.3.8 (Différentiation). Un cas particulier de l'exemple précédent est donné par l'opérateur de différenciation $I - S^{-1}$ où I dénote l'opérateur identité. Le processus obtenu en sortie s'écrit

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

On pourra itérer l'opérateur de différenciation, ainsi $Y = (I - S^{-1})^k X$ est donnée par

$$Y_t = \sum_{j=0}^k \binom{k}{j} (-1)^j X_{t-j}, \quad t \in \mathbb{Z}.$$

Exemple 1.3.9 (Retournement du temps). Etant donné un processus $X = \{X_t, t \in \mathbb{Z}\}$, on appellera *processus retourné* le processus obtenu par *retournement du temps* défini par

$$Y_t = X_{-t}, \quad t \in \mathbb{Z}.$$

Exemple 1.3.10 (Intégration). Etant donné un processus $X = (X_t)_{t \in \mathbb{Z}}$ qui vérifie $\sum_{t=-\infty}^0 |X_t| < \infty$ p.s., on appellera *processus intégré* le processus défini par

$$Y_t = \sum_{s=0}^{\infty} X_{t-s}, \quad t \in \mathbb{Z}.$$

Contrairement aux exemples précédents, l'application ϕ qui définit ce filtrage doit être définie avec quelques précautions. Il faut en effet tout d'abord définir ϕ sur

$$A = \left\{ x = (x_t)_{t \in \mathbb{Z}} \in E^{\mathbb{Z}} : \sum_{t=-\infty}^0 |x_t| < \infty \right\},$$

par $\phi(x) = \sum_{s=0}^{\infty} x_{t-s}$. Comme A est un espace vectoriel, on peut prolonger ϕ linéairement sur $(E^{\mathbb{Z}}, \mathcal{E}^{\otimes \mathbb{Z}})$. Le point important est que ce filtrage ne sera appliqué à X que sous l'hypothèse $\sum_{t=-\infty}^0 |X_t| < \infty$ p.s. et que ce prolongement est donc défini de façon *quelconque*.

On remarque que dans tous les exemples précédents les opérateurs introduits préservent la stationnarité stricte, c'est-à-dire, si X est strictement stationnaire alors Y l'est aussi. Il est facile de construire des filtrages linéaires qui ne préservent pas la stationnarité stricte, par exemple, $y = \phi(x)$ avec $y_t = x_t$ pour t pair et $y_t = x_t + 1$ pour t impaire. Une propriété plus forte que la conservation de la stationnarité est donnée par la définition suivante.

Définition 1.3.11. Un filtrage linéaire est *invariant par translation* s'il commute avec S : $\phi \circ S = S \circ \phi$.

Cette propriété implique la préservation de la stationnarité mais ne lui est pas équivalente. Le retournement du temps est en effet un exemple de filtrage qui ne commute pas avec S puisque dans ce cas on a $\phi \circ S = S^{-1} \circ \phi$. En revanche tous les autres exemples ci-dessus satisfont la propriété d'invariance par translation.

Remarque 1.3.12. Un filtrage ϕ *invariant par translation* est entièrement déterminé par sa composition avec sa composition avec la projection canonique Π_0 , voir (1.5). En effet, notons $\phi_0 = \Pi_0 \circ \phi$. Alors pour tout $s \in \mathbb{Z}$, $\Pi_s \circ \phi = \Pi_0 \circ S^s \circ \phi = \Pi_0 \circ \phi \circ S^s$. Il suffit enfin d'observer que pour tout $x \in E^{\mathbb{Z}}$, $\phi(x)$ est la suite $(\pi_s \circ \phi)_{s \in \mathbb{Z}}$.

Problèmes

1.3.1 (Stationnarité de processus). Soit $(Z_t)_{t \in \mathbb{Z}}$ est un processus i.i.d. à valeurs réelles. Trouver les processus stationnaires parmi les processus $(X_t)_{t \in \mathbb{Z}}$ et $(Y_t)_{t \in \mathbb{Z}}$ suivants :

- a) $X_t = Z_1 + \cdots + Z_t$.
- b) $X_t = Z_t + \theta Z_{t-1}$, où $\theta \in \mathbb{R}$ est une constante.
- c) $X_t = Z_t Z_{t-1}$.
- d) $Y_t = (-1)^t Z_t$ et $X_t = Y_t + Z_t$. (Distinguer différentes hypothèses possibles sur la loi de Z_0).
- e) $X_t = \sum_{k \geq 0} \theta^k Z_{t-k}$ pour $|\theta| < 1$ en supposant que $\mathbb{E}[|Z_1|] < \infty$.

1.3.2 (Processus harmonique réel). Soit $(X_t)_{t \in \mathbb{Z}}$ défini par $X_t = A \cos(\theta t) + B \sin(\theta t)$ où A et B sont des variables aléatoires indépendantes de loi $\mathcal{N}_1(0, 1)$ et $\theta \in \mathbb{R}$ est une constante. Le processus $(X_t)_{t \in \mathbb{Z}}$ est-il stationnaire? Cette propriété est-elle conservée si A et B sont centrées à valeurs dans $\{-1, 1\}$?

1.3.3 (Invariance par translation). Montrer que les filtrages des exemples 1.3.6, 1.3.7 et 1.3.10 sont invariants par translation. Qu'en est-il du filtrage défini dans l'exemple 1.3.9? Ce filtrage préserve-t-il la stationnarité?

1.3.4 (Marche aléatoire). Soit $(Z_t)_{t \in \mathbb{Z}}$ est un processus i.i.d. à valeurs réelles.

- a) Construire un processus $(X_t)_{t \in \mathbb{Z}}$ tel que $Z_t = X_t - X_{t-1}$ pour tout $t \in \mathbb{Z}$.
- b) Plusieurs solutions sont-elles possible?
- c) En existe-t-il une qui soit stationnaire?

Chapitre 2

Eléments d'Analyse Hilbertienne

Les séries chronologiques exploitent beaucoup de notions d'analyse Hilbertienne. Ce chapitre n'est pas un cours d'analyse Hilbertienne, mais rassemble les notions essentielles que nous aurons à manipuler dans ce livre. La plupart des résultats sont élémentaires et sont démontrés.

2.1 Définitions

Définition 2.1.1 (Espace pré-hilbertien). Soit \mathcal{H} un espace vectoriel sur l'ensemble des nombres complexes \mathbb{C} . L'espace \mathcal{H} est appelé *pré-hilbertien* si \mathcal{H} est muni d'un produit scalaire :

$$\langle \cdot, \cdot \rangle : x, y \in \mathcal{H} \times \mathcal{H} \mapsto \langle x, y \rangle \in \mathbb{C}$$

qui vérifie les propriétés suivantes :

- a) pour tout $(x, y) \in \mathcal{H} \times \mathcal{H}$, $\langle x, y \rangle = \overline{\langle y, x \rangle}$
- b) pour tout $(x, y) \in \mathcal{H} \times \mathcal{H}$ et tout $(\alpha, \beta) \in \mathbb{C} \times \mathbb{C}$, $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$
- c) pour tout $x \in \mathcal{H}$, $\langle x, x \rangle \geq 0$, et $\langle x, x \rangle = 0$ si et seulement si $x = 0$.

L'application :

$$\| \cdot \| : x \in \mathcal{H} \mapsto \sqrt{\langle x, x \rangle} \geq 0$$

définit alors une norme sur \mathcal{H} .

Exemple 2.1.2 (Espace \mathbb{C}^n). L'ensemble des vecteurs colonnes $x = [x_1 \cdots x_n]^T$, où $x_k \in \mathbb{C}$, est un espace vectoriel dans lequel la relation :

$$\langle x, y \rangle = y^H x = \sum_{k=1}^n x_k \overline{y_k}$$

définit un produit scalaire.

Exemple 2.1.3 (Espace ℓ^2). L'ensemble des suites numériques complexes $\{x_k\}_{k \in \mathbb{N}}$ vérifiant $\sum_{k=0}^{\infty} |x_k|^2 < \infty$ est un espace vectoriel sur \mathbb{C} . On définit pour tout x et y de cet espace :

$$\langle x, y \rangle = \sum_{k=0}^{\infty} x_k \overline{y_k}.$$

Cette somme est bien définie puisque $|x_k \overline{y_k}| \leq (|x_k|^2 + |y_k|^2)/2$. De plus, on vérifie aisément les propriétés (i-iii) de la définition 2.1.1. L'espace ainsi défini est donc un espace pré-Hilbertien, que l'on note ℓ^2 .

Exemple 2.1.4 (Fonctions de carré intégrable). L'ensemble $\mathcal{L}^2(T)$ des fonctions boréliennes définies sur un intervalle T de \mathbb{R} , à valeurs complexes et de module de carré intégrable par rapport à la mesure de Lebesgue

$(\int_T |f(t)|^2 dt < \infty)$ est un espace vectoriel. Considérons alors le produit intérieur :

$$(f, g) \in \mathcal{L}^2(T) \times \mathcal{L}^2(T) \mapsto \langle f, g \rangle = \int_T f(t) \overline{g(t)} dt$$

Cette intégrale est bien définie puisque $|f(t)\overline{g(t)}| \leq (|f(t)|^2 + |g(t)|^2)/2$ et l'on montre aisément que les propriétés (i) et (ii) de la définition 2.1.1. Par contre la propriété (iii) n'est pas vérifiée puisque :

$$\langle f, f \rangle = 0 \not\Rightarrow \forall t \in T \ f(t) = 0$$

En effet une fonction f qui est nulle sauf sur un ensemble de mesure nulle pour la mesure de Lebesgue, vérifie $\langle f, f \rangle = 0$. L'espace \mathcal{H} muni du produit (f, g) n'est donc pas un espace pré-Hilbertien. C'est pourquoi on définit l'ensemble $L^2(T)$ des classes d'équivalence de $\mathcal{L}^2(T)$ pour la *relation d'équivalence* définie par l'égalité presque partout entre deux fonctions. Par construction, $L^2(T)$ est alors un espace pré-Hilbertien.

Exemple 2.1.5 (Variables aléatoires de variance finie). De façon similaire à l'exemple 2.1.4, pour tout espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, on définit $\mathcal{H} = \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ (noté $\mathcal{L}^2(\Omega)$ s'il n'y a pas de confusion possible) comme l'ensemble des v.a. X définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs complexes telles que

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

Sur cet ensemble, on définit

$$(X, Y) \in \mathcal{L}^2(\Omega) \times \mathcal{L}^2(\Omega) \mapsto \langle X, Y \rangle = \mathbb{E}[X\overline{Y}] .$$

Pour les mêmes raisons que dans l'exemple 2.1.4, on définit l'espace pré-Hilbertien $L^2(\Omega, \mathcal{F}, \mathbb{P})$ (ou $L^2(\Omega)$) comme l'ensemble des classes d'équivalences de $\mathcal{L}^2(\Omega)$ pour la relation d'équivalence définie par l'égalité presque sûre entre deux v.a. Cet exemple et l'exemple 2.1.4 se généralisent en fait à tout espace mesuré $(\Omega, \mathcal{F}, \mu)$ en posant

$$(f, g) \in \mathcal{L}^2(\Omega, \mathcal{F}, \mu) \times \mathcal{L}^2(\Omega, \mathcal{F}, \mu) \mapsto \langle f, g \rangle = \int f \overline{g} d\mu .$$

On montre aisément les propriétés suivantes :

Théorème 2.1.6. *Pour tout $x, y \in \mathcal{H} \times \mathcal{H}$, nous avons :*

- a) *Inégalité de Cauchy-Schwarz :* $|\langle x, y \rangle| \leq \|x\| \|y\|$,
- b) *Inégalité triangulaire :* $|\|x\| - \|y\|| \leq \|x - y\| \leq \|x\| + \|y\|$,
- c) *Identité du parallélogramme :*

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$$

Définition 2.1.7 (Convergence forte dans \mathcal{H}). Soit (x_n) une suite de vecteurs et x un vecteur d'un espace préhilbertien \mathcal{H} . On dit que (x_n) tend fortement vers x si et seulement si $\|x_n - x\| \rightarrow 0$ quand $n \rightarrow +\infty$. On note $x_n \rightarrow x$.

Proposition 2.1.8. *Si dans un espace de Hilbert la suite $x_n \rightarrow x$, alors (x_n) est bornée.*

Preuve. D'après l'inégalité triangulaire, on a :

$$\|x_n\| = \|(x_n - x) + x\| \leq \|x_n - x\| + \|x\|$$

□

Définition 2.1.9 (Convergence faible dans \mathcal{H}). Soit (x_n) une suite de vecteurs et x un vecteur d'un espace préhilbertien \mathcal{H} . On dit que (x_n) tend faiblement vers x si et seulement si, pour tout $y \in \mathcal{H}$, $\langle x_n, y \rangle \rightarrow \langle x, y \rangle$ quand $n \rightarrow \infty$. On note $x_n \rightharpoonup x$.

Théorème 2.1.10. *Une suite (x_n) fortement convergente converge aussi faiblement vers la même limite.*

Preuve. Supposons que (x_n) converge fortement vers x , $\lim_{n \rightarrow \infty} \|x_n - x\| = 0$. Alors, pour tout $y \in \mathcal{H}$,

$$|\langle x_n - x, y \rangle| \leq \|x_n - x\| \|y\| \rightarrow 0, \quad \text{quand } n \rightarrow \infty.$$

□

En général, la convergence faible n'entraîne pas la convergence forte. Pour tout $y \in \mathcal{H}$, l'application $\langle \cdot, y \rangle : \mathcal{H} \rightarrow \mathbb{C}$, $x \mapsto \langle x, y \rangle$ est une forme linéaire. Le théorème 2.1.10 montre que cette forme linéaire est continue.

Théorème 2.1.11 (Continuité du produit scalaire). *Soit $x_n \rightarrow x$ et $y_n \rightarrow y$ deux suites convergentes de vecteurs d'un espace pré-hilbertien \mathcal{H} . Alors quand $n \rightarrow +\infty$: $\langle x_n, y_n \rangle \rightarrow \langle x, y \rangle$. En particulier, si $x_n \rightarrow x$, $\|x_n\| \rightarrow \|x\|$.*

Preuve. D'après l'inégalité triangulaire puis l'inégalité de Cauchy-Schwarz, nous avons :

$$\begin{aligned} \langle x, y \rangle - \langle x_n, y_n \rangle &= \langle (x - x_n) + x_n, (y - y_n) + y_n \rangle - \langle x_n, y_n \rangle \\ &= \langle x - x_n, y - y_n \rangle + \langle x - x_n, y_n \rangle + \langle x_n, y - y_n \rangle \\ &\leq \|x_n - x\| \|y_n - y\| + \|x_n - x\| \|y_n\| + \|y_n - x\| \|x_n\| \end{aligned}$$

On conclut en utilisant la proposition 2.1.8 qui montre que les suites (x_n) et (y_n) sont bornées. □

Définition 2.1.12 (Suite de Cauchy). Soit (x_n) une suite de vecteurs d'un espace vectoriel normé $(\mathcal{H}, \|\cdot\|)$. On dit que (x_n) est une suite de Cauchy si et seulement si :

$$\|x_n - x_m\| \rightarrow 0$$

quand $n, m \rightarrow +\infty$.

Notons qu'en vertu de l'inégalité triangulaire toute suite convergente est une suite de Cauchy. La réciproque est fautive : une suite de Cauchy peut ne pas être convergente. Un contre-exemple est donné par l'exemple 2.1.18.

Définition 2.1.13 (Espace complet, espace de Hilbert). On dit qu'un espace vectoriel normé $(\mathcal{H}, \|\cdot\|)$ est *complet* si toute suite de Cauchy d'éléments de cet espace converge dans cet espace. On dit \mathcal{H} est un *espace de Hilbert* si \mathcal{H} est pré-hilbertien et complet.

Proposition 2.1.14 (Suites normalement convergentes). *Un espace vectoriel normé $(\mathcal{H}, \|\cdot\|)$ est complet si et seulement si toute série normalement convergente est convergente.*

Ce résultat classique, voir [4, proposition 5 du chapitre 6, page 124], est utile pour montrer que les espaces L^p sont complets.

Exemple 2.1.15 (Espace de suite). L'espace ℓ^2 est un espace de Hilbert. Soit (a_n) une suite de Cauchy dans ℓ^2 . Si nous notons

$$a_n = (a_{n,1}, a_{n,2}, \dots),$$

alors, pour tout $\epsilon > 0$, il existe N tel que, pour tout $n, m \geq N$,

$$\sum_{k=1}^{\infty} |a_{m,k} - a_{n,k}| \leq \epsilon^2, \quad (2.1)$$

pour tout $n, m \geq N$. Fixons tout d'abord k . La relation précédente montre que la suite $(a_{n,k})$ est une suite de Cauchy dans \mathbb{C} . Cette suite converge donc vers α_k . Nous notons $a = (\alpha_k)$. Nous allons montrer que $a \in \ell^2$ et que $\lim_{n \rightarrow \infty} \|a_n - a\| = 0$. Comme l'espace ℓ^2 est stable par différence, nous allons montrer que pour tout n , $a_n - a \in \ell^2$. Comme $a = a_n - (a_n - a)$ et $a_n \in \ell^2$, cette propriété implique donc que $a \in \ell^2$.

En utilisant (2.1), nous avons pour tout $p \in \mathbb{N}$, et tout $m, n \geq N$,

$$\sum_{k=1}^p |a_{m,k} - a_{n,k}|^2 \leq \sum_{k=1}^{\infty} |a_{m,k} - a_{n,k}|^2 \leq \epsilon^2.$$

Par conséquent, pour tout $p \in \mathbb{N}$ et tout $n \geq N$, $\lim_{m \rightarrow \infty} \sum_{k=1}^p |a_{m,k} - a_{n,k}|^2 = \sum_{k=1}^p |a_k - a_{n,k}|^2 \leq \epsilon^2$. En prenant, la limite en p , nous obtenons donc, pour tout $n \geq N$,

$$\|a - a_n\|^2 = \sum_{k=1}^{\infty} |a_k - a_{n,k}|^2 \leq \epsilon^2,$$

ce qui montre que $(a - a_n) \in \ell^2$. Comme ϵ est arbitraire, nous avons aussi $\lim_{n \rightarrow \infty} \|a - a_n\| = 0$.

Proposition 2.1.16 (Espaces L^2). *Pour tout espace mesurable $(\Omega, \mathcal{F}, \mu)$, l'espace $L^2(\Omega, \mathcal{F}, \mu)$ (voir l'exemple 2.1.5) des fonctions de carré intégrable pour la mesure μ est un espace de Hilbert.*

Un résultat plus général sur les espaces L^p est fourni par [4, proposition 6 du chapitre 6, page 126].

Définition 2.1.17 (Sous-espace fermé). Soit \mathcal{E} un sous-espace d'un espace de Hilbert \mathcal{H} . On dit que \mathcal{E} est fermé, si toute suite (x_n) de \mathcal{E} , qui converge, converge dans \mathcal{E} .

Exemple 2.1.18 (Sous-espace non-fermé, espace non-complet). Soit $\mathcal{C}([-\pi, \pi])$ l'espace des fonctions continues sur $[-\pi, \pi]$. Cet espace est un sous-espace de l'espace de Hilbert $L^2([-\pi, \pi])$. Considérons la suite de fonctions :

$$f_n(x) = \sum_{k=1}^n \frac{1}{k} \cos(kx)$$

Les fonctions $f_n(x)$, qui sont indéfiniment continûment différentiables, appartiennent à $\mathcal{C}(-\pi, \pi)$. Montrons que cette suite est une suite de Cauchy. En effet, pour $m > n$, on a :

$$\|f_n - f_m\|^2 = \pi \sum_{k=n+1}^m \frac{1}{k^2} \longrightarrow 0 \quad \text{quand } (n, m) \rightarrow \infty$$

D'autre part on montre aisément que la limite de cette suite $f_{\infty}(x) = \sum_{k=1}^{\infty} k^{-1} \cos(kx) = \log |\sin(x/2)|$ n'est pas continue et n'appartient donc pas à $\mathcal{C}([-\pi, \pi])$. Le sous-espace $\mathcal{C}([-\pi, \pi])$ est donc un sous-espace vectoriel de $L^2([-\pi, \pi])$ mais n'est pas fermé. C'est donc aussi un espace pré-hilbertien qui n'est pas complet.

Définition 2.1.19 (Sous espace engendré par un sous-ensemble). Soit \mathcal{X} un sous-ensemble de \mathcal{H} . Nous notons $\text{Vect}(\mathcal{X})$ le sous-espace vectoriel des combinaisons linéaires finies d'éléments de \mathcal{X} et $\overline{\text{Vect}(\mathcal{X})}$ l'adhérence de $\text{Vect}(\mathcal{X})$ dans \mathcal{H} , c'est-à-dire le plus petit sous-ensemble fermé de \mathcal{H} contenant $\text{Vect}(\mathcal{X})$, ou encore l'espace obtenu par l'ensemble $\text{Vect}(\mathcal{X})$ complété de toutes les limites de suite d'éléments de $\text{Vect}(\mathcal{X})$.

Définition 2.1.20 (Orthogonalité). Deux vecteurs $x, y \in \mathcal{H}$ sont dit *orthogonaux*, si $\langle x, y \rangle = 0$, ce que nous notons $x \perp y$. Si \mathcal{S} est un sous-ensemble de \mathcal{H} , la notation $x \perp \mathcal{S}$, signifie que $x \perp s$ pour tout $s \in \mathcal{S}$. Nous notons $\mathcal{S} \perp \mathcal{T}$ si tout élément de \mathcal{S} est orthogonal à tout élément de \mathcal{T} .

Supposons qu'il existe deux sous-espaces \mathcal{A} et \mathcal{B} tels que $\mathcal{H} = \mathcal{A} \oplus \mathcal{B}$, dans le sens où, pour tout vecteur $h \in \mathcal{H}$, il existe $a \in \mathcal{A}$ et $b \in \mathcal{B}$, tel que $h = a + b$. Si en plus $\mathcal{A} \perp \mathcal{B}$ nous dirons que \mathcal{H} est la *somme directe* de \mathcal{A} et \mathcal{B} , ce que nous notons $\mathcal{H} = \mathcal{A} \overset{\perp}{\oplus} \mathcal{B}$.

Définition 2.1.21 (Complément orthogonal). Soit \mathcal{E} un sous-ensemble d'un espace de Hilbert \mathcal{H} . On appelle ensemble orthogonal de \mathcal{E} , l'ensemble défini par :

$$\mathcal{E}^\perp = \{x \in \mathcal{H} : \forall y \in \mathcal{E} \quad \langle x, y \rangle = 0\}$$

2.2 Famille orthogonale et orthonormal

Définition 2.2.1 (Famille orthogonale, orthonormale). Soit E un sous ensemble de \mathcal{H} . On dit que E est une famille orthogonale si et seulement si pour tout $(x, y) \in E \times E$, $x \neq y$, $\langle x, y \rangle = 0$. Si de plus $\|x\| = 1$ pour tout $x \in E$, on dira que E est une famille orthonormale.

Une famille orthogonale a la propriété remarquable que dans le développement de la norme quadratique des combinaisons linéaires finies, les termes de doubles produits sont nuls. Soit E une famille orthogonale, $(x_1, \dots, x_n) \in E^n$ et $(\alpha_1, \dots, \alpha_n) \in \mathbb{C}^n$. Alors

$$\left\| \sum_{k=1}^n \alpha_k x_k \right\|^2 = \sum_{k=1}^n |\alpha_k|^2 \|x_k\|^2. \quad (2.2)$$

Il s'en suit donc le théorème élémentaire suivant.

Théorème 2.2.2. *Toute famille orthogonale d'éléments non-nuls est libre.*

La relation (2.2) est bien connue en géométrie euclidienne. L'avantage du cadre hilbertien est qu'il permet d'obtenir des résultats pour une somme infinie.

Théorème 2.2.3. *Soit $(e_i)_{i \geq 1}$ une suite orthonormale d'un espace de Hilbert \mathcal{H} et soit $(\alpha_i)_{i \geq 1}$ une suite de nombre complexes. La série*

$$\sum_{i=1}^{\infty} \alpha_i e_i \quad (2.3)$$

converge dans \mathcal{H} si et seulement si $\sum_i |\alpha_i|^2 < \infty$ auquel cas

$$\left\| \sum_{i=1}^{\infty} \alpha_i e_i \right\|^2 = \sum_{i=1}^{\infty} |\alpha_i|^2. \quad (2.4)$$

Preuve. Pour tout $m > k > 0$, de même que pour l'équation (2.2), nous avons

$$\left\| \sum_{i=k}^m \alpha_i e_i \right\|^2 = \sum_{i=k}^m |\alpha_i|^2.$$

Comme $\sum_{i=1}^{\infty} |\alpha_i|^2 < \infty$, la suite $s_m = \sum_{i=1}^m \alpha_i e_i$ est une suite de Cauchy dans \mathcal{H} . Comme \mathcal{H} est complet, cette suite de Cauchy converge. L'identité (2.4) est obtenue par passage à la limite.

Réciproquement, si la série $\sum_{i=1}^{\infty} \alpha_i e_i$ converge, ce passage à la limite reste valide et (2.4) montre que la série de terme $(|\alpha_i|^2)_{i \geq 1}$ est convergente. \square

La question se pose aussi de savoir comment approcher un élément x de \mathcal{H} par une décomposition en série de la forme (2.3). Pour cela le résultat suivant pour une famille finie sera utile.

Proposition 2.2.4. *Si x est un vecteur d'un espace de Hilbert \mathcal{H} et si $E = \{e_1, \dots, e_n\}$ est une famille orthonormale finie, alors :*

$$\left\| x - \sum_{k=1}^n \langle x, e_k \rangle e_k \right\|^2 = \|x\|^2 - \sum_{k=1}^n |\langle x, e_k \rangle|^2. \quad (2.5)$$

De plus $\sum_{k=1}^n \langle x, e_k \rangle e_k$ est l'élément de $\text{Vect}(e_1, \dots, e_n)$ le plus proche de x : la quantité (2.5) est aussi égale à

$$\inf \{ \|x - y\|^2 : y \in \text{Vect}(e_1, \dots, e_n) \}.$$

Preuve. On remarque que pour tout $j = 1, \dots, n$,

$$\left\langle x - \sum_{k=1}^n \langle x, e_k \rangle e_k, e_j \right\rangle = \langle x, e_j \rangle - \langle x, e_j \rangle = 0.$$

Il s'en suit que la décomposition

$$x = \left(x - \sum_{k=1}^n \langle x, e_k \rangle e_k \right) + \sum_{i=1}^n \langle x, e_k \rangle e_k$$

est une décomposition en la somme de deux termes orthogonaux. L'identité de Pythagore et l'égalité (2.2) avec $x_k = e_k$ et $\alpha_k = \langle x, e_k \rangle$

On montre de même que, pour tout $(\alpha_1, \dots, \alpha_n) \in \mathbb{C}^n$,

$$\left\| x - \sum_{k=1}^n \alpha_k e_k \right\|^2 = \left\| x - \sum_{k=1}^n \langle x, e_k \rangle e_k \right\|^2 + \sum_{k=1}^n |\langle x, e_k \rangle - \alpha_k|^2,$$

et donc que $\sum_{k=1}^n \langle x, e_k \rangle e_k$ est la meilleure approximation de x par une combinaison linéaire des vecteurs e_1, \dots, e_n . \square

Cette propriété d'approximation des familles orthonormales joue un rôle essentiel.

Exemple 2.2.5 (procédé de Gram-Schmidt). Soit $(y_i)_{i \geq 1}$ une famille d'éléments d'un espace de Hilbert \mathcal{H} . Le procédé de Gram-Schmidt est un procédé par récurrence qui permet alors de construire une famille orthogonale qui vérifie la propriété $\text{Vect}(e_1, \dots, e_n) = \text{Vect}(y_1, \dots, y_n)$ pour tout $n \geq 1$. Nous donnons ici la construction de la suite $(e_i)_{i \geq 1}$. La preuve de ses propriétés est laissée à titre d'exercice.

En partant d'une suite orthonormale (e_i) et en appliquant la proposition 2.2.4 pour toute sous-suite finie, on obtient aussi le résultat suivant.

Corollaire 2.2.6 (Inégalité de Bessel). *Soit $(e_i)_{i \geq 1}$ une suite orthonormale d'un espace de Hilbert \mathcal{H} . Alors*

$$\sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2 \leq \|x\|^2.$$

L'inégalité de Bessel implique que pour tout $x \in \mathcal{H}$, $\lim_{n \rightarrow \infty} \langle x, e_n \rangle = 0$ (la suite de vecteurs (e_n) converge *faiblement* vers 0) mais aussi que la suite $(\langle x, e_i \rangle)_{i \geq 1}$ est un élément de l'espace ℓ^2 des suites de carrés sommables. En appliquant le théorème 2.2.3, on obtient que le développement en série

$$\sum_{i=1}^{\infty} \langle x, e_i \rangle e_i \tag{2.6}$$

est toujours convergent. On l'appelle le *développement de Fourier généralisé* de x ; les coefficients $\langle x, e_i \rangle$ sont appelés *coefficients de Fourier généralisés* par rapport à la suite orthonormale (e_i) . Il faut prendre garde toutefois au fait que si la série $\sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$ converge, sa limite n'est pas nécessairement égale à x .

Exemple 2.2.7. Considérons $\mathcal{H} = L^2(\mathbb{T})$ et soit $e_n(t) = \pi^{-1/2} \sin(nt)$ pour $n = 1, 2, \dots$. La suite (e_n) est orthonormale dans \mathcal{H} , mais pour $x(t) = \cos(t)$, nous avons

$$\begin{aligned} \sum_{n=1}^{\infty} \langle x, e_n \rangle e_n(t) &= \sum_{n=1}^{\infty} \left[\pi^{-1/2} \int_{\mathbb{T}} \cos(t) \sin(nt) dt \right] \pi^{-1/2} \sin(nt) \\ &= \sum_{n=1}^{\infty} 0 \cdot \sin(nt) = 0 \neq \cos t. \end{aligned}$$

En fait, pour obtenir x , il faut une propriété supplémentaire.

Définition 2.2.8 (Famille complète, base hilbertienne). Une famille E d'éléments d'un espace de Hilbert \mathcal{H} est dite *complète* si $\overline{\text{Vect}(E)} = \mathcal{H}$. Une *suite orthonormale complète* s'appelle une *base hilbertienne*.

La complétude signifie donc n'importe quel élément de \mathcal{H} s'écrit comme la limite d'une suite de combinaisons linéaires finies de la famille considérées. Pour les bases hilbertiennes, cette suite se construit aisément sous la forme de la série (2.6), comme l'indique le résultat suivant.

Théorème 2.2.9. *Soit $(e_i)_{i \geq 1}$ une base hilbertienne de l'espace de Hilbert \mathcal{H} . Alors pour tout $x \in \mathcal{H}$,*

$$x = \sum_{i=1}^{\infty} \langle x, e_i \rangle e_i . \quad (2.7)$$

Preuve. Nous savons que la série (2.6) converge. D'autre part la suite étant complète, il existe un tableau $(\alpha_{p,n})_{1 \leq i \leq n}$ tel que

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \alpha_{i,n} e_i = x .$$

Or d'après la proposition 2.2.4, on a

$$\left\| x - \sum_{i=1}^n \sum_{i=1}^n \langle x, e_i \rangle e_i \right\| \leq \left\| x - \sum_{i=1}^n \sum_{i=1}^n \alpha_{i,n} e_i \right\| .$$

On a donc aussi convergence du développement de Fourier généralisé de x vers x . \square

Le théorème 2.2.9 montre en particulier qu'une famille orthonormale est une base hilbertienne si et seulement si l'identité (2.7) entre un élément et son développement de Fourier est vérifié pour tout élément. Ceci implique aisément le résultat suivant dont la preuve est laissée à titre d'exercice.

Théorème 2.2.10. *Soit $(e_i)_{i \geq 1}$ une suite orthonormale d'un espace de Hilbert \mathcal{H} . Les trois propositions suivantes sont équivalentes.*

- a) $(e_i)_{i \geq 1}$ est une base hilbertienne.
- b) L'élément nul est l'unique élément qui satisfait

$$\langle x, e_i \rangle = 0 \quad \text{pour tout } i \geq 1 .$$

- c) Pour tout $x \in \mathcal{H}$,

$$\|x\|^2 = \sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2 . \quad (2.8)$$

Exemple 2.2.11 (Base de Fourier). Le système de fonctions

$$e_n(x) = (2\pi)^{-1/2} e^{inx}, n \in \mathbb{Z}$$

est une suite orthonormale complète de $L^2(\mathbb{T})$. La preuve de l'orthogonalité est élémentaire, mais la preuve de la complétude est plus délicate. Nous l'établirons dans le paragraphe 2.3.

Définition 2.2.12 (Espace de Hilbert séparable). On dit qu'un espace de Hilbert est séparable s'il contient un sous-ensemble dénombrable dense.

L'intérêt d'un espace de Hilbert séparable est qu'il existe une base hilbertienne. C'est d'ailleurs une condition suffisante.

Théorème 2.2.13. *Un espace de Hilbert \mathcal{H} est séparable si et seulement si il existe une base hilbertienne.*

Preuve. Soit (e_i) une base hilbertienne de \mathcal{H} . L'ensemble $S = \bigcup_{n=1}^{\infty} S_n$ avec, pour $n \in \mathbb{N}$,

$$S_n \stackrel{\text{def}}{=} \left\{ \sum_{l=1}^n (\alpha_k + i\beta_k) e_k, (\alpha_k, \beta_k) \in \mathbb{Q} \times \mathbb{Q}, k = 1, \dots, n \right\}$$

est dénombrable (comme union dénombrable d'ensembles dénombrables). Comme, pour $x \in \mathcal{H}$,

$$\lim_{n \rightarrow \infty} \left\| \sum_{k=1}^n \langle x, e_k \rangle e_k - x \right\| = 0,$$

l'ensemble S est dense dans \mathcal{H} .

Si \mathcal{H} est séparable alors il existe une suite $(y_i)_{i \geq 1}$ dense et donc aussi complète. Le procédé de Gram-Schmidt décrit à l'exemple 2.2.5 permet alors de construire une famille orthogonale $(e_i)_{i \geq 1}$ telle que $\text{Vect}(e_1, \dots, e_n) = \text{Vect}(y_1, \dots, y_n)$ pour tout n . En retirant les éléments nuls de cette suite et en renormalisant ses termes non-nuls pour qu'ils soient de norme 1, on obtient alors une base hilbertienne. \square

2.3 Séries de Fourier

Dans cette partie, nous allons établir que

$$\phi_n(x) = (2\pi)^{-1/2} e^{inx}, \quad n \in \mathbb{Z}, \quad (2.9)$$

est une famille complète de $L^2(\mathbb{T}, \mathcal{B}(\mathbb{T}), \mu)$ quelque soit la mesure finie μ sur les boréliens du tore \mathbb{T} . En particulier si μ est proportionnelle à la mesure de Lebesgue, on obtient que cette suite forme une famille orthogonale complète.

Notons $L^1(\mathbb{T})$ l'ensemble des fonctions 2π -périodiques localement intégrables par rapport à la mesure de Lebesgue \mathbb{R} . Pour $f \in L^1(\mathbb{T})$, posons

$$f_n = \sum_{k=-n}^n \langle g, \phi_k \rangle \phi_k, \quad n = 0, 1, 2, \dots$$

Nous avons

$$f_n(x) = \sum_{k=-n}^n \frac{1}{2\pi} \int_{\mathbb{T}} f(t) e^{-ikt} dt = \sum_{k=-n}^n \frac{1}{2\pi} \int_{\mathbb{T}} f(t) e^{ik(x-t)} dt.$$

Nous allons établir que, pour tout $f \in L^1(\mathbb{T})$,

$$\frac{1}{n+1} \sum_{k=0}^n f_k,$$

qui est dans $\text{Vect}(\phi_n, n \in \mathbb{Z})$ est une *bonne approximation* de f , en précisant la notion d'approximation utilisée, suivant les hypothèses supplémentaires sur f . Remarquons que pour tout $x \in \mathbb{R}$,

$$\begin{aligned} \frac{1}{n+1} \sum_{k=0}^n f_k(x) &= \sum_{k=-n}^n \left(1 - \frac{|k|}{n+1}\right) \langle f, \phi_k \rangle \phi_k(x) \\ &= \frac{1}{2\pi} \int_{\mathbb{T}} f(t) \left[\sum_{k=-n}^n \left(1 - \frac{|k|}{n+1}\right) e^{ik(x-t)} \right] dt. \end{aligned}$$

On note la fonction entre crochets $K_n(x-t)$, d'où finalement, pour tout $x \in \mathbb{R}$,

$$\frac{1}{n+1} \sum_{k=0}^n f_k(x) = \frac{1}{2\pi} \int_{\mathbb{T}} f(t) K_n(x-t) dt. \quad (2.10)$$

Un calcul élémentaire donne

$$K_n(u) = \frac{1}{n+1} \frac{\sin^2 \frac{(n+1)u}{2}}{\sin^2 \frac{u}{2}}. \quad (2.11)$$

Définition 2.3.1 (Noyau de sommabilité). Nous dirons qu'une suite de fonctions (κ_n) de fonctions 2π -périodique continue est un *noyau de sommabilité* si, pour tout $n \in \mathbb{N}$

$$\int_{\mathbb{T}} \kappa_n(t) dt = 2\pi \quad (2.12)$$

$$\int_{\mathbb{T}} |\kappa_n(t)| dt \leq M, \quad \text{où } M \text{ est une constante} \quad (2.13)$$

$$\lim_{n \rightarrow \infty} \int_{\delta}^{2\pi-\delta} |\kappa_n(t)| dt = 0, \quad \text{pour tout } \delta \in]0, \pi[. \quad (2.14)$$

Lemme 2.3.2. *La fonction $x \rightarrow K_n(x)$ donnée par (2.11) est un noyau de sommabilité, appelé noyau de Fejer.*

Preuve. Comme $\int_{\mathbb{T}} e^{ikt} dt = 0$ si $k \neq 0$, nous avons $\int_{\mathbb{T}} K_n(t) dt = 2\pi$. Comme $K_n(t) \geq 0$ pour tout $t \in \mathbb{T}$, nous avons de même $\int_{\mathbb{T}} |K_n(t)| dt = 2\pi$. Finalement, soit $\delta \in]0, \pi[$. Pour tout $t \in]\delta, 2\pi - \delta[$, $\sin t/2 \geq \sin \delta/2$, ce qui implique

$$K_n(t) \leq \frac{1}{(n+1) \sin^2 \delta/2}.$$

Par conséquent

$$\int_{\delta}^{2\pi-\delta} K_n(t) dt \leq \frac{2\pi}{(n+1) \sin^2 \delta/2} \xrightarrow{n \rightarrow \infty} 0.$$

□

Le résultat suivant montre que la convolution avec un noyau de sommabilité fournit une approximation uniforme d'une fonction continue. On appelle ce procédé d'approximation une *régularisation*.

Lemme 2.3.3. *Soient $f : \mathbb{R} \rightarrow \mathbb{C}$ une fonction 2π -périodique continue sur \mathbb{R} et (κ_n) un noyau de sommabilité. Alors*

$$\sup_{x \in \mathbb{R}} \left| \frac{1}{2\pi} \int_{\mathbb{T}} f(t) \kappa_n(x-t) dt - f(x) \right| \rightarrow 0 \quad \text{quand } n \rightarrow \infty.$$

Preuve. En utilisant (2.12), on a, pour tout x ,

$$\left| \frac{1}{2\pi} \int_{\mathbb{T}} f(t) \kappa_n(x-t) dt - f(x) \right| = \left| \frac{1}{2\pi} \int_{\mathbb{T}} [f(t) - f(x)] \kappa_n(x-t) dt \right|.$$

Il suffit donc de montrer que

$$\sup_{x \in \mathbb{R}} \left| \int_{-\pi}^{\pi} [f(x-u) - f(x)] \kappa_n(u) du \right| \rightarrow 0. \quad (2.15)$$

Comme f est continue sur \mathbb{R} et périodique, f est uniformément continue. Soit $\epsilon > 0$. Il existe alors $\delta \in (0, \pi)$ tel que $|f(x) - f(t)| \leq \epsilon$ pour $|x - t| \leq \delta$. Il s'en suit donc en séparant l'intégrale en 2 parties suivant que $|u| \leq \delta$ ou l'inverse :

$$\left| \int_{-\pi}^{\pi} [f(x-u) - f(x)] \kappa_n(u) du \right| \leq \int_{\delta < |u| < \pi} |\kappa_n(u)| du + \epsilon \int_{-\delta}^{\delta} |\kappa_n(u)| du.$$

On remarque alors que cette majoration ne dépend pas de x , que son premier terme tend vers 0 quand $n \rightarrow \infty$ en vertu de (2.14) et que son deuxième terme est borné par $M\epsilon$ en utilisant (2.13). Comme ϵ peut être pris arbitrairement petit, on en conclut (2.15). □

Corollaire 2.3.4. *Soit μ une mesure finie sur les boréliens du tore. La suite $(\phi_n)_{n \in \mathbb{Z}}$ définie par (2.9) génère l'espace de Hilbert $L^2(\mathbb{T}, \mathcal{B}(\mathbb{T}), \mu)$, ce qu'on écrit $\text{Vect}(\phi_n, n \in \mathbb{Z}) = L^2(\mathbb{T}, \mathcal{B}(\mathbb{T}), \mu)$.*

Preuve. D'après [4, proposition 8 du chapitre 6, page 128], pour tout $\epsilon > 0$ et toute $f \in L^2(\mathbb{T}, \mathcal{B}(\mathbb{T}), \mu)$, il existe une fonction f_ϵ continue sur $[0, 2\pi]$ telle que

$$\int_{[0, 2\pi)} |f - f_\epsilon|^2 d\mu \leq \epsilon^2.$$

Soit $\epsilon' > 0$ arbitrairement petit. On note $g_{\epsilon'}$ la fonction continue et 2π -périodique qui coïncide avec f_ϵ sur $[\epsilon', 2\pi - \epsilon']$ et qui est linéaire sur $[-\epsilon', \epsilon']$. Alors

$$\int_{[0, 2\pi)} |f_\epsilon - g_{\epsilon'}|^2 d\mu \leq 2\mu([-\epsilon', \epsilon']) \sup_{t \in [0, 2\pi]} |f_\epsilon(t)|^2.$$

On remarque que ce majorant tend vers 0 quand $\epsilon' \rightarrow 0$, en particulier, il existe $\epsilon' > 0$ tel que ce majorant est inférieur à ϵ^2 et donc tel que

$$\left(\int_{\mathbb{T}} |f - g_{\epsilon'}|^2 d\mu \right)^{1/2} \leq 2\epsilon.$$

Autrement dit les fonctions 2π -périodiques continues sur \mathbb{R} approchent n'importe quelle fonction de $L^2(\mathbb{T}, \mathcal{B}(\mathbb{T}), \mu)$. D'après (2.10) les lemmes 2.3.3 et 2.3.2, on a d'autre part que toute fonction continue sur le tore peut être arbitrairement approchée par une fonction de $\text{Vect}(\phi_n, n \in \mathbb{Z})$ en norme sup, et donc aussi en norme $L^2(\mathbb{T}, \mathcal{B}(\mathbb{T}), \mu)$, ce qui conclut la preuve. \square

Dans le cas où μ est la mesure de Lebesgue sur le tore, on obtient le résultat suivant.

Corollaire 2.3.5. *La suite $(\phi_n)_{n \in \mathbb{Z}}$ définie par (2.9) forme une famille orthogonale complète de $L^2(\mathbb{T})$.*

Le corollaire 2.3.5 montre que pour tout $f \in L^2(\mathbb{T})$,

$$f = \sum_{k=-\infty}^{\infty} \alpha_k \phi_k \quad \text{avec} \quad \alpha_k = (2\pi)^{-1/2} \int_{\mathbb{T}} f(x) e^{-ikx} dx,$$

où la somme infinie converge au sens $L^2(\mathbb{T})$. En général, cette condition n'implique pas la convergence ponctuelle. Les coefficients (α_k) sont appelés les coefficients de Fourier de f . L'identité de Parseval s'écrit dans ce cas

$$\int_{\mathbb{T}} |f(x)|^2 dx = \sum_{k=-\infty}^{\infty} |\alpha_k|^2.$$

2.4 Projection et principe d'orthogonalité

Le théorème suivant, appelé *théorème de projection*, joue un rôle central en analyse Hilbertienne.

Théorème 2.4.1. *Soit \mathcal{E} est un sous-ensemble convexe fermé d'un espace de Hilbert \mathcal{H} et soit x un élément quelconque de \mathcal{H} , alors :*

a) *il existe un unique élément noté $\text{proj}(x|\mathcal{E}) \in \mathcal{E}$ tel que :*

$$\|x - \text{proj}(x|\mathcal{E})\| = \inf_{w \in \mathcal{E}} \|x - w\|$$

b) *Si de plus \mathcal{E} est un espace vectoriel, $\text{proj}(x|\mathcal{E})$ est l'unique élément $\hat{x} \in \mathcal{E}$ tel que $x - \hat{x} \in \mathcal{E}^\perp$.*

On appelle $\text{proj}(x|\mathcal{E})$ la projection orthogonale de x sur \mathcal{E} .

Preuve. a) Soit $x \in \mathcal{H}$. On note $h = \inf_{w \in \mathcal{E}} \|x - w\| \geq 0$. Alors il existe une suite w_1, w_2, \dots , de vecteurs de \mathcal{E} tels que :

$$\lim_{m \rightarrow +\infty} \|x - w_m\|^2 = h^2 \geq 0 \quad (2.16)$$

L'identité du parallélogramme, $\|a - b\|^2 + \|a + b\|^2 = 2\|a\|^2 + 2\|b\|^2$ avec $a = w_m - x$ et $b = w_n - x$, montre que :

$$\|w_m - w_n\|^2 + \|w_m + w_n - 2x\|^2 = 2\|w_m - x\|^2 + 2\|w_n - x\|^2$$

Comme $(w_m + w_n)/2 \in \mathcal{E}$, nous avons $\|w_m + w_n - 2x\|^2 = 4\|(w_m + w_n)/2 - x\|^2 \geq 4h^2$. D'après 2.16, pour tout $\epsilon > 0$, il existe N tel que et $\forall m, n > N$:

$$\|w_m - w_n\|^2 \leq 2(h^2 + \epsilon) + 2(h^2 + \epsilon) - 4h^2 = 4\epsilon.$$

qui montre que $\{w_n, n \in \mathbb{N}\}$ est une suite de Cauchy et donc que la suite $\{w_n, n \in \mathbb{N}\}$ tend vers une limite dans \mathcal{E} , puisque l'espace \mathcal{E} est fermé. On note y cette limite. On en déduit, par continuité de la norme, que $\|y - x\| = h$. Montrons que cet élément est unique. Supposons qu'il existe un autre élément $z \in \mathcal{E}$ tel que $\|x - z\|^2 = \|x - y\|^2 = h^2$. Alors l'identité du parallélogramme donne :

$$\begin{aligned} 0 \leq \|y - z\|^2 &= -4\|(y + z)/2 - x\|^2 + 2\|x - y\|^2 + 2\|x - z\|^2 \\ &\leq -4h^2 + 2h^2 + 2h^2 = 0 \end{aligned}$$

où nous avons utilisé que $(y + z)/2 \in \mathcal{E}$ par hypothèse de convexité et donc $\|(y + z)/2 - x\|^2 \geq h^2$. Il s'en suit que $y = z$, d'où l'unicité.

b) Soit \hat{x} la projection orthogonale de x sur \mathcal{E} . Alors, si il existe $u \in \mathcal{E}$ tel que $x - u \perp \mathcal{E}$, on peut écrire :

$$\begin{aligned}
\|x - \hat{x}\|^2 &= \langle x - u + u - \hat{x}, x - u + u - \hat{x} \rangle \\
&= \|x - u\|^2 + \|u - \hat{x}\|^2 + 2 \langle u - \hat{x}, x - u \rangle \\
&= \|x - u\|^2 + \|u - \hat{x}\|^2 + 0 \geq \|x - u\|^2
\end{aligned}$$

et donc $u = \hat{x}$. Réciproquement supposons que $u \in \mathcal{E}$ et $x - u \notin \mathcal{E}$. Alors choisissons $y \in \mathcal{E}$ tel que $\|y\| = 1$ et tel que $c = \langle x - u, y \rangle \neq 0$ et notons $\tilde{x} = u + cy \in \mathcal{E}$. On a :

$$\begin{aligned}
\|x - \tilde{x}\|^2 &= \langle x - u + u - \tilde{x}, x - u + u - \tilde{x} \rangle \\
&= \|x - u\|^2 + \|u - \tilde{x}\|^2 + 2 \langle u - \tilde{x}, x - u \rangle \\
&= \|x - u\|^2 + c^2 - 2c \langle y, x - u \rangle = \|x - u\|^2 - c^2 < \|x - u\|^2
\end{aligned}$$

Par conséquent $\tilde{x} \in \mathcal{E}$ est strictement plus proche de x que ne l'est u . \square

Le point (b)) est particulièrement important pour calculer la projection puisqu'il permet de remplacer un problème de minimisation par la résolution d'une équation linéaire.

Exemple 2.4.2 (Projection sur un vecteur). Soit \mathcal{H} un espace de Hilbert, $\mathcal{C} = \text{Vect}(v)$ le sous-espace engendré par un vecteur $v \in \mathcal{H}$ et x un vecteur quelconque de \mathcal{H} . On a alors $\text{proj}(x|\mathcal{C}) = \alpha v$ avec $\alpha = \langle x, v \rangle / \|v\|^2$. Si on note $\epsilon = x - \text{proj}(x|\mathcal{C})$, on a :

$$\|\epsilon\|^2 = \|x\|^2 (1 - \|\rho\|^2) \quad \text{où} \quad \rho = \frac{\langle x, v \rangle}{\|x\| \|v\|} \quad \text{avec} \quad |\rho| \leq 1$$

La projection définie par la proposition 2.4.3 satisfait les propriétés intéressantes suivantes.

Proposition 2.4.3. *Soit \mathcal{H} un espace de Hilbert et $\text{proj}(\cdot|\mathcal{E})$ la projection orthogonale sur le sous-espace fermé \mathcal{E} . On a :*

1. *l'application $x \in \mathcal{H} \mapsto \text{proj}(x|\mathcal{E}) \in \mathcal{E}$ est linéaire :*

$$\forall (\alpha, \beta) \in \mathbb{C} \times \mathbb{C}, \quad \text{proj}(\alpha x + \beta y|\mathcal{E}) = \alpha \text{proj}(x|\mathcal{E}) + \beta \text{proj}(y|\mathcal{E}).$$

2. $\|x\|^2 = \|\text{proj}(x|\mathcal{E})\|^2 + \|x - \text{proj}(x|\mathcal{E})\|^2$ (Pythagore),

3. *La fonction $\text{proj}(\cdot|\mathcal{E}) : \mathcal{H} \rightarrow \mathcal{H}$ est continue,*

4. $x \in \mathcal{E}$ si et seulement si $\text{proj}(x|\mathcal{E}) = x$,

5. $x \in \mathcal{E}^\perp$ si et seulement si $\text{proj}(x|\mathcal{E}) = 0$,

6. *Soient \mathcal{E}_1 et \mathcal{E}_2 deux sous espaces vectoriels fermés de \mathcal{H} , tels que $\mathcal{E}_1 \subset \mathcal{E}_2$.*

Alors :

$$\forall x \in \mathcal{H}, \quad \text{proj}(\text{proj}(x|\mathcal{E}_2)|\mathcal{E}_1) = \text{proj}(x|\mathcal{E}_1).$$

7. Soient \mathcal{E}_1 et \mathcal{E}_2 deux sous-espaces vectoriels fermés de \mathcal{H} , tels que $\mathcal{E}_1 \perp \mathcal{E}_2$.
Alors :

$$\forall x \in \mathcal{H}, \quad \text{proj} \left(x | \mathcal{E}_1 \oplus^\perp \mathcal{E}_2 \right) = \text{proj} (x | \mathcal{E}_1) + \text{proj} (x | \mathcal{E}_2) .$$

Théorème 2.4.4. Si \mathcal{E} est un sous-ensemble d'un espace de Hilbert \mathcal{H} , alors \mathcal{E}^\perp est un sous-espace fermé.

Preuve. Soit $(x_n)_{n \geq 0}$ une suite convergente d'éléments de \mathcal{E}^\perp . Notons x la limite de cette suite. Par continuité du produit scalaire nous avons, pour tout $y \in \mathcal{E}$,

$$\langle x, y \rangle = \lim_{n \rightarrow \infty} \langle x_n, y \rangle = 0$$

et donc $x \in \mathcal{E}^\perp$. □

Théorème 2.4.5. Soit $(\mathcal{M}_n)_{n \in \mathbb{Z}}$ une suite croissante de sous-espaces vectoriels (s.e.v.) fermés d'un espace de Hilbert \mathcal{H} .

a) Soit $\mathcal{M}_{-\infty} = \bigcap_n \mathcal{M}_n$. Alors, pour tout $h \in \mathcal{H}$, nous avons

$$\text{proj} (h | \mathcal{M}_{-\infty}) = \lim_{n \rightarrow -\infty} \text{proj} (h | \mathcal{M}_n)$$

b) Soit $\mathcal{M}_\infty = \overline{\bigcup_{n \in \mathbb{Z}} \mathcal{M}_n}$. Alors, pour tout $h \in \mathcal{H}$,

$$\text{proj} (h | \mathcal{M}_\infty) = \lim_{n \rightarrow \infty} \text{proj} (h | \mathcal{M}_n) .$$

Preuve. Remarquons tout d'abord (b)) se déduit aisément en appliquant (a)) et en remarquant que

$$\mathcal{M}_\infty^\perp = \bigcap_n \mathcal{M}_n^\perp ,$$

et donc, comme \mathcal{M}_∞ et les \mathcal{M}_n sont fermés, on a d'après la propriété 7 de la proposition 2.4.3 : $\text{proj} (h | \mathcal{M}_{-\infty}) = h - \text{proj} (h | \mathcal{M}_\infty^\perp)$ et de même pour \mathcal{M}_n . De plus comme les \mathcal{M}_n^\perp sont fermés d'après le théorème 2.4.4, on peut bien appliquer (a)).

Il reste donc à montrer (a)). Comme \mathcal{M}_n est un s.e.v. fermé de \mathcal{H} , $\mathcal{M}_{-\infty}$ est un s.e.v. fermé de \mathcal{H} . Le théorème de projection 2.4.1 prouve que $\text{proj} (h | \mathcal{M}_{-\infty})$ existe. Pour $m < n$, définissons $\mathcal{M}_n \ominus \mathcal{M}_m$ le complément orthogonal de \mathcal{M}_m dans \mathcal{M}_n , c'est à dire $\mathcal{M}_m^\perp \cap \mathcal{M}_n$. Cet ensemble est un s.e.v fermé de \mathcal{H} d'après le théorème 2.4.4. Notons que d'après la propriété 7 de la proposition 2.4.3,

$$\text{proj} (h | \mathcal{M}_n \ominus \mathcal{M}_m) = \text{proj} (h | \mathcal{M}_n) - \text{proj} (h | \mathcal{M}_m) .$$

Il s'en suit que, pour tout $m \geq 1$,

$$\sum_{n=-m+1}^0 \|\text{proj}(h|\mathcal{M}_n \ominus \mathcal{M}_{n-1})\|^2 = \|\text{proj}(h|\mathcal{M}_0 \ominus \mathcal{M}_{-m})\|^2 \leq \|h\|^2 < \infty.$$

On obtient que la série de termes positifs $(\|\text{proj}(h|\mathcal{M}_n \ominus \mathcal{M}_{n-1})\|^2)_{n \leq 0}$ est convergente et comme pour tout $m \leq p \leq 0$,

$$\|\text{proj}(h|\mathcal{M}_p) - \text{proj}(h|\mathcal{M}_n)\|^2 = \sum_{n=-m+1}^p \|\text{proj}(h|\mathcal{M}_n \ominus \mathcal{M}_{n-1})\|^2,$$

on voit que la suite $\{\text{proj}(h|\mathcal{M}_n), n = 0, -1, -2, \dots\}$ est une suite de Cauchy. Comme \mathcal{H} est complet, $\text{proj}(h|\mathcal{M}_n)$ converge dans \mathcal{H} . Notons z sa limite. Il reste à prouver que $z = \text{proj}(h|\mathcal{M}_{-\infty})$. En appliquant le théorème de projection 2.4.1, nous devons donc démontrer que $z \in \mathcal{M}_{-\infty}$ et $h - z \perp \mathcal{M}_{-\infty}$. Comme $\text{proj}(h|\mathcal{M}_n) \in \mathcal{M}_p$ pour tout $n \leq p$, nous avons donc $z \in \mathcal{M}_p$ pour tout p et donc $z \in \mathcal{M}_{-\infty}$. Prenons maintenant $p \in \mathcal{M}_{-\infty}$. Alors $p \in \mathcal{M}_n$ pour tout $n \in \mathbb{Z}$, et donc, pour tout $n \in \mathbb{Z}$, $\langle h - \text{proj}(h|\mathcal{M}_n), p \rangle = 0$ et $\langle h - z, p \rangle = 0$ en passant à la limite, ce qui conclut la preuve. \square

Corollaire 2.4.6. *Soit $\{e_k, k \in \mathbb{N}\}$ une famille orthonormale d'un espace de Hilbert \mathcal{H} . On note $\mathcal{E}_{\infty} = \text{Vect}(e_k, k \in \mathbb{N})$. Alors*

$$\text{proj}(h|\mathcal{E}_{\infty}) = \sum_{k=0}^{\infty} \langle h, e_k \rangle e_k.$$

La preuve de ce corollaire est laissée à titre d'exercice.

2.5 Isométries et isomorphismes d'espaces de Hilbert

Définition 2.5.1 (Isométrie). Soient \mathcal{H} et \mathcal{I} deux espaces de Hilbert complexes et \mathcal{G} un sous-espace vectoriel de \mathcal{H} . Une *isométrie* S de \mathcal{G} dans \mathcal{I} est une application linéaire $S : \mathcal{G} \rightarrow \mathcal{I}$ telle que $\langle Sv, Sw \rangle_{\mathcal{I}} = \langle v, w \rangle_{\mathcal{H}}$ pour tout $(v, w) \in \mathcal{G}$.

On peut montrer qu'une isométrie entre deux espaces de Hilbert est nécessairement linéaire. On remarque aussi qu'une isométrie est toujours une application continue.

Définition 2.5.2 (isomorphisme d'espaces de Hilbert). Un espace de Hilbert \mathcal{H} est *isomorphe* à un espace de Hilbert \mathcal{I} s'il existe une isométrie bijective T de \mathcal{H} dans \mathcal{I} .

Théorème 2.5.3. *Soit \mathcal{H} un espace de Hilbert séparable.*

a) *Si \mathcal{H} est infini dimensionnel, alors il est isomorphe à ℓ^2 .*

b) Si \mathcal{H} est de dimension finie, alors \mathcal{H} est isomorphe à \mathbb{C}^n .

Preuve. Soit (e_i) une suite orthonormale complète de \mathcal{H} . Si \mathcal{H} est infini-dimensionnel, alors (e_i) est une suite infinie. Soit $x \in \mathcal{H}$. Pour $x \in \mathcal{H}$, définissons $Tx = (\alpha_i)$, où $\alpha_i = \langle x, e_i \rangle$. Le théorème 2.2.3 montre que T est une isométrie de \mathcal{H} dans ℓ^2 . \square

Comme tous les espaces de Hilbert infini-dimensionnels sont isomorphes à l'espace des suites ℓ^2 , deux espaces de Hilbert infini-dimensionnels séparables quelconques sont isomorphes.

Le résultat suivant permet de construire facilement des isométries.

Théorème 2.5.4. Soit $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ et $(\mathcal{I}, \langle \cdot, \cdot \rangle_{\mathcal{I}})$ deux espaces de Hilbert complexes. Soit \mathcal{G} un sous-espace vectoriel de \mathcal{H} .

- a) soit $S : \mathcal{G} \rightarrow \mathcal{I}$ une isométrie de \mathcal{G} dans \mathcal{I} . Alors, S se prolonge de façon unique en une isométrie $\bar{S} : \bar{\mathcal{G}} \rightarrow \mathcal{I}$ et $\bar{S}(\bar{\mathcal{G}})$ est l'adhérence de $S(\mathcal{G})$ dans \mathcal{I} .
- b) Soit $(v_t, t \in T)$ et $(w_t, t \in T)$ deux familles de vecteurs de \mathcal{H} et \mathcal{I} indexées par un ensemble d'indices T quelconque. Supposons que pour tout $(s, t) \in T \times T$, $\langle v_t, v_s \rangle_{\mathcal{H}} = \langle w_t, w_s \rangle_{\mathcal{I}}$. Alors, il existe une unique isométrie $S : \text{Vect}(v_t, t \in T) \rightarrow \text{Vect}(w_t, t \in T)$ telle que pour tout $t \in T$, $Sv_t = w_t$. De plus, on a $S(\overline{\text{Vect}(v_t, t \in T)}) = \overline{\text{Vect}(w_t, t \in T)}$.

Dans la suite, nous utiliserons la même notation pour S et son prolongement \bar{S} .

Preuve. Soit $v \in \bar{\mathcal{G}}$. Pour toute suite $(v_n) \in \mathcal{G}$ convergeant vers v , la suite (Sv_n) est une suite de Cauchy dans \mathcal{I} (car (v_n) est une suite de Cauchy dans \mathcal{G} et S est isométrique). Il existe donc $w \in \mathcal{I}$ telle que $w = \lim_{n \rightarrow \infty} Sv_n$. Si (v'_n) est une autre suite convergeant vers v , nous avons $\|v'_n - v_n\|_{\mathcal{H}} \rightarrow 0$ et, comme S est isométrique, $\|Sv_n - Sv'_n\|_{\mathcal{I}} \rightarrow 0$, ce qui montre que la limite w ne dépend pas du choix de la suite. Posons $\bar{S}v = w$. Les propriétés de linéarité et de conservation du produit scalaire sont conservées par passage à la limite et $\bar{S} : \bar{\mathcal{G}} \rightarrow \mathcal{I}$ est une isométrie prolongeant S .

Par construction $\bar{S}(\bar{\mathcal{G}})$ est inclus dans l'adhérence de $S(\mathcal{G})$. Inversement, soit $w \in \overline{S(\mathcal{G})}$. Il existe une suite $(v_n) \in \mathcal{G}$ telle que $w = \lim_{n \rightarrow \infty} Sv_n$. La suite (Sv_n) est de Cauchy et comme S est une isométrie, la suite (v_n) est aussi de Cauchy dans \mathcal{G} . Soit $v \in \bar{\mathcal{G}}$ sa limite. Nous avons $\bar{S}v = \lim_{n \rightarrow \infty} Sv_n$ et donc $\bar{S}v = w$, ce qui montre $\overline{S(\mathcal{G})} \subseteq \bar{S}(\bar{\mathcal{G}})$. Ceci établit le point (a) de la proposition.

Pour toute partie finie J de T et pour tous coefficients complexes $(a_t)_{t \in J}$ et $(b_t)_{t \in J}$, nous avons

$$\sum_{t \in J} a_t v_t = \sum_{t \in J} b_t v_t \Rightarrow \sum_{t \in J} a_t w_t = \sum_{t \in J} b_t w_t$$

puisqu'en posant $c_t = a_t - b_t$,

$$\left\| \sum_{t \in J} c_t v_t \right\|_{\mathcal{H}}^2 = \sum_{t \in J} \sum_{t' \in J} c_t \overline{c_{t'}} \langle v_t, v_{t'} \rangle_{\mathcal{H}} = \sum_{t \in J} \sum_{t' \in J} c_t \overline{c_{t'}} \langle w_t, w_{t'} \rangle_{\mathcal{I}} = \left\| \sum_{t \in I} c_t w_t \right\|_{\mathcal{I}}^2,$$

par linéarité et conservation du produit scalaire. Ceci permet de définir $Sf = \sum_{t \in I} a_t w_t$ pour tout f tel que $f = \sum_{t \in I} a_t v_t$ avec I partie finie de \mathbb{T} . Nous avons donc défini S sur $\mathcal{G} = \text{Vect}(v_t, t \in \mathbb{T})$ et c'est une isométrie. Cette isométrie, en vertu de (a)) se prolonge de façon unique en une isométrie $\tilde{S} : \tilde{\mathcal{G}} \rightarrow \mathcal{I}$ telle que $\tilde{S}(\tilde{\mathcal{G}}) = \overline{S(\mathcal{G})}$. Par construction, $\tilde{\mathcal{G}} = \overline{\text{Vect}(v_t, t \in \mathbb{T})}$ et $S(\mathcal{G}) = \text{Vect}(w_t, t \in \mathbb{T})$. \square

Chapitre 3

Processus stationnaires au second ordre

3.1 Processus du second ordre

On a vu dans l'exemple 2.1.5 que l'espace $L^2(\Omega)$ des variables aléatoires (v.a.) de variance finie est un espace de Hilbert. Pour profiter des propriétés de ces espaces il est donc naturel de travailler sur des processus faisant intervenir des v.a. de cet espace.

Définition 3.1.1 (Processus du second ordre). Le processus $X = (X_t)_{t \in T}$ à valeurs dans \mathbb{C}^d est dit du second ordre, si $\mathbb{E}[|X_t|^2] < \infty$ pour tout $t \in T$, où $|x|$ est la norme hermitienne de $x \in \mathbb{C}^d$.

Notons que la *fonction moyenne* définie sur T par $\mu(t) = \mathbb{E}[X_t]$ est à valeurs dans \mathbb{C}^d et que la *fonction d'autocovariance* définie sur $T \times T$ par

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

Elle prend ses valeurs dans l'espace des matrices de dimension $d \times d$. Pour tout $s \in T$, $\Gamma(s, s)$ est une matrice d'autocovariance. C'est donc une matrice hermitienne positive. Plus généralement, toute fonction d'autocovariance vérifie les propriétés suivantes.

Proposition 3.1.2. Soit Γ la fonction d'autocovariance d'un processus du second ordre indexé par T à valeurs dans \mathbb{C}^d . Elle vérifie alors les propriétés suivantes.

1. *Symétrie hermitienne* : pour tout $s, t \in T$,

$$\Gamma(s, t) = \Gamma(t, s)^H \quad (3.1)$$

2. *Type positif* : pour tout $n \geq 1$, pour tout $t_1, \dots, t_n \in T$ et pour tout $a_1, \dots, a_n \in \mathbb{C}^d$,

$$\sum_{1 \leq k, m \leq n} a_k^H \Gamma(t_k, t_m) a_m \geq 0 \quad (3.2)$$

Preuve. La propriété (3.1) est immédiate par définition de la covariance. Pour montrer (3.2), formons la combinaison linéaire $Y = \sum_{k=1}^n a_k^H X_{t_k}$. Y est une variable aléatoire complexe. En utilisant les propriétés de forme hermitienne de la covariance, on obtient

$$\mathbb{V}(Y) = \sum_{1 \leq k, m \leq n} a_k^H \Gamma(t_k, t_m) a_m$$

ce qui établit (3.2). □

Dans le cas scalaire ($d = 1$), on note en général $\gamma(s, t)$ la covariance, en réservant la notation $\Gamma(s, t)$ au cas des processus vectoriels ($d > 1$).

3.2 Covariance d'un processus stationnaire au second ordre

Dorénavant, dans ce chapitre, on prend $T = \mathbb{Z}$. On définit la stationnarité au second ordre en ne retenant que les propriétés du second ordre (moyenne et covariance) d'un processus stationnaire au sens strict indexé par \mathbb{Z} . En effet, soit $X = (X_t)_{t \in \mathbb{Z}}$ un processus stationnaire au sens strict à valeurs dans \mathbb{C}^d . Supposons de plus qu'il est du second ordre. Alors sa fonction moyenne est constante puisque la loi marginale l'est, et sa fonction d'autocovariance Γ vérifie $\Gamma(s, t) = \Gamma(s - t, 0)$ pour tout $s, t \in \mathbb{Z}$ puisque les lois bi-dimensionnelles sont invariantes par translation. Cela donne la définition suivante.

Définition 3.2.1 (Stationnarité au second ordre). Soit $\mu \in \mathbb{C}^d$ et $\Gamma : \mathbb{Z} \rightarrow \mathbb{C}^{d \times d}$. Un processus $(X_t)_{t \in \mathbb{Z}}$ à valeurs dans \mathbb{C}^d est dit *stationnaire au second ordre* (ou *faiblement stationnaire*) de moyenne μ et de fonction d'auto-covariance Γ si :

- a) X est un processus du second ordre, i.e. $\mathbb{E}[|X_t|^2] < +\infty$,
- b) pour tout $t \in \mathbb{Z}$, $\mathbb{E}[X_t] = \mu$,
- c) pour tout couple $(s, t) \in \mathbb{Z} \times \mathbb{Z}$, $\text{cov}(X_s, X_t) = \Gamma(s - t)$.

Par convention la fonction d'autocovariance d'un processus stationnaire au second ordre indexé par T est définie sur T au lieu de $T \times T$ pour le cas général.

Comme expliqué en préambule de la définition, un processus du second ordre stationnaire au sens strict est stationnaire au second ordre. L'implication inverse est vraie pour la classe des processus gaussiens définies au paragraphe 1.2.3 d'après la proposition 1.2.10.

On remarque qu'un processus $(X_t)_{t \in \mathbb{Z}}$ à valeurs dans \mathbb{C}^d est stationnaire au second ordre de moyenne μ et de fonction d'auto-covariance Γ si et seulement si pour tout $\lambda \in \mathbb{C}^d$, le processus $(\lambda^H X_t)_{t \in \mathbb{Z}}$ à valeurs dans \mathbb{C} est stationnaire au second ordre de moyenne $\lambda^H \mu$ et de fonction d'auto-covariance $\lambda^H \Gamma \lambda$. L'étude des processus stationnaires au second ordre peut donc se restreindre au cas $d = 1$ sans grande perte de généralité.

3.2.1 Propriétés

Les propriétés de la proposition 3.1.2 se déclinent pour un processus stationnaire au second ordre de la façon suivante.

Proposition 3.2.2. *La fonction d'autocovariance $\gamma : \mathbb{Z} \rightarrow \mathbb{C}$ d'un processus stationnaire au second ordre à valeurs complexes vérifie les propriétés suivantes qui sont une conséquence directe de la proposition 3.1.2.*

1. *Symétrie hermitienne* : Pour tout $s \in \mathbb{Z}$,

$$\gamma(-s) = \overline{\gamma(s)}$$

2. *Type positif* : Pour tout entier $n \geq 1$ et tout vecteur (a_1, \dots, a_n) de valeurs complexes,

$$\sum_{s=1}^n \sum_{t=1}^n \overline{a_s} \gamma(s-t) a_t \geq 0$$

La matrice de covariance de n valeurs consécutives X_1, \dots, X_n du processus possède de plus une structure particulière, dite de *Toeplitz*, caractérisée par le fait que $(\Gamma_n)_{ij} = \gamma(i-j)$. On obtient une matrice de la forme

$$\begin{aligned} \Gamma_n &= \text{cov}([X_1 \ \dots \ X_n]^T) \\ &= \begin{bmatrix} \gamma(0) & \gamma(-1) & \dots & \gamma(1-n) \\ \gamma(1) & \gamma(0) & \dots & \gamma(2-n) \\ \vdots & & & \\ \gamma(n-1) & \gamma(n-2) & \dots & \gamma(0) \end{bmatrix} \end{aligned} \quad (3.3)$$

Lorsque $\gamma(0)$ est non-nul il peut être pratique de normaliser la fonction d'autocovariance. On obtient la définition suivante.

Définition 3.2.3 (Fonction d'autocorrélation). Pour un processus stationnaire au second ordre de variance non nulle, on appelle fonction d'autocorrélation ρ la fonction définie sur $s \in \mathbb{Z}$ par $\rho(s) = \gamma(s)/\gamma(0)$. Il s'agit d'une quantité normalisée dans le sens où $\rho(0) = 1$ et $|\rho(s)| \leq 1$ pour tout $s \in \mathbb{Z}$.

En effet, l'inégalité de Cauchy-Schwarz (voir le théorème 2.1.6) appliquée à γ implique

$$|\gamma(s)| = |\text{cov}(X_s, X_0)| \leq \sqrt{\mathbb{V}(X_s) \mathbb{V}(X_0)} = \gamma(0)$$

la dernière inégalité découlant de l'hypothèse de stationnarité.

Exemple 3.2.4 (Retournement du temps (suite)). Soit $(X_t)_{t \in \mathbb{Z}}$ un processus aléatoire stationnaire au second ordre à valeurs réelles de moyenne μ_X et de fonction d'autocovariance γ_X . On note, pour tout $t \in \mathbb{Z}$, $Y_t = X_{-t}$ le processus *retourné*, comme dans l'exemple 1.3.9. Alors Y_t est un processus stationnaire au second ordre de même moyenne et de même fonction d'autocovariance que le processus X_t . En effet on a :

$$\begin{aligned} \mathbb{E}[Y_t] &= \mathbb{E}[X_{-t}] = \mu_X \\ \text{cov}(Y_{t+h}, Y_t) &= \text{cov}(X_{-t-h}, X_{-t}) = \gamma_X(-h) = \gamma_X(h) \end{aligned}$$

Définition 3.2.5 (Bruit blanc faible). On appelle bruit blanc faible un processus aléatoire stationnaire au second ordre à valeurs complexes ou réelles, centré, de fonction d'autocovariance γ définie par $\gamma(0) = \sigma^2 > 0$ et $\gamma(s) = 0$ pour tout $s \neq 0$. On le notera $(X_t) \sim \text{BB}(0, \sigma^2)$.

Définition 3.2.6 (Bruit blanc fort). On appelle bruit blanc fort une suite de variables aléatoires (X_t) , centrées, indépendantes et identiquement distribuées (i.i.d.) de variance $\mathbb{E}[X_t^2] = \sigma^2 < \infty$. On le notera $(X_t) \sim \text{IID}(0, \sigma^2)$.

Par définition un bruit blanc fort est un bruit blanc faible. La structure de bruit blanc fort est clairement plus contraignante que celle du bruit blanc faible. Notons que, de même que la stationnarité stricte d'un processus gaussien découle de la stationnarité faible, un bruit blanc faible gaussien est un bruit blanc fort.

Exemple 3.2.7 (Processus MA(1)). Soit (X_t) le processus stationnaire au second ordre défini par :

$$X_t = Z_t + \theta Z_{t-1}, \quad (3.4)$$

où $(Z_t) \sim \text{BB}(0, \sigma^2)$ réel et $\theta \in \mathbb{R}$. On vérifie aisément que $\mathbb{E}[X_t] = 0$ et que sa fonction d'autocovariance est définie par

$$\gamma(s) = \begin{cases} \sigma^2(1 + \theta^2) & \text{si } s = 0, \\ \sigma^2\theta & \text{si } s = \pm 1, \\ 0 & \text{sinon.} \end{cases} \quad (3.5)$$

Le processus (X_t) est donc bien stationnaire au second ordre. Un tel processus est appelé *processus à moyenne ajustée d'ordre 1*. Cette propriété se généralise, sans difficulté, à un processus MA(q). Nous reviendrons plus en détail, paragraphe 4.2, sur la définition et les propriétés de ces processus.

Exemple 3.2.8 (Processus harmonique réel). Soient $(A_k)_{1 \leq k \leq N}$ N v.a. réelles de variance finie. On note $\sigma_k^2 = \mathbb{V}(A_k)$. Soient $(\Phi_k)_{1 \leq k \leq N}$, N variables aléatoires indépendantes et identiquement distribuées (i.i.d.), de loi uniforme sur $[-\pi, \pi]$, et indépendantes de $(A_k)_{1 \leq k \leq N}$. On définit :

$$X_t = \sum_{k=1}^N A_k \cos(\lambda_k t + \Phi_k), \quad (3.6)$$

où $(\lambda_k)_{1 \leq k \leq N} \in [-\pi, \pi]$ sont N pulsations. Le processus (X_t) est appelé processus harmonique. On vérifie aisément que $\mathbb{E}[X_t] = 0$ et que, pour tout $s, t \in \mathbb{Z}$,

$$\mathbb{E}[X_s X_t] = \frac{1}{2} \sum_{k=1}^N \sigma_k^2 \cos(\lambda_k(s - t)).$$

Le processus harmonique est donc stationnaire au second ordre.

Exemple 3.2.9 (Marche aléatoire). Soit (S_t) le processus défini sur $t \in \mathbb{N}$ par $S_t = X_0 + X_1 + \dots + X_t$, où (X_t) est un bruit blanc fort réel. Un tel processus est appelé *une marche aléatoire*. On en déduit que $\mathbb{E}[S_t] = 0$, $\mathbb{E}[S_t^2] = t\sigma^2$ et, pour $s \leq t \in \mathbb{N}$, on a :

$$\mathbb{E}[S_s S_t] = \mathbb{E}[(S_s + X_{s+1} + \dots + X_t)S_s] = s\sigma^2$$

Le processus (S_t) n'est donc pas stationnaire au second ordre.

Exemple 3.2.10. Nous allons montrer que la fonction χ définie sur \mathbb{Z} , par

$$\chi(s) = \begin{cases} 1 & \text{si } s = 0, \\ \rho & \text{si } s = \pm 1, \\ 0 & \text{sinon.} \end{cases} \quad (3.7)$$

est la fonction d'autocovariance d'un processus stationnaire au second ordre réel si et seulement si $\rho \in [-1/2, 1/2]$. Nous avons déjà montré exemple 3.2.7 que la fonction d'autocovariance γ d'un processus MA(1) est donnée par (3.5). La fonction χ est donc la fonction d'autocovariance d'un processus MA(1) si et seulement si $\sigma^2(1 + \theta^2) = 1$ et $\sigma^2\theta = \rho$. Lorsque $|\rho| \leq 1/2$, ce système d'équations admet comme solution :

$$\theta = (2\rho)^{-1}(1 \pm \sqrt{1 - 4\rho^2}) \quad \text{et} \quad \sigma^2 = (1 + \theta^2)^{-1}.$$

Lorsque $|\rho| > 1/2$, ce système d'équations n'admet pas de solution réelles et la fonction χ n'est donc pas la fonction d'autocovariance d'un processus MA(1). Plus généralement, on vérifiera dans l'exercice 3.4.4 que si $|\rho| > 1/2$, χ n'est en fait pas de type positif et n'est donc pas une fonction de covariance, voir 3.2.2.

3.2.2 Interprétation de la fonction d'autocovariance

Dans les exemples précédents, nous avons été amenés à évaluer la fonction d'autocovariance de processus pour quelques exemples simples de séries temporelles. Dans la plupart des problèmes d'intérêt pratique, nous ne parlons pas de modèles de série temporelle définis *a priori*, mais d'*observations*, $\{x_1, \dots, x_n\}$ associées à une *réalisation* du processus. Afin de comprendre la structure de dépendance entre les différentes observations, nous serons amenés à *estimer* la loi du processus, ou du moins des caractéristiques de ces lois. Pour un processus stationnaire au second ordre, nous pourrions, à titre d'exemple, estimer sa moyenne par la *moyenne empirique* :

$$\hat{\mu}_n = n^{-1} \sum_{k=1}^n x_k$$

et les fonctions d'autocovariance et d'autocorrélation par les fonctions d'autocorrélation et d'autocovariance *empiriques*

$$\hat{\gamma}(h) = n^{-1} \sum_{k=1}^{n-|h|} (x_k - \hat{\mu}_n)(x_{k+|h|} - \hat{\mu}_n) \quad \text{et} \quad \hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0) .$$

Lorsqu'il est *a priori* raisonnable de penser que la série considérée est stationnaire au second ordre, la moyenne empirique, la fonction d'autocovariance empirique et la fonction d'autocorrélation empirique sont de "bons" estimateurs, dans un sens que nous préciserons dans les chapitres 9 et 10. L'analyse de la fonction d'autocovariance empirique est un élément permettant de guider le choix d'un modèle approprié pour les observations. Par exemple, le fait que la fonction d'autocovariance empirique soit *proche* de zéro pour tout $h \neq 0$ (proximité qu'il faudra définir dans un sens statistique précis) indique par exemple qu'un bruit blanc est un modèle adéquat pour les données. La figure 3.1 représente les 100 premières valeurs de la fonction d'autocorrélation empirique de la série des battements cardiaques représentée figure 1.1. On observe que cette série est *positivement corrélée* c'est-à-dire que les fonctions coefficients d'autocorrélation sont positifs et significativement non nuls. Nous avons, à titre de comparaison, représenté aussi la fonction d'autocorrélation empirique d'une trajectoire de même longueur d'un bruit blanc gaussien. Une forte corrélation peut être interprétée comme l'indice d'une dépendance

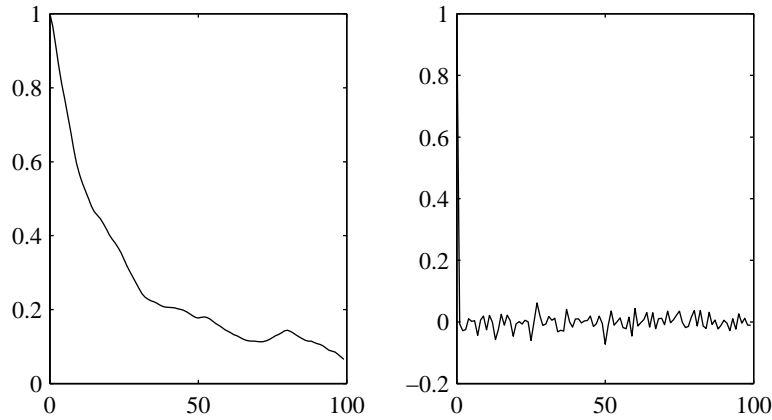


Fig. 3.1 Courbe de gauche : fonction d'autocorrélation empirique de la série des battements cardiaques (figure 1.1). Courbe de droite : fonction d'autocorrélation empirique d'une trajectoire de même longueur d'un bruit blanc gaussien.

linéaire. Ainsi la figure 3.2 montre que le fait que $\hat{\rho}(1) = 0.966$ pour la

série des battements cardiaques se traduit par une très forte prédictabilité de X_{t+1} en fonction de X_t (les couples de points successifs s'alignent quasiment sur une droite). Nous montrerons au chapitre 5, que dans un tel contexte, $\mathbb{E}[(X_{t+1} - \mu) - \rho(1)(X_t - \mu)] = (1 - \rho^2)\text{cov}(X_t)$, c'est-à-dire, compte tenu de la valeur estimée pour $\rho(1)$, que la variance de "l'erreur de prédiction" $X_{t+1} - [\mu + \rho(1)(X_t - \mu)]$ est 15 fois plus faible que celle du signal original.

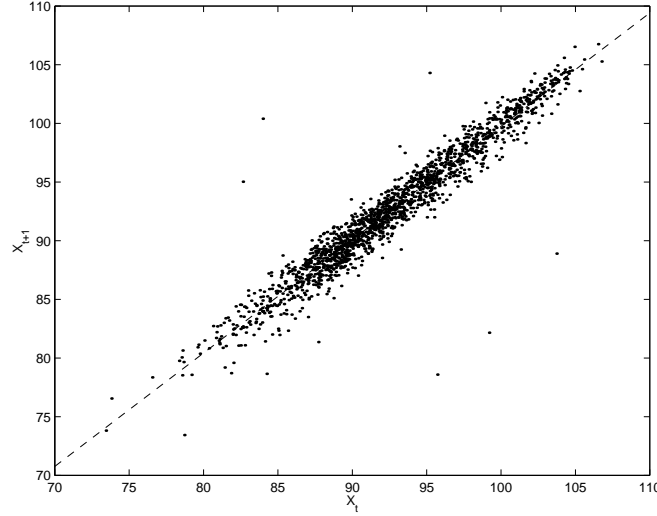


Fig. 3.2 X_{t+1} en fonction de X_t pour la série des battements cardiaques de la figure 1.1). Les tirets représentent la meilleure droite de régression linéaire de X_{t+1} sur X_t .

L'indice S&P500 tracé (fig. 1.4) présente un cas de figure plus difficile, d'une part parce que la série de départ ne saurait être tenue pour stationnaire et qu'il nous faudra considérer la série des évolutions journalières ; d'autre part, parce que selon le choix de la transformation des données considérées, la série transformée présente ou non des effets de corrélation. On définit tout d'abord les *log-retours* de l'indice S&P500 comme les différences des logarithmes de l'indice à deux dates successives :

$$X_t = \log(S_t) - \log(S_{t-1}) = \log\left(1 + \frac{S_t - S_{t-1}}{S_{t-1}}\right)$$

La série des log-retours de la série S&P 500 est représentée dans la figure 3.3. Les coefficients d'autocorrélation empiriques de la série des log-retours sont représentés dans la figure 3.4. On remarque qu'ils sont approximativement nuls pour $h \neq 0$ ce qui suggère de modéliser la série des log-retours par un bruit blanc faible. Il est intéressant d'étudier aussi la série des log-retours

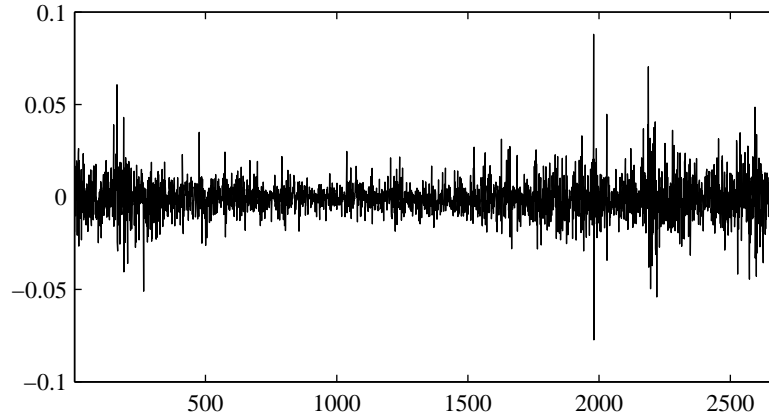


Fig. 3.3 Log-Retours de la série S&P 500 (figure 1.4).

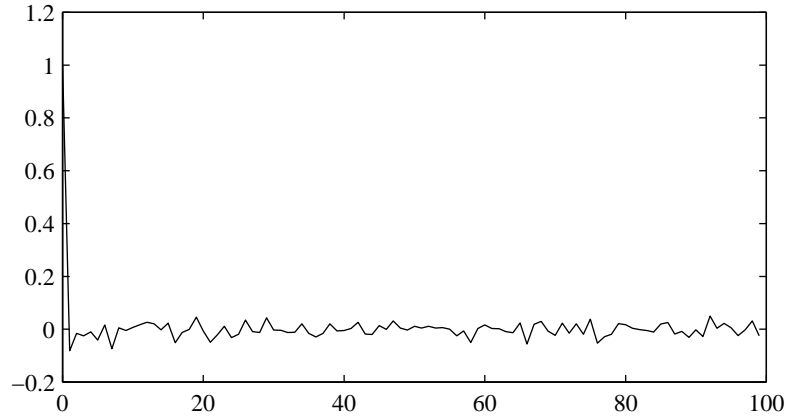


Fig. 3.4 Fonction d'autocorrélation empirique de la série des log-retours de l'indice S&P 500.

absolus, $A(t) = |X_t|$. On peut, de la même façon, déterminer la suite des coefficients d'autocorrélation empirique de cette série, qui est représentée dans la figure 3.5. On voit, qu'à l'inverse de la série des log-retours, la série des valeurs absolues des log-retours est positivement corrélée, les valeurs d'autocorrélation étant significativement non nulles pour $|h| \leq 100$. On en déduit, en particulier, que la suite des log-retours peut être modélisée comme un bruit blanc, mais pas un bruit blanc fort : en effet, pour un bruit blanc fort X_t , nous avons, pour toute fonction f telle que $\mathbb{E}[f(X_t)^2] = \sigma_f^2 < \infty$,

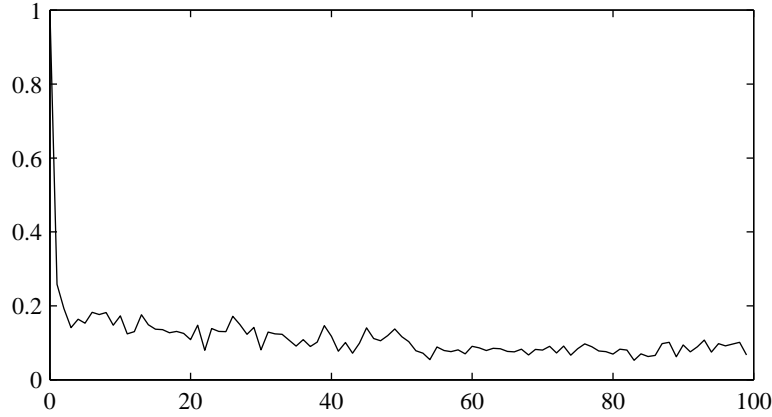


Fig. 3.5 *Fonction d'autocorrélation empirique de la série des valeurs absolues des log-retours de l'indice S&P 500.*

$\text{cov}(f(X_{t+h}), f(X_t)) = 0$ pour $h \neq 0$ (les variables $f(X_{t+h})$ et $f(X_t)$ étant indépendantes, elles sont a fortiori non corrélées).

3.3 Mesure spectrale d'un processus stationnaire

Dans toute la suite, \mathbb{T} désigne le tore $(-\pi, \pi]$ et $\mathcal{B}(\mathbb{T})$ la tribu borélienne associée. Le théorème d'Herglotz ci dessous établit l'équivalence entre la fonction d'autocovariance et une mesure finie définie sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$. Cette mesure, appelée *mesure spectrale du processus*, joue un rôle analogue à celui de la transformation de Fourier pour les fonctions de carré intégrable.

Théorème 3.3.1 (Herglotz). *Une suite $(\gamma(h))_{h \in \mathbb{Z}}$ est de type positif si et seulement si il existe une unique mesure positive ν sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ telle que :*

$$\gamma(h) = \int_{\mathbb{T}} e^{ih\lambda} \nu(d\lambda), \quad \forall h \in \mathbb{Z}. \quad (3.8)$$

Remarque 3.3.2. Lorsque γ est la fonction d'autocovariance d'un processus stationnaire au second ordre, on sait d'après la proposition 3.2.2 que $\{\gamma(h)\}_{h \in \mathbb{Z}}$ est de type positif. Les hypothèses du théorème de Herglotz sont donc vérifiées et dans ce cas la mesure ν est appelée la *mesure spectrale* du processus. Si la mesure ν possède une densité f par rapport à la mesure de Lebesgue sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ alors f est appelée la *densité spectrale de puissance* du processus.

Preuve. Si $\gamma(n)$ a la représentation (3.8), montrons que $\gamma(n)$ est de type positif. En effet, pour tout n et toute suite $\{a_k \in \mathbb{C}\}_{1 \leq k \leq n}$,

$$\sum_{k,m} a_k \overline{a_m} \gamma(k-m) = \int_{\mathbb{T}} \sum_{k,m} a_k \overline{a_m} e^{ik\lambda} e^{-im\lambda} \nu(d\lambda) = \int_{\mathbb{T}} \left| \sum_k a_k e^{ik\lambda} \right|^2 \nu(d\lambda) \geq 0.$$

Réciproquement, supposons que $\gamma(n)$ soit une suite de type positif et considérons la suite de fonctions indexée par n :

$$f_n(\lambda) = \frac{1}{2\pi n} \sum_{k=1}^n \sum_{m=1}^n \gamma(k-m) e^{-ik\lambda} e^{im\lambda} = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) \gamma(k) e^{-ik\lambda}.$$

γ étant de type positif, $f_n(\lambda) \geq 0$, pour tout $\lambda \in \mathbb{T}$. Notons ν_n la mesure (positive) de densité f_n par rapport à la mesure de Lebesgue sur \mathbb{T} . On a alors

$$\begin{aligned} \int_{\mathbb{T}} e^{ih\lambda} \nu_n(d\lambda) &= \int_{\mathbb{T}} e^{ih\lambda} f_n(\lambda) d\lambda = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) \gamma(k) \int_{\mathbb{T}} e^{i(h-k)\lambda} d\lambda \\ &= \begin{cases} \left(1 - \frac{|h|}{n}\right) \gamma(h), & \text{si } |h| < n, \\ 0, & \text{sinon.} \end{cases} \quad (3.9) \end{aligned}$$

Quitte à renormaliser ν_n pour en faire une mesure de probabilité, le théorème 8.2.2 implique qu'il existe une mesure positive ν et une sous-suite ν_{n_k} de ν_n telle que

$$\int_{\mathbb{T}} e^{ih\lambda} \nu_{n_k}(d\lambda) \longrightarrow \int_{\mathbb{T}} e^{ih\lambda} \nu(d\lambda), \text{ lorsque } k \rightarrow \infty.$$

En remplaçant n par n_k dans (3.9) et en faisant tendre k vers l'infini, on a

$$\gamma(h) = \int_{\mathbb{T}} e^{ih\lambda} \nu(d\lambda), \quad \forall h \in \mathbb{Z}.$$

Montrons à présent que ν est unique. En effet, s'il existait une autre mesure μ telle que pour tout $h \in \mathbb{Z}$: $\int_{\mathbb{T}} e^{ih\lambda} \nu(d\lambda) = \int_{\mathbb{T}} e^{ih\lambda} \mu(d\lambda)$ alors d'après le lemme 2.3.3, $\int_{\mathbb{T}} g(\lambda) \nu(d\lambda) = \int_{\mathbb{T}} g(\lambda) \mu(d\lambda)$ pour toute fonction continue g telle que $g(\pi) = g(-\pi)$. On en déduit donc que $\nu = \mu$. \square

Corollaire 3.3.3 (Corollaire du théorème d'Herglotz). *Une suite $(\gamma(h))_{h \in \mathbb{Z}}$ à valeurs complexes telle que $\sum_{h \in \mathbb{Z}} |\gamma(h)|^2 < \infty$ est de type positif si et seulement si la fonction définie par*

$$f(\lambda) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma(h) e^{-ih\lambda}$$

est positive pour tout $\lambda \in \mathbb{T}$.

Preuve. D'après le théorème de Herglotz (Théorème 3.3.1), $(\gamma(h))_{h \in \mathbb{Z}}$ est de type positif si et seulement si il existe une mesure positive ν sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ telle que :

$$\gamma(h) = \int_{\mathbb{T}} e^{ih\lambda} \nu(d\lambda) .$$

D'après le théorème 2.2.3 et le corollaire 2.3.5, comme $\sum_{h \in \mathbb{Z}} |\gamma(h)|^2 < \infty$, on peut considérer la série de Fourier associée convergente dans $L^2(\mathbb{T}, \lambda^{\text{Leb}})$: $(2\pi)^{-1} \sum_{k \in \mathbb{Z}} \gamma(k) e^{-ik\lambda} \stackrel{\text{def}}{=} f(\lambda)$. Ainsi, $\gamma(h) = \int_{\mathbb{T}} e^{ih\lambda} f(\lambda) d\lambda$ et donc la positivité de ν revient à la positivité de f , ce qui conclut la preuve. \square

Exemple 3.3.4. En reprenant l'exemple 3.2.10, on vérifie immédiatement que $(\chi(h))$ est de module sommable et que :

$$f(\lambda) = \frac{1}{2\pi} \sum_h \chi(h) e^{-ih\lambda} = \frac{1}{2\pi} (1 + 2\rho \cos(\lambda))$$

et donc que la séquence est une fonction d'autocovariance uniquement lorsque $|\rho| \leq 1/2$.

Exemple 3.3.5 (Densité spectrale de puissance du bruit blanc). La fonction d'autocovariance d'un bruit blanc est donnée par $\gamma(h) = \sigma^2 \delta(h)$, d'où l'expression de la densité spectrale correspondante

$$f(\lambda) = \frac{\sigma^2}{2\pi}$$

La densité spectrale d'un bruit blanc est donc constante. Cette propriété est à l'origine de la terminologie "bruit blanc" qui provient de l'analogie avec le spectre de la lumière blanche constant dans toute la bande de fréquences visibles.

Exemple 3.3.6 (Densité spectrale de puissance du processus MA(1)). Le processus MA(1) introduit dans l'exemple 3.2.7 possède une séquence d'autocovariance donnée par $\gamma(0) = \sigma^2(1 + \theta^2)$, $\gamma(1) = \gamma(-1) = \sigma^2\theta$ et $\gamma(h) = 0$ sinon (cf. exemple 3.2.7). D'où l'expression de sa densité spectrale :

$$f(\lambda) = \frac{\sigma^2}{2\pi} (2\theta \cos(\lambda) + (1 + \theta^2)) = \frac{\sigma^2}{2\pi} |1 + \theta e^{-i\lambda}|^2$$

La densité spectrale d'un tel processus est représentée figure 3.6 pour $\theta = -0.9$ et $\sigma^2 = 1$ avec une échelle logarithmique (dB).

Exemple 3.3.7 (Mesure spectrale du processus harmonique). La fonction d'autocovariance du processus harmonique $X_t = \sum_{k=1}^N A_k \cos(\lambda_k t + \Phi_k)$ (voir exemple 3.2.8) est donnée par :

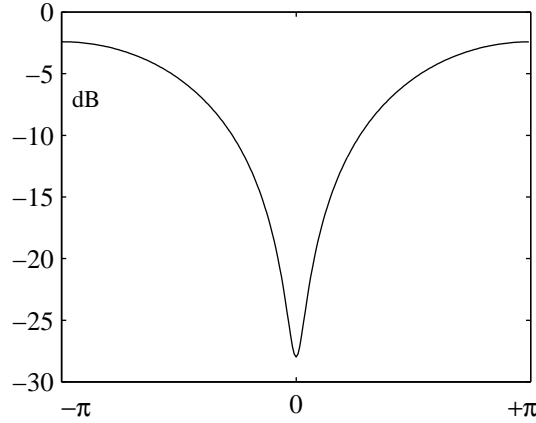


Fig. 3.6 Densité spectrale (en dB) d'un processus MA-1, défini par l'équation (3.4) pour $\sigma = 1$ et $\theta = -0.9$.

$$\gamma(h) = \frac{1}{2} \sum_{k=1}^N \sigma_k^2 \cos(\lambda_k h) \quad (3.10)$$

où $\sigma_k^2 = \mathbb{E}[A_k^2]$. Cette suite de coefficients d'autocovariance n'est pas sommable et la mesure spectrale n'admet pas de densité. En notant cependant que :

$$\cos(\lambda_k h) = \frac{1}{2} \int_{-\pi}^{\pi} e^{ih\lambda} (\delta_{\lambda_k}(\mathrm{d}\lambda) + \delta_{-\lambda_k}(\mathrm{d}\lambda))$$

où $\delta_{x_0}(\mathrm{d}\lambda)$ désigne la mesure de Dirac au point x_0 (cette mesure associe la valeur 1 à tout borélien de $[-\pi, \pi]$ contenant x_0 et la valeur 0 sinon), la mesure spectrale du processus harmonique peut s'écrire :

$$\nu(\mathrm{d}\lambda) = \frac{1}{4} \sum_{k=1}^N \sigma_k^2 \delta_{\lambda_k}(\mathrm{d}\lambda) + \frac{1}{4} \sum_{k=1}^N \sigma_k^2 \delta_{-\lambda_k}(\mathrm{d}\lambda)$$

Elle apparaît donc comme une somme de mesures de Dirac, dont les masses σ_k^2 sont localisées aux pulsations des différentes composantes harmoniques.

Contrairement aux autres exemples étudiés, le processus harmonique possède une fonction d'autocovariance, donnée par 3.10, non absolument sommable ($\gamma(h)$ ne tend pas même vers 0 pour les grandes valeurs de h). Par suite, il admet une mesure spectrale mais pas une densité spectrale. La propriété suivante, à démontrer à titre d'exercice, implique que le processus harmonique est en fait entièrement prédictible à partir de quelques-unes de ses valeurs passées.

Proposition 3.3.8. *S'il existe un rang n pour lequel la matrice de covariance Γ_n définie en (3.3) est non inversible, le processus correspondant X_t est prédictible dans le sens où il existe une combinaison linéaire a_1, \dots, a_l avec $l \leq n-1$ telle que $X_t = \sum_{k=1}^l a_k X_{t-k}$, l'égalité ayant lieu presque sûrement.*

L'expression de la fonction d'autocovariance, obtenue en (3.10) pour le processus harmonique, montre que les matrices de covariances associées s'écrivent comme la somme de $2N$ matrices complexes de rang 1. Par conséquent, les matrices Γ_n ne sont pas inversibles dès que $n > 2N$, ce qui implique que le processus harmonique est prédictible dès lors que l'on en a observé $2N$ valeurs. Ce résultat est sans surprise compte tenu du fait que les trajectoires de ce processus sont des sommes de sinusoides de fréquences $\lambda_1, \dots, \lambda_N$ dont seules les amplitudes et les phases sont aléatoires. La propriété suivante donne une condition suffisante simple pour éviter ce type de comportements "extrêmes". Cette propriété implique en particulier que, pour une fonction d'autocovariance absolument sommable (tous les exemples vus ci-dessus en dehors du processus harmoniques), les valeurs futures du processus correspondant ne sont pas prédictibles sans erreur à partir d'un ensemble fini de valeurs passées du processus. Nous reviendrons en détail sur ces problèmes de prédiction au chapitre 5.

Proposition 3.3.9. *Soit $\gamma(h)$ la fonction d'autocovariance d'un processus stationnaire au second ordre. On suppose que $\gamma(0) > 0$ et que $\gamma(h) \rightarrow 0$ quand $h \rightarrow \infty$. Alors, quel que soit n , la matrice de covariance définie en (3.3) est de rang plein et donc inversible.*

Preuve. Supposons qu'il existe une suite de valeurs complexes (a_1, \dots, a_n) non toutes nulles, telle que $\sum_{k=1}^n \sum_{m=1}^n a_k \overline{a_m} \gamma(k-m) = 0$. En notant ν_X la mesure spectrale de X_t , on peut écrire :

$$0 = \sum_{k=1}^n \sum_{m=1}^n a_k \overline{a_m} \int_{\mathbb{T}} e^{i(k-m)\lambda} \nu_X(d\lambda) = \int_{\mathbb{T}} \left| \sum_{k=1}^n a_k e^{ik\lambda} \right|^2 \nu_X(d\lambda)$$

Ce qui implique que $\left| \sum_{k=1}^n a_k e^{ik\lambda} \right|^2 = 0$ ν_X presque partout, c'est à dire que $\nu_X(\{\lambda : \left| \sum_{k=1}^n a_k e^{ik\lambda} \right|^2 \neq 0\}) = \nu_X(\mathbb{T} - Z) = 0$ où $Z = \{\lambda_1, \dots, \lambda_M : \sum_{k=1}^n a_k e^{ik\lambda_m} = 0\}$ désigne l'ensemble fini ($M < n$) des racines $x \in \mathbb{T}$ du polynôme trigonométrique $\sum_{k=1}^n a_k e^{ik\lambda}$. Par conséquent, les seuls éléments de $\mathcal{B}(\mathbb{T})$, qui peuvent être de mesure non nulle pour ν_X , sont les singletons $\{\lambda_m\}$. Ce qui implique que $\nu_X = \sum_{m=1}^M a_m \delta_{\lambda_m}$ (où $a_m \geq 0$ ne peuvent être tous nuls si $\gamma(0) \neq 0$). Mais, dans ce cas, $\gamma(h) = \sum_{m=1}^M a_m e^{ih\lambda_m}$, ce qui contredit l'hypothèse que $\gamma(h)$ tend vers 0 quand n tend vers l'infini. \square

3.4 Représentation harmonique des processus

3.4.1 Champ aléatoire non corrélé et intégrale stochastique

Définition 3.4.1 (Champ aléatoire non corrélé). Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Un *champ aléatoire non-corrélé* W sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ est une application $W : A \mapsto W(A)$ de $\mathcal{B}(\mathbb{T})$ dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ telle que

1. Pour tout $A \in \mathcal{B}(\mathbb{T})$, $\mathbb{E}[W(A)] = 0$.
2. Pour tout $A, B \in \mathcal{B}(\mathbb{T})$ tels que $A \cap B = \emptyset$, les variables $W(A)$ et $W(B)$ sont décorréliées et $W(A \cup B) = W(A) + W(B)$;
3. Pour toute suite $(A_n)_{n \geq 0} \in \mathcal{B}(\mathbb{T})$ décroissante telle que $\bigcap_{n=0}^{\infty} A_n = \emptyset$, $\lim_{n \rightarrow \infty} \mathbb{V}(W(A_n)) = 0$.

Lemme 3.4.2. Soit W un champ aléatoire non corrélé. Pour $A \in \mathcal{B}(\mathbb{T})$, posons $\eta_W(A) = \mathbb{V}(W(A))$. Alors η_W est une mesure finie sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$. De plus, pour tout $A, B \in \mathcal{B}(\mathbb{T})$, $\text{Cov}(W(A), W(B)) = \eta_W(A \cap B)$.

Preuve. Pour montrer que η_W est une mesure, il suffit d'établir que η_W est additive et que, pour toute suite $(A_n)_{n \geq 0}$ décroissante d'éléments de $\mathcal{B}(\mathbb{T})$ telle que $\bigcap_{n=0}^{\infty} A_n = \emptyset$, $\lim_{n \rightarrow \infty} \eta_W(A_n) = 0$. L'additivité découle de la propriété 2 et la continuité de 3.

Notons que : $A = (A \setminus B) \cup (A \cap B)$ et $B = (B \setminus A) \cup (A \cap B)$ et que $A \setminus B$, $B \setminus A$ et $A \cap B$ sont des ensembles disjoints. En vertu de 2, les variables $W(A) = W(A \setminus B) + W(A \cap B)$, $W(B) = W(B \setminus A) + W(A \cap B)$ et $W(A \setminus B)$, $W(B \setminus A)$ et $W(A \cap B)$ sont décorréliées. Par conséquent,

$$\text{Cov}(W(A), W(B)) = \mathbb{V}(W(A \cap B)) = \eta_W(A \cap B).$$

□

La mesure η_W est appelée la *base du champ aléatoire* W . Le Lemme précédent admet la réciproque suivante.

Lemme 3.4.3. Soit $W : A \mapsto W(A)$ de $\mathcal{B}(\mathbb{T})$ à valeurs dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ telle que, pour tout $A \in \mathcal{B}(\mathbb{T})$, $\mathbb{E}[W(A)] = 0$. Supposons qu'il existe une mesure η_W sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ telle que, pour tout $A, B \in \mathcal{B}(\mathbb{T})$, $\text{Cov}(W(A), W(B)) = \eta_W(A \cap B)$. Alors W est un champ aléatoire décorrélé de mesure de base η_W .

Preuve. Soit $A, B \in \mathcal{B}(\mathbb{T})$ telle que $A \cap B = \emptyset$. Nous avons :

$$\begin{aligned} & \mathbb{V}(W(A \cup B) - W(A) - W(B)) \\ &= \eta_W(A \cup B) + \eta_W(A) + \eta_W(B) - 2\eta_W(A) - 2\eta_W(B) + 2\eta_W(A \cap B) = 0, \end{aligned}$$

en utilisant $\eta_W(A \cup B) = \eta_W(A) + \eta_W(B)$ et $\eta_W(A \cap B) = 0$. Nous avons donc $W(A \cup B) = W(A) + W(B)$ et la propriété d'additivité est satisfaite.

Soit $(A_n)_{n \geq 0}$ une suite décroissantes d'éléments de $\mathcal{B}(\mathbb{T})$ telle que $\bigcap_{n=0}^{\infty} A_n = \emptyset$. Comme η_W est une mesure, nous avons $\lim_{n \rightarrow \infty} \eta_W(A_n) = 0$ ce qui implique (définition 3.4.1 (3)) $\lim_{n \rightarrow \infty} \mathbb{V}(W(A_n)) = 0$. La propriété de continuité est aussi satisfaite. \square

Exemple 3.4.4 (Champ aléatoire atomique). Soit $(Y_k)_{k \geq 0}$ une suite de variables aléatoires décorrélées de $L^2(\Omega, \mathcal{A}, \mathbb{P})$. Posons pour $k \geq 0$, $\sigma_k^2 = \mathbb{V}(Y_k)$ et supposons que $\sum_{k=0}^{\infty} \sigma_k^2 < \infty$. Soit $(\lambda_k)_{k \geq 0}$ une suite d'éléments distincts de \mathbb{T} . Considérons l'application W de $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ définie par

$$W : A \mapsto W(A) = \sum_{k=0}^{\infty} Y_k \mathbb{1}_A(\lambda_k) = \sum_{k=0}^{\infty} Y_k \delta_{\lambda_k}(A),$$

où δ_x est la mesure de Dirac. Nous avons

$$\text{Cov}(W(A), W(B)) = \sum_{k=0}^{\infty} \sigma_k^2 \mathbb{1}_A(\lambda_k) \mathbb{1}_B(\lambda_k) = \eta_W(A \cap B),$$

où

$$\eta_W(A) = \sum_{k=0}^{\infty} \sigma_k^2 \delta_{\lambda_k}(A).$$

Le lemme 3.4.3 montre que W est un champ aléatoire décorrélé de mesure de base η_W .

3.4.2 Intégrale stochastique par rapport à un champ

Théorème 3.4.5. *Soit W un champ aléatoire décorrélé sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ de base η_W . Alors il existe une unique isométrie $W : f \mapsto W(f)$ de $L^2(\mathbb{T}, \eta_W)$ à valeurs dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ telle que $W(\mathbb{1}_A) = W(A)$ pour tout $A \in \mathcal{B}(\mathbb{T})$.*

Pour tout $f \in L^2(\mathbb{T}, \eta_W)$ on a $\mathbb{E}[W(f)] = 0$ et l'image de $L^2(\mathbb{T}, \eta_W)$ par W est égale au sous-espace \mathcal{H}^W fermé de $L^2(\Omega, \mathcal{A}, \mathbb{P})$ engendré par la famille $\{W(A), A \in \mathcal{B}(\mathbb{T})\}$.

Preuve. On pose $\mathcal{H} = L^2(\mathbb{T}, \eta_W)$ et $\mathcal{I} = L^2(\Omega, \mathcal{A}, \mathbb{P})$. Pour $A, B \in \mathcal{B}(\mathbb{T})$, nous avons

$$\langle \mathbb{1}_A, \mathbb{1}_B \rangle_{\mathcal{H}} = \int \mathbb{1}_A \mathbb{1}_B d\eta_W = \langle W(A), W(B) \rangle_{\mathcal{I}}.$$

Comme les fonctions étagées sont denses dans $L^2(\mathbb{T}, \eta_W)$, $\overline{\text{Vect}(\mathbb{1}_A, A \in \mathcal{B}(\mathbb{T}))} = L^2(\mathbb{T}, \eta_W)$, le résultat découle du théorème de prolongement des isométries (théorème 2.5.4). \square

Dans la suite, nous utiliserons la même notation pour le champ aléatoire W et pour l'isométrie canoniquement associée W . Nous noterons, pour $f \in L^2(\mathbb{T}, \eta_W)$,

$$W(f) \stackrel{\text{def}}{=} \int_{\mathbb{T}} f dW = \int_{\mathbb{T}} f(\lambda) dW(\lambda). \quad (3.11)$$

Comme, par construction, $f \rightarrow W(f)$ est une isométrie et que les isométries entre espaces de Hilbert sont automatiquement linéaires, nous avons pour tout $(f, g) \in L^2(\mathbb{T}, \eta_W)$, et tout $(u, v) \in \mathbb{C} \times \mathbb{C}$,

$$\int_{\mathbb{T}} (uf + vg) dW = u \int_{\mathbb{T}} f dW + v \int_{\mathbb{T}} g dW.$$

et

$$\mathbb{E} \left[\left(\int_{\mathbb{T}} f dW \right) \overline{\left(\int_{\mathbb{T}} g dW \right)} \right] = \int_{\mathbb{T}} f \bar{g} d\eta_W.$$

Pour toute variable aléatoire $Y \in \overline{\text{Vect}(W(A), A \in \mathcal{B}(\mathbb{T}))}$, $\mathbb{E}[Y] = 0$, nous avons $\mathbb{E} \left[\int_{\mathbb{T}} f dW \right] = 0$.

Le théorème suivant montre en fait que toutes les isométries de $L^2(\mathbb{T}, \mu)$ à valeurs dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ peuvent se représenter sous la forme d'une intégrale stochastique par rapport à un champ spectral non corrélé de mesure de base μ .

Théorème 3.4.6. *Soient μ une mesure finie sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ et J une isométrie de $L^2(\mathbb{T}, \mu)$ à valeurs dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ telle que pour tout $f \in L^2(\mathbb{T}, \mu)$, $\mathbb{E}[J(f)] = 0$. Alors, il existe un champ spectral non corrélé centré W de base μ tel que, pour tout $f \in L^2(\mathbb{T}, \mu)$, $J(f) = \int_{\mathbb{T}} f dW$. De plus, J détermine W de façon unique.*

Preuve. Pour $A \in \mathcal{B}(\mathbb{T})$, définissons le champ $W(A) = J(\mathbb{1}_A)$. Comme J est une isométrie, pour tout $A, B \in \mathcal{B}(\mathbb{T})$,

$$\text{Cov}(W(A), W(B)) = \langle J(\mathbb{1}_A), J(\mathbb{1}_B) \rangle_{L^2(\Omega, \mathcal{A}, \mathbb{P})} = \langle \mathbb{1}_A, \mathbb{1}_B \rangle_{L^2(\mathbb{T}, \mu)} = \mu(A \cap B).$$

et par le lemme 3.4.3, W est un champ aléatoire non corrélé sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ de mesure de base μ . Puisque W est une isométrie et $W(A) = J(\mathbb{1}_A)$ pour tout $A \in \mathcal{B}(\mathbb{T})$, alors on a $J(f) = \int f dW$ pour tout $f \in L^2(\mathbb{T}, \mu)$. \square

3.4.3 Champ spectral et représentation spectrale

Nous allons maintenant relier les champs aléatoires décorrelés sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ et les processus stationnaires au second ordre.

Proposition 3.4.7. *Soit W un champ aléatoire décorrelé sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ de base η_W , où η_W est une mesure finie sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$. Alors, la transformée de Fourier du champ $X_n = \int_{\mathbb{T}} e^{in\lambda} dW(\lambda)$, $n \in \mathbb{N}$ est un processus stationnaire au second ordre, centré, de mesure spectrale η_W .*

Preuve. Posons $f_n(\lambda) = e^{in\lambda}$ pour $n \in \mathbb{N}$. Nous avons $f_n \in L^2(\mathbb{T}, \eta_W)$. Comme $W : f_n \mapsto W(f_n)$ est une isométrie, nous avons pour tout $(n, m) \in \mathbb{N}^2$,

$$\begin{aligned} \text{Cov}(X_n, X_m) &= \mathbb{E}[X_n \bar{X}_m] = \langle W(f_n), W(f_m) \rangle_{L^2(\Omega, \mathcal{A}, \mathbb{P})} \\ &= \langle f_m, f_n \rangle_{L^2(\mathbb{T}, \eta_W)} = \int_{\mathbb{T}} e^{i(m-n)\lambda} \eta_W(d\lambda). \end{aligned}$$

□

Nous allons maintenant montrer que tout processus stationnaire (X_n) est la transformée de Fourier d'un champ aléatoire décorrélé, appelé *champ spectral* du processus.

Définition 3.4.8 (Enveloppe linéaire d'un processus). Soit $X = \{X_t, t \in \mathbb{Z}\}$ un processus du second ordre. On appelle *enveloppe linéaire* du processus X et l'on note \mathcal{H}_∞^X la fermeture dans $L^2(\Omega)$ du sous-espace engendré par les v.a. $\{X_t, t \in \mathbb{Z}\}$,

$$\mathcal{H}_\infty^X = \overline{\text{Vect}(X_t, t \in \mathbb{Z})}.$$

Cet ensemble est alors le sous-espace de $L^2(\Omega, \mathcal{A}, \mathbb{P})$ contenant toute v.a. Y pour lesquelles il existe une suite d'éléments $(Y_n)_{n \geq 1}$ de $\text{Vect}(X_t, t \in \mathbb{Z})$ (l'espace des combinaisons linéaires finies formées d'éléments de $\{X_t, t \in \mathbb{Z}\}$) qui converge vers Y au sens L^2 quand $n \rightarrow \infty$, *i.e.*

$$\lim_{n \rightarrow \infty} \mathbb{E}|Y - Y_n|^2 \rightarrow 0.$$

Théorème 3.4.9. Soit (X_n) un processus stationnaire au second-ordre centré de mesure spectrale μ_X . Alors il existe un unique champ aléatoire décorrélé \hat{X} sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ de base μ_X , appelé *champ spectral* de X , tel que, pour tout $n \in \mathbb{N}$, $X_n = \int_{\mathbb{T}} e^{in\lambda} d\hat{X}(\lambda)$.

Ce champ spectral définit une unique isométrie entre $L^2(\mathbb{T}, \mu_X)$ et \mathcal{H}_∞^X , l'enveloppe linéaire du processus (X_n) .

Preuve. Posons $\mathcal{H} = L^2(\mathbb{T}, \mu_X)$, $\mathcal{I} = L^2(\Omega, \mathcal{A}, \mathbb{P})$ et pour $n \in \mathbb{Z}$, $f_n(\lambda) = e^{in\lambda}$. Considérons les suites $\{f_n, n \in \mathbb{Z}\}$ et $\{X_n, n \in \mathbb{Z}\}$. Remarquons que $\overline{\text{Vect}(f_n, n \in \mathbb{Z})} = L^2(\mathbb{T}, \mu_X)$. Par définition de la mesure spectrale, pour tout $(n, m) \in \mathbb{Z}^2$,

$$\langle f_n, f_m \rangle_{L^2(\mathbb{T}, \mu_X)} = \int_{\mathbb{T}} e^{in\lambda} e^{-im\lambda} \mu_X(d\lambda) = \text{Cov}(X_n, X_m) = \langle X_n, X_m \rangle_{L^2(\Omega, \mathcal{A}, \mathbb{P})}.$$

D'après le théorème 2.5.4 il existe une unique isométrie S_X de $\overline{\text{Vect}(f_n, n \in \mathbb{Z})} = L^2(\mathbb{T}, \mu_X)$ à valeurs dans $\overline{\text{Vect}(X_n, n \in \mathbb{Z})} = \mathcal{H}_\infty^X$, l'enveloppe linéaire du processus (X_n) telle que, pour tout $n \in \mathbb{Z}$, $S_X(f_n) = X_n$. Pour $A \in \mathcal{B}(\mathbb{T})$, considérons le champ aléatoire $\hat{X}(A) = S_X(\mathbb{1}_A) \in \mathcal{H}_\infty^X$. Comme (X_n) est un

processus centré, toute variable aléatoire $Y \in \mathcal{H}_\infty^X$ vérifie $\mathbb{E}[Y] = 0$. Donc, pour tout $A \in \mathcal{B}(\mathbb{T})$, $\mathbb{E}[\hat{X}(A)] = 0$. D'autre part, pour $A, B \in \mathcal{B}(\mathbb{T})$,

$$\mu_X(A \cap B) = \langle 1_A, 1_B \rangle_{L^2(\mathbb{T}, \mu_X)} = \left\langle \hat{X}(A), \hat{X}(B) \right\rangle_{L^2(\Omega, \mathcal{A}, \mathbb{P})},$$

ce qui montre, en utilisant le Lemme 3.4.3, que \hat{X} est un champ aléatoire décorrélé.

Le théorème 3.4.5 montre que $\hat{X} : f \mapsto \int f d\hat{X}$ définit une isométrie de $L^2(\mathbb{T}, \mu_X)$ dans \mathcal{H}_∞^X . Mais comme, pour tout $A \in \mathcal{B}(\mathbb{T})$, nous avons, $\int 1_A d\hat{X} = S_X(1_A)$ et que $L^2(\mathbb{T}, \mathcal{B}(\mathbb{T})) = \overline{\text{Vect}(1_A, A \in \mathcal{B}(\mathbb{T}))}$, nous avons $\int f d\hat{X} = S_X(f)$ pour tout $f \in L^2(\mathbb{T}, \mu_X)$. Par conséquent,

$$X_n = \hat{X}(f_n) = \int_{\mathbb{T}} e^{in\lambda} d\hat{X}(\lambda).$$

□

Problèmes

3.4.1 (Relation entre stationnarité au second ordre et stationnarité stricte). Soit X une variable aléatoire normalement distribuée, de moyenne nulle et de variance 1. Soit $Y = X\mathbf{1}_{\{U=1\}} - X\mathbf{1}_{\{U=0\}}$ où U est une variable de Bernoulli de paramètre $1/2$, indépendante de X .

1. Montrer que Y suit une loi gaussienne centrée réduite.
2. Montrer que $\text{cov}(X, Y) = 0$ mais que X et Y ne sont pas indépendants.
3. En déduire un processus qui soit un bruit blanc faible mais qui ne soit pas un bruit blanc fort.

3.4.2 (Somme de processus stationnaires au second ordre). Soient $(X_t)_{t \in \mathbb{Z}}, (Y_t)_{t \in \mathbb{Z}}$ deux processus stationnaires au second ordre décorrés c'est-à-dire tels que X_t et Y_s soient décorrés pour tout t, s (*i.e.* $\text{Cov}(X_t, Y_s) = 0$). Montrer que $(Z_t)_{t \in \mathbb{Z}}$ défini par $Z_t = X_t + Y_t$ est un processus stationnaire au second ordre. Comment sa fonction d'autocovariance se déduit-elle de celles des deux processus de départ ?

3.4.3 (Stationnarité de processus). Trouver les processus stationnaires au second ordre parmi les processus suivants :

- a) $X_t = Z_t$ si t est pair, et $X_t = Z_t + 1$ si t est impair, et $(Z_t)_{t \in \mathbb{Z}}$ est un processus stationnaire au second ordre.
- b) $X_t = Z_t Z_{t-1}$ où $(Z_t)_{t \in \mathbb{Z}}$ est un bruit blanc fort.
- c) $Y_t = (-1)^t Z_t$ et $X_t = Y_t + Z_t$, où $(Z_t)_{t \in \mathbb{Z}}$ est un bruit blanc fort.

3.4.4. On considère la fonction χ définie par (3.7) pour $\rho \in \mathbb{R}$ donné.

- a) On suppose $\rho > 1/2$. Montrer que χ n'est pas de type positif. On utilisera la suite $a_k = (-1)^k$ pour tout $k \geq 1$.
- b) Idem avec $\rho < -1/2$ et $a_k = 1$.
- c) Redémontrer en utilisant le corollaire du Théorème de Herglotz que χ est la fonction d'autocovariance d'un processus réel stationnaire au second ordre si et seulement si $\rho \in [-1/2, 1/2]$.

3.4.5 (Stationnarité, fonction d'autocovariance, densité spectrale). Soient $\{X_t\}$ et $\{Y_t\}$ deux processus stationnaires, centrés et indépendants *i.e.* pour tout s, t , X_s et Y_t sont indépendants. On note γ_X et γ_Y leur fonction d'autocovariance, que l'on suppose absolument sommables. On note f_X et f_Y leur densité spectrale.

1. Justifier la stationnarité du processus $S_t = X_t + Y_t$ ainsi que l'existence de sa densité spectrale f_S .
2. Exprimer f_S en fonction de f_X et f_Y .
3. Montrer que le processus Z_t défini par $Z_t = X_t Y_t$ est stationnaire et calculer sa fonction d'autocovariance γ_Z en fonction de γ_X et de γ_Y .

4. On définit à présent pour deux fonctions f et g 2π -périodiques leur produit de convolution par

$$f \star g(x) = \int_{-\pi}^{\pi} f(u)g(x-u)du .$$

Calculer $f_X \star f_Y$ et en déduire f_Z en fonction de f_X et de f_Y .

3.4.6 (Processus harmonique bruité). Soit $\{X_t\}$ le processus défini par

$$X_t = A \cos(\pi t/3) + B \sin(\pi t/3) + Y_t ,$$

où A et B sont deux variables aléatoires indépendantes de moyenne nulle, de variance σ^2 et indépendantes de $\{Y_t\}$.

1. Calculer la fonction d'autocovariance et la mesure spectrale de $\{X_t\}$ lorsque $\{Y_t\}$ est un bruit blanc de variance ν^2 .
2. Même question lorsque $\{Y_t\}$ est tel que $Y_t = Z_t + 2.5Z_{t-1}$, où $\{Z_t\}$ est un bruit blanc de variance ν^2 .

3.4.7 (Propriétés de la fonction d'autocovariance). Les fonctions suivantes sont-elles des fonctions d'autocovariance d'un processus stationnaire au second ordre ?

1. $\gamma(h) = 1$ si $h = 0$ et $\gamma(h) = 1/h$ si $h \neq 0$.
2. $\gamma(h) = 1 + \cos(\pi h/2)$.

3.4.8 (Elimination de tendance et de saisonnalité). On considère le processus modélisé par

$$Y_t = \beta t + S_t + U_t ,$$

où $\beta \in \mathbb{R}$, S_t est une fonction périodique de période 4 et $(U_t)_{t \in \mathbb{Z}}$ est un processus stationnaire au second ordre.

1. (Y_t) est-il stationnaire au second ordre ?
2. On considère dorénavant le processus (Z_t) obtenu par application de $1 - B^4$ à (Y_t) , où B désigne l'opérateur de retard. Montrer que (Z_t) est stationnaire au second ordre et calculer sa fonction d'autocovariance en fonction de celle de (U_t) .

3.4.9 (Processus linéaires). Soit $\{X_t\}$ donné par $X_t = \sum_{k \in \mathbb{Z}} \psi_k Z_{t-k}$, où $\{Z_t\}$ est un bruit blanc de variance σ^2 et $\{\psi_k\}_k$ est une famille de réels absolument sommables $\sum_{k \in \mathbb{Z}} |\psi_k| < \infty$.

1. Calculer la fonction d'auto-covariance de $\{X_t\}$.
2. Montrer que la famille des coefficients d'auto-covariance est absolument sommable : $\sum_h |\gamma_X(h)| < \infty$.
3. Exprimer la densité spectrale f_X de $\{X_t\}$.

4. Calculer la densité spectrale du processus $\{X_t\}$ défini par $X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$ où $\{Z_t\}$ est un bruit blanc de variance σ^2 et les θ_i sont des réels.

3.4.10 (Exemples de densités spectrales). Parmi les fonctions suivantes lesquelles peuvent être la densité spectrale d'un processus stationnaire :

1. $f(\lambda) = 1 - (1/2)\lambda^2$, $\lambda \in [-\pi, \pi]$?
2. $f(\lambda) = 1 + (1/2)\lambda$, $\lambda \in [-\pi, \pi]$?
3. $f(\lambda) = 476 + \cos(14\lambda)$, $\lambda \in [-\pi, \pi]$?

3.4.11 (Bornes pour les valeurs propres d'une matrice de covariance). Soit un processus (Y_t) de densité spectrale $f(\lambda)$ telle que $0 \leq m \leq f(\lambda) \leq M < \infty$. Pour $n \geq 1$, on note Γ_n la matrice de covariance de $(Y_1, \dots, Y_n)'$. Montrer que les valeurs propres de Γ_n sont dans un intervalle à préciser.

Partie II

Filtrage et prédiction

Chapitre 4

Filtrage des processus stationnaires au second ordre

Dans ce chapitre nous nous intéressons à une classe très importante de processus du second ordre, les processus autorégressifs à moyenne ajustée ou processus ARMA. Afin de pouvoir étudier leurs propriétés, nous allons tout d'abord établir les propriétés des processus obtenus par un filtrage linéaire de processus stationnaires au second ordre.

4.1 Filtrages linéaires de processus au second ordre

On s'intéresse dans ce paragraphe aux propriétés du processus (Y_t) obtenu comme image du processus (X_t) par le filtre linéaire suivant :

$$Y_t = \sum_{k \in \mathbb{Z}} \psi_k X_{t-k} , \quad (4.1)$$

où (ψ_k) est une suite de nombres complexes. Lorsqu'il n'y a qu'un nombre fini de ψ_k non nuls, la somme (4.1) est bien définie. On dit dans ce cas-là que le filtre est à réponse impulsionnelle finie. La question devient plus délicate lorsque l'on considère des filtres à réponse impulsionnelle infinie c'est à dire lorsque le nombre de ψ_k non nuls est infini. En effet, Y_t défini par (4.1) est la limite dans un sens à préciser, d'une suite de variables aléatoires. Le théorème 4.1.1 donne un sens précis à cette limite.

Théorème 4.1.1. *Soit $(\psi_k)_{k \in \mathbb{Z}}$ une suite absolument sommable, i.e. $\sum_{k=-\infty}^{\infty} |\psi_k| < \infty$ et soit (X_t) un processus aléatoire tel que $\sup_{t \in \mathbb{Z}} \mathbb{E}[|X_t|] < \infty$. Alors, pour tout $t \in \mathbb{Z}$, la suite :*

$$Y_{n,t} = \sum_{k=-n}^n \psi_k X_{t-k}$$

converge presque sûrement, quand n tend vers l'infini, vers une limite Y_t que nous notons

$$Y_t = \sum_{k=-\infty}^{\infty} \psi_k X_{t-k} .$$

De plus, la variable aléatoire Y_t est intégrable, i.e. $\mathbb{E}[|Y_t|] < \infty$ et la suite $(Y_{n,t})_{n \geq 0}$ converge vers Y_t dans $L^1(\Omega, \mathcal{A}, \mathbb{P})$, i.e.

$$\lim_{n \rightarrow \infty} \mathbb{E}[|Y_{n,t} - Y_t|] = 0 .$$

Supposons que $\sup_{t \in \mathbb{Z}} \mathbb{E}[|X_t|^2] < \infty$ alors $\mathbb{E}[|Y_t|^2] < \infty$ et la suite $(Y_{n,t})_{n \geq 0}$ converge en moyenne quadratique vers la variable aléatoire Y_t , i.e.

$$\lim_{n \rightarrow \infty} \mathbb{E}[|Y_{n,t} - Y_t|^2] = 0 .$$

Preuve. Notons pour tout $t \in \mathbb{Z}$ et $n \in \mathbb{N}$, $U_{n,t} = \sum_{k=-n}^n |\psi_k| |X_{t-k}|$. La suite $(U_{n,t})_{n \geq 0}$ est une suite de variables aléatoires intégrables. Puisque $\lim_{n \rightarrow \infty} \uparrow U_{n,t} = \sum_{k \in \mathbb{Z}} |\psi_k| |X_{t-k}|$, on en déduit que (théorème de Beppo-Levi)

$$\lim_{n \rightarrow \infty} \uparrow \mathbb{E}[U_{n,t}] = \mathbb{E} \left[\sum_{k \in \mathbb{Z}} |\psi_k| |X_{t-k}| \right],$$

où $\lim_{n \rightarrow \infty} \uparrow$ signifie qu'il s'agit d'une limite croissante. Comme

$$\mathbb{E}[U_{n,t}] \leq \sum_{k=-n}^n |\psi_k| \mathbb{E}[|X_{t-k}|] \leq \sup_{t \in \mathbb{Z}} \mathbb{E}[|X_t|] \sum_{k \in \mathbb{Z}} |\psi_k| < \infty,$$

on en déduit que

$$\mathbb{E} \left[\sum_{k \in \mathbb{Z}} |\psi_k| |X_{t-k}| \right] < \infty.$$

Par conséquent, il existe un ensemble $\Omega_0 \in \mathcal{F}$ tel que $\mathbb{P}(\Omega_0) = 1$ et tel que, pour tout $\omega \in \Omega_0$,

$$\sum_{k \in \mathbb{Z}} |\psi_k| |X_{t-k}(\omega)| < \infty.$$

Donc pour tout $\omega \in \Omega_0$,

$$|Y_{n,t}(\omega) - Y_t(\omega)| \leq \sum_{|k| > n} |\psi_k| |X_{t-k}(\omega)| \rightarrow 0, \text{ lorsque } n \rightarrow \infty.$$

Ainsi, pour tout $\omega \in \Omega_0$, $Y_{n,t}(\omega)$ est convergente et converge vers $Y_t(\omega)$, ce qui montre que $\lim_{n \rightarrow \infty} Y_{n,t} = Y_t$ p.s. . Le lemme de Fatou montre que

$$\mathbb{E}[|Y_t|] = \mathbb{E} \left[\liminf_n |Y_{n,t}| \right] \leq \liminf_n \mathbb{E}[Y_{n,t}] \leq \sup_t \mathbb{E}[|X_t|] \sum_{j=-\infty}^{\infty} |\psi_j| < \infty,$$

et donc que $Y_t \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. Comme $|Y_{n,t} - Y_t| \leq \sum_{k \in \mathbb{Z}} |\psi_k| |X_{t-k}|$, le théorème de convergence dominée montre que $\lim_n \mathbb{E}[|Y_{n,t} - Y_t|] = 0$ et donc que la suite $\{Y_{n,t}\}$ converge vers Y_t dans $L^1(\Omega, \mathcal{A}, \mathbb{P})$.

Considérons maintenant le cas où $\sup_{t \in \mathbb{Z}} \mathbb{E}[|X_t|^2] < \infty$. Remarquons tout d'abord que $\mathbb{E}[|X_t|] \leq (\mathbb{E}[|X_t|^2])^{1/2}$ et donc que cette condition implique que $\sup_{t \in \mathbb{Z}} \mathbb{E}[|X_t|] < \infty$. La suite $(Y_{m,t})_{m \geq 0}$ est une suite de Cauchy dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$. En effet, pour $p \geq q$, nous avons en notant $\|X\|_2 = (\mathbb{E}[|X|^2])^{1/2}$

$$\|Y_{p,t} - Y_{q,t}\|_2 = \left\| \sum_{|k|=q+1}^p \psi_k X_{t-k} \right\|_2 \leq \sup_t \|X_t\|_2 \sum_{|k|=q+1}^p |\psi_k| \xrightarrow{q,p \rightarrow \infty} 0.$$

Comme $L^2(\Omega, \mathcal{A}, \mathbb{P})$ est complet, la suite $(Y_{n,t})$ converge vers une variable Y_t^* . En utilisant le lemme de Fatou, nous avons

$$\mathbb{E}[|Y_t - Y_t^*|^2] = \mathbb{E}\left[\liminf_n |Y_{n,t} - Y_t^*|^2\right] \leq \liminf_n \mathbb{E}[|Y_{n,t} - Y_t^*|^2] = 0,$$

ce qui montre que les limites p.s. Y_t et $L^2(\Omega, \mathcal{A}, \mathbb{P})$, Y_t^* coïncident p.s. . \square

Le résultat suivant établit que le processus (Y_t) obtenu par filtrage linéaire d'un processus stationnaire au second ordre (X_t) via l'équation (4.1) est lui-même stationnaire au second ordre, à condition que la suite des (ψ_k) soit absolument sommable *i.e.* $\sum_{k \in \mathbb{Z}} |\psi_k| < \infty$.

Théorème 4.1.2 (Filtrage des processus stationnaires au second ordre). *Soit (ψ_k) une suite telle que $\sum_{k=-\infty}^{\infty} |\psi_k| < \infty$ et soit (X_t) un processus stationnaire au second ordre de moyenne $\mu_X = \mathbb{E}[X_t]$ et de fonction d'autocovariance $\gamma_X(h) = \text{cov}(X_{t+h}, X_t)$ alors le processus $Y_t = \sum_{k=-\infty}^{\infty} \psi_k X_{t-k}$ est stationnaire au second ordre de moyenne :*

$$\mu_Y = \mu_X \sum_{k=-\infty}^{\infty} \psi_k, \quad (4.2)$$

de fonction d'autocovariance :

$$\gamma_Y(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \bar{\psi}_k \gamma_X(h + k - j), \quad (4.3)$$

et de mesure spectrale :

$$\nu_Y(d\lambda) = |\psi(e^{-i\lambda})|^2 \nu_X(d\lambda), \quad (4.4)$$

où $\psi(e^{-i\lambda}) = \sum_{k \in \mathbb{Z}} \psi_k e^{-ik\lambda}$.

Preuve. D'après la continuité du produit scalaire dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$, voir théorème 2.1.11, on a

$$\begin{aligned} \mathbb{E}\left[\sum_{k \in \mathbb{Z}} \psi_k X_{t-k}\right] &= \mathbb{E}\left[\lim_{n \rightarrow \infty} \sum_{k=-n}^n \psi_k X_{t-k}\right] = \lim_{n \rightarrow \infty} \mathbb{E}\left[\sum_{k=-n}^n \psi_k X_{t-k}\right] \\ &= \mu_X \left(\lim_{n \rightarrow \infty} \sum_{k=-n}^n \psi_k\right) = \mu_X \sum_{k \in \mathbb{Z}} \psi_k. \end{aligned}$$

Montrons à présent le résultat sur la fonction d'auto-covariance. D'après la continuité du produit scalaire dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$, voir théorème 2.1.11, on a

$$\begin{aligned} \text{Cov}(Y_t, Y_{t+h}) &= \lim_{n \rightarrow \infty} \sum_{k,j=-n}^n \psi_k \bar{\psi}_j \text{Cov}(X_{t-k}, X_{t+h-j}) \\ &= \sum_{k \in \mathbb{Z}} \psi_k \bar{\psi}_j \gamma_X(h+k-j), \end{aligned}$$

ce qui montre Eq. (4.3)

D'après le théorème 3.3.1, $\gamma_X(h) = \int_{\mathbb{T}} e^{ih\lambda} \nu_X(d\lambda)$ où ν_X désigne la mesure spectrale du processus (X_t) . En reportant cette expression de $\gamma_X(h)$ dans (4.3), nous obtenons

$$\gamma_Y(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \bar{\psi}_k \int_{\mathbb{T}} e^{i(h+k-j)\lambda} \nu_X(d\lambda). \quad (4.5)$$

Puisque

$$\sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \int_{\mathbb{T}} |\psi_j| |\psi_k| \nu_X(d\lambda) \leq \gamma_X(0) \left(\sum_{j=-\infty}^{\infty} |\psi_j| \right)^2 < \infty,$$

on peut appliquer le théorème de Fubini et permuter les signes somme et intégrale dans (4.5). Ce qui donne :

$$\gamma_Y(h) = \int_{\mathbb{T}} e^{ih\lambda} \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \bar{\psi}_k e^{ik\lambda} e^{-ij\lambda} = \int_{\mathbb{T}} e^{ih\lambda} |\psi(e^{-i\lambda})|^2 \nu_X(d\lambda),$$

où $\psi(e^{-i\lambda}) = \sum_{k \in \mathbb{Z}} \psi_k e^{-ik\lambda}$. On en déduit, d'après le théorème 3.3.1 que $\nu_Y(d\lambda) = |\psi(e^{-i\lambda})|^2 \nu_X(d\lambda)$. \square

Nous définissons à présent une classe très importante de processus obtenus par filtrage : les *processus linéaires* qui sont obtenus en filtrant un bruit blanc.

Définition 4.1.3 (Processus linéaire). Nous dirons que (X_t) est un *processus linéaire* s'il existe un bruit blanc $Z_t \sim \text{BB}(0, \sigma^2)$ et une suite de coefficients $(\psi_k)_{k \in \mathbb{Z}}$ absolument sommable telle que :

$$X_t = \mu + \sum_{k=-\infty}^{\infty} \psi_k Z_{t-k}, \quad t \in \mathbb{Z}, \quad (4.6)$$

où μ est un nombre complexe. On dira que $(X_t)_{t \in \mathbb{Z}}$ est un processus linéaire *causal* par rapport à $(Z_t)_{t \in \mathbb{Z}}$ si (4.6) est vérifiée avec $\psi_k = 0$ pour tout $k < 0$. On dira que $(X_t)_{t \in \mathbb{Z}}$ est un processus linéaire *inversible* par rapport à $(Z_t)_{t \in \mathbb{Z}}$ si (4.6) est vérifiée et qu'il existe de plus une suite $(\pi_k)_{k \geq 0}$ absolument sommable telle que

$$Z_t = \sum_{k=0}^{\infty} \pi_k (X_{t-k} - \mu), \quad t \in \mathbb{Z}. \quad (4.7)$$

D'après le théorème 4.1.2, un processus linéaire est stationnaire au second ordre de moyenne μ , de fonction d'autocovariance :

$$\gamma_X(h) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \bar{\psi}_{j+h} = \sigma^2 \sum_{\ell=-\infty}^{\infty} \psi_{\ell-h} \bar{\psi}_{\ell}, \quad (4.8)$$

et dont la mesure spectrale admet une densité donnée par :

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} |\psi(e^{-i\lambda})|^2, \quad (4.9)$$

où $\psi(e^{-i\lambda}) = \sum_{k \in \mathbb{Z}} \psi_k e^{-ik\lambda}$.

4.2 Processus ARMA

Avant de passer au cas général des processus ARMA, nous nous intéressons à deux classes de processus ARMA particuliers : les processus à moyenne ajustée (MA) et les processus autorégressifs (AR).

4.2.1 Processus MA(q)

Définition 4.2.1 (Processus MA(q)). On dit que le processus (X_t) est à moyenne ajustée d'ordre q (ou MA(q)) si X_t est donné par :

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \quad (4.10)$$

où $Z_t \sim \text{BB}(0, \sigma^2)$ et les θ_i sont des nombres complexes.

Le terme “moyenne ajustée” est la traduction assez malheureuse du nom anglo-saxon “moving average” (moyenne mobile). Observons que $X_t = \sum_{k=0}^q \theta_k Z_{t-k}$, avec la convention $\theta_0 = 1$. En utilisant les résultats du théorème 4.1.2, on obtient $\mathbb{E}[X_t] = 0$, et

$$\gamma_X(h) = \begin{cases} \sigma^2 \sum_{t=0}^{q-h} \theta_t \bar{\theta}_{t+h}, & \text{si } 0 \leq h \leq q, \\ \sigma^2 \sum_{t=0}^{q+h} \bar{\theta}_t \theta_{t-h}, & \text{si } -q \leq h \leq 0, \\ 0, & \text{sinon.} \end{cases} \quad (4.11)$$

Enfin, d'après la formule (4.9), le processus admet une densité spectrale dont l'expression est :

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \left| 1 + \sum_{k=1}^q \theta_k e^{-ik\lambda} \right|^2 .$$

Un exemple de densité spectrale pour le processus MA(1) est représenté sur la figure 3.6.

4.2.2 Processus AR(p)

Définition 4.2.2 (Processus AR(p)). On dit que le processus $\{X_t\}$ est un processus autorégressif d'ordre p (ou AR(p)) si $\{X_t\}$ est un processus stationnaire au second ordre et s'il est solution de l'équation de récurrence :

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t , \quad (4.12)$$

où $Z_t \sim \text{BB}(0, \sigma^2)$ est un bruit blanc et les ϕ_k sont des nombres complexes.

Le terme “autorégressif” provient de la forme de l'équation (4.12) dans laquelle la valeur courante du processus s'exprime sous la forme d'une régression des p valeurs précédentes du processus plus un bruit additif.

L'existence et l'unicité d'une solution stationnaire au second ordre de l'équation (4.12) sont des questions délicates (qui ne se posaient pas lorsque nous avons défini les modèles MA). Nous détaillons ci-dessous la réponse à cette question dans le cas $p = 1$.

4.2.2.1 Cas : $|\phi_1| < 1$

L'équation de récurrence (4.12) s'écrit dans le cas $p = 1$:

$$X_t = \phi_1 X_{t-1} + Z_t , \quad (4.13)$$

où $(Z_t) \sim \text{BB}(0, \sigma^2)$. En itérant (4.13), on obtient :

$$\begin{aligned} X_t &= \phi_1(\phi_1 X_{t-2} + Z_{t-1}) + Z_t = \phi_1^2 X_{t-2} + \phi_1 Z_{t-1} + Z_t \\ &= \phi_1^{k+1} X_{t-k-1} + \phi_1^k Z_{t-k} + \cdots + \phi_1^2 Z_{t-2} + \phi_1 Z_{t-1} + Z_t . \end{aligned}$$

En prenant la limite quand $k \rightarrow \infty$, on en déduit que

$$X_t = \sum_{j=0}^{\infty} \phi_1^j Z_{t-j} , \quad (4.14)$$

la série convergeant dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ et p.s. . En effet, si on suppose que X_t une solution stationnaire,

$$\mathbb{E} \left[\left| X_t - \sum_{j=0}^k \phi_1^j Z_{t-j} \right|^2 \right] = |\phi_1|^{2k+2} \mathbb{E} [|X_{t-k-1}|^2] = |\phi_1|^{2k+2} \mathbb{E} [|X_0|^2] \rightarrow 0, \quad k \rightarrow \infty,$$

puisque $|\phi_1| < 1$. De plus, d'après la définition 4.1.3, (X_t) défini par (4.14) est un processus linéaire et est donc stationnaire au second ordre. On peut vérifier que (X_t) défini par (4.14) est bien solution de (4.13) en notant que :

$$X_t = Z_t + \phi_1 \sum_{k=0}^{+\infty} \phi_1^k Z_{t-1-k} = Z_t + \phi_1 X_{t-1}.$$

Remarquons que la solution donnée par (4.14) peut être obtenue en utilisant le développement la fraction rationnelle $\psi(z) = (1 - \phi_1 z)^{-1}$ en série entière

$$\psi(z) = \frac{1}{1 - \phi_1 z} = \sum_{k=0}^{+\infty} \phi_1^k z^k$$

convergeant sur le disque $\{z \in \mathbb{C} : |z| < |\phi_1|^{-1}\}$. Ce lien n'a rien de fortuit, comme nous le verrons dans le paragraphe 4.2.3.

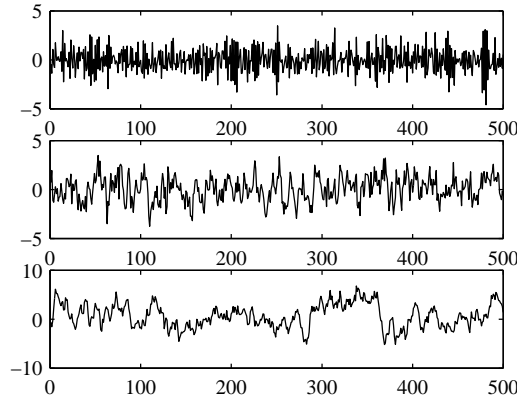


Fig. 4.1 Trajectoires de longueur 500 d'un processus AR(1) gaussien. Courbe du haut : $\phi_1 = -0.7$. Courbe du milieu : $\phi_1 = 0.5$. Courbe du bas : $\phi_1 = 0.9$

La fonction d'autocovariance de (X_t) solution stationnaire de (4.13) est donnée par la formule (4.8) qui s'écrit ;

$$\gamma_X(h) = \sigma^2 \sum_{k=0}^{\infty} \phi_1^k \bar{\phi}_1^{k+h} = \sigma^2 \frac{\bar{\phi}_1^h}{1 - |\phi_1|^2}, \quad \text{si } h \geq 0, \quad (4.15)$$

$$= \overline{\gamma(-h)}, \quad \text{sinon.} \quad (4.16)$$

Lorsque ϕ_1 est un réel strictement positif, le processus (X_t) est positivement corrélé, dans le sens où tous ses coefficients d'auto-covariance sont positifs. Les exemples de trajectoires représentées sur la figure 4.1 montrent que des valeurs de ϕ_1 proches de 1 correspondent à des trajectoires “persistantes”. Inversement, des valeurs de ϕ_1 réelles et négatives conduisent à des trajectoires où une valeur positive a tendance à être suivie par une valeur négative. la densité spectrale de (X_t) est donnée par

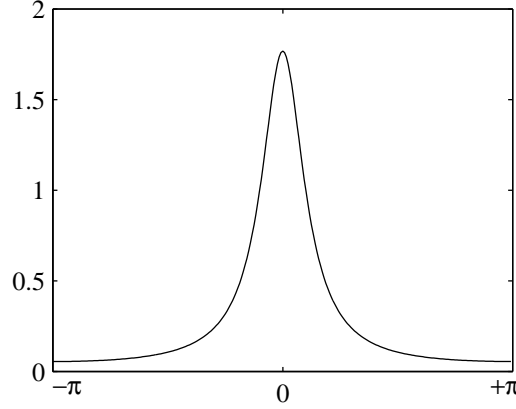


Fig. 4.2 Densité spectrale d'un processus AR(1), défini par (4.13) pour $\sigma = 1$ et $\phi_1 = 0.7$.

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{k=0}^{\infty} \phi_1^k e^{-ik\lambda} \right|^2 = \frac{\sigma^2}{2\pi} \frac{1}{|1 - \phi_1 e^{-i\lambda}|^2} . \quad (4.17)$$

4.2.2.2 Cas $|\phi_1| > 1$

Dans ce cas-là, $\mathbb{E} \left[|X_t - \sum_{j=0}^k \phi_1^j Z_{t-j}|^2 \right] = |\phi_1|^{2k+2} \mathbb{E} [X_{t-k-1}^2]$ diverge lorsque k tend vers l'infini. Par contre, on peut réécrire l'équation définissant X_t en fonction de Z_t comme suit

$$X_t = -\phi_1^{-1} Z_{t+1} + \phi_1^{-1} X_{t+1}.$$

En itérant l'équation précédente, on obtient

$$\begin{aligned} X_t &= -\phi_1^{-1} Z_{t+1} - \phi_1^{-2} Z_{t+2} + \phi_1^{-2} X_{t+2} = \dots \\ &= -\phi_1^{-1} Z_{t+1} - \phi_1^{-2} Z_{t+2} - \dots - \phi_1^{-k-1} Z_{t+k+1} + \phi_1^{-k-1} X_{t+k+1} . \end{aligned}$$

En utilisant exactement les mêmes arguments que ceux employés précédemment, on déduit que la solution stationnaire dans ce cas vaut

$$X_t = - \sum_{j \geq 1} \phi_1^{-j} Z_{t+j} . \quad (4.18)$$

Cette solution est **non causale** : elle dépend uniquement du “futur” du processus (Z_t) .

Remarquons que, comme précédemment, la solution donnée par (4.18) est obtenu en choisissant le développement fraction rationnelle $\psi(z) = (1 - \phi_1 z)^{-1}$

$$\psi(z) = \frac{1}{1 - \phi_1 z} = \frac{-(\phi_1 z)^{-1}}{1 - (\phi_1 z)^{-1}} = -(\phi_1 z)^{-1} \sum_{k=0}^{+\infty} (\phi_1 z)^{-k} = - \sum_{k \geq 1} \phi_1^{-k} z^{-k} ,$$

qui converge sur la couronne $\{z \in \mathbb{C}, |z| > |\phi_1|^{-1}\}$. Nous remarquons que nous avons choisi dans les deux cas $|\phi_1| < 1$ et $|\phi_1| > 1$ les développements convergeant dans des domaines incluant le cercle unité, $\{z \in \mathbb{C}, |z| = 1\}$. Ce lien est expliqué précisément dans le paragraphe 4.2.3.

4.2.2.3 Cas $|\phi_1| = 1$

Supposons qu’il existe une solution stationnaire dans ce cas alors, par stationnarité de X_t ,

$$\mathbb{E} \left[\left| X_t - \sum_{j=0}^{k-1} \phi_1^j Z_{t-j} \right|^2 \right] = |\phi_1|^{2k} \mathbb{E} [|X_{t-k}|^2] = |\phi_1|^{2k} \mathbb{E} [|X_t|^2] = \mathbb{E} [|X_t|^2] .$$

Or, le terme de gauche est aussi égal à

$$\mathbb{E} [|X_t|^2] + \mathbb{E} \left[\left| \sum_{j=0}^{k-1} \phi_1^j Z_{t-j} \right|^2 \right] - 2 \mathbb{E} \left[\bar{X}_t \sum_{j=0}^{k-1} \phi_1^j Z_{t-j} \right] .$$

Ainsi, $\mathbb{E} \left[\left| \sum_{j=0}^{k-1} \phi_1^j Z_{t-j} \right|^2 \right] = 2 \mathbb{E} \left[\bar{X}_t \sum_{j=0}^{k-1} \phi_1^j Z_{t-j} \right]$. De plus, $\mathbb{E} \left[\left| \sum_{j=0}^{k-1} \phi_1^j Z_{t-j} \right|^2 \right] = \sum_{j=0}^{k-1} |\phi_1|^{2j} \sigma^2 = k \sigma^2$. D’où, en utilisant l’inégalité de Cauchy-Schwarz,

$$k \sigma^2 \leq 2 \mathbb{E} [|X_t|^2]^{1/2} \mathbb{E} \left[\left| \sum_{j=0}^{k-1} \phi_1^j Z_{t-j} \right|^2 \right]^{1/2} \leq 2(\gamma_X(0) + |\mu_X|^2)^{1/2} k^{1/2} \sigma ,$$

ce qui est impossible pour k grand. Donc, dans ce cas, **il n’existe pas de solution stationnaire**.

4.2.2.4 Conclusion

Nous avons donc montré, dans le cas $p = 1$, que l'équation de récurrence (4.12) n'admettait pas de solution stationnaire lorsque $|\phi_1| = 1$ et qu'elle admettait une solution stationnaire lorsque $|\phi_1| \neq 1$, donnée par :

$$X_t = \sum_{j \geq 0} \phi_1^j Z_{t-j}, \text{ si } |\phi_1| < 1,$$

et

$$X_t = - \sum_{j \geq 1} \phi_1^{-j} Z_{t+j}, \text{ si } |\phi_1| > 1.$$

4.2.3 Cas général

Avant d'énoncer le théorème 4.2.4 qui donne une condition nécessaire et suffisante d'existence d'une solution stationnaire à l'équation récurrente (4.24) définissant un processus ARMA(p, q), nous introduisons un nouvel opérateur qui sera utile dans la preuve du théorème 4.2.4.

Soit $\mathcal{S}(\Omega, \mathcal{A}, \mathbb{P})$ l'ensemble des processus indexés par \mathbb{Z} stationnaires au second ordre et à valeurs complexes. A toute suite de coefficients complexes (α_k) vérifiant : $\sum_{k \in \mathbb{Z}} |\alpha_k| < \infty$, on associe un opérateur F_α qui à $X \in \mathcal{S}(\Omega, \mathcal{A}, \mathbb{P})$ associe le processus Y défini par :

$$F_\alpha : X \mapsto Y = (Y_t)_{t \in \mathbb{Z}} = \left(\sum_{k \in \mathbb{Z}} \alpha_k X_{t-k} \right)_{t \in \mathbb{Z}}.$$

D'après le théorème 4.1.2, Y est aussi dans $\mathcal{S}(\Omega, \mathcal{A}, \mathbb{P})$.

Le lemme 4.2.3 montre comment composer deux opérateurs de type F_α .

Lemme 4.2.3. *Soient (α_k) et (β_k) des suites de coefficients complexes telles que : $\sum_{k \in \mathbb{Z}} |\alpha_k| < \infty$ et $\sum_{k \in \mathbb{Z}} |\beta_k| < \infty$. Si $X \in \mathcal{S}(\Omega, \mathcal{A}, \mathbb{P})$ alors*

$$F_\alpha[F_\beta X] = [F_{\alpha\beta}]X, \text{ dans } L^2(\Omega, \mathcal{A}, \mathbb{P}),$$

$$\text{où } (\alpha\beta)_k = \sum_{j \in \mathbb{Z}} \alpha_j \beta_{k-j}.$$

Preuve. Soit $Y = F_\beta X$. D'après le théorème 4.1.2, puisque $\sum_k |\beta_k| < \infty$, Y est dans $\mathcal{S}(\Omega, \mathcal{A}, \mathbb{P})$. Pour les mêmes raisons, $F_\alpha Y$ est lui aussi dans $\mathcal{S}(\Omega, \mathcal{A}, \mathbb{P})$. Soient $Z = F_\alpha[F_\beta X]$ et $W = [F_{\alpha\beta}]X$, on a alors, pour tout $t \in \mathbb{Z}$, $Z_t = \sum_{j \in \mathbb{Z}} \alpha_j Y_{t-j}$, où $Y_t = \sum_{k \in \mathbb{Z}} \beta_k X_{t-k}$ et $W_t = \sum_{k \in \mathbb{Z}} (\sum_{j \in \mathbb{Z}} \alpha_j \beta_{k-j}) X_{t-k}$. Ainsi, $Z_t = \sum_{j \in \mathbb{Z}} \alpha_j (\sum_{k \in \mathbb{Z}} \beta_k X_{t-j-k})$.

Définissons $Z_{t,m,n}$ et $W_{t,m,n}$ par : $Z_{t,m,n} = \sum_{j=-m}^m \alpha_j (\sum_{k=-n}^n \beta_k X_{t-j-k})$ et $W_{t,m,n} = \sum_{k=-m}^m (\sum_{j=-n}^n \alpha_j \beta_{k-j}) X_{t-k}$. En posant $\ell = j+k$, on en déduit

que

$$Z_{t,m,n} = \sum_{\ell=-(m+n)}^{m+n} \left(\sum_{j=-m}^m \alpha_j \beta_{\ell-j} \right) X_{t-\ell} = W_{t,m+n,m} . \quad (4.19)$$

En notant $\|X\|_2 = (\mathbb{E}[|X|^2])^{1/2}$, nous pouvons écrire en utilisant l'inégalité triangulaire que :

$$\|Z_t - W_t\|_2 \leq \|Z_t - Z_{t,m,n}\|_2 + \|Z_{t,m,n} - W_{t,m+n,m}\|_2 + \|W_{t,m+n,m} - W_t\|_2 , \quad (4.20)$$

le deuxième terme du membre de droite de (4.20) étant nul d'après (4.19). D'autre part, avec : $Z_{t,m} = \sum_{j=-m}^m \alpha_j (\sum_{k \in \mathbb{Z}} \beta_k X_{t-j-k}) = \sum_{j=-m}^m \alpha_j Y_{t-j}$, on a :

$$\|Z_t - Z_{t,m,n}\|_2 \leq \|Z_t - Z_{t,m}\|_2 + \|Z_{t,m} - Z_{t,m,n}\|_2 . \quad (4.21)$$

En utilisant l'inégalité de Cauchy-Schwarz, le fait que Y est dans $\mathcal{S}(\Omega, \mathcal{A}, \mathbb{P})$ et $\sum_{k \in \mathbb{Z}} |\alpha_k| < \infty$, on a

$$\begin{aligned} \|Z_t - Z_{t,m}\|_2^2 &= \left\| \sum_{|j|>m} \alpha_j Y_{t-j} \right\|_2^2 \leq \mathbb{E} \left[\sum_{|j|>m, |j'|>m} |\alpha_j| |\alpha_{j'}| |Y_{t-j}| |Y_{t-j'}| \right] \\ &\leq \mathbb{E}[Y_0^2] \left(\sum_{|j|>m} |\alpha_j| \right)^2 \rightarrow 0, \quad m \rightarrow \infty . \end{aligned} \quad (4.22)$$

D'autre part, en utilisant l'inégalité de Cauchy-Schwarz, le fait que X est dans $\mathcal{S}(\Omega, \mathcal{A}, \mathbb{P})$ et l'absolue sommabilité de (α_k) et (β_k) :

$$\begin{aligned} \|Z_{t,m} - Z_{t,m,n}\|_2^2 &= \left\| \sum_{|j| \leq m} \alpha_j \left(\sum_{|k|>n} \beta_k X_{t-j-k} \right) \right\|_2^2 \\ &\leq \mathbb{E}[X_0^2] \left(\sum_{-m \leq j, j' \leq m} |\alpha_j| |\alpha_{j'}| \right) \left(\sum_{|k|, |k'|>n} |\beta_k| |\beta_{k'}| \right) \\ &\leq \mathbb{E}[X_0^2] \left(\sum_{j \in \mathbb{Z}} |\alpha_j| \right)^2 \left(\sum_{|k|>n} |\beta_k| \right)^2 \rightarrow 0, \quad m, n \rightarrow \infty . \end{aligned} \quad (4.23)$$

En utilisant (4.21), (4.22) et (4.23), on obtient que le premier terme du membre de droite de (4.20) tend vers 0 lorsque m et n tendent vers l'infini. On peut montrer en utilisant le même type d'arguments que $\|W_{t,m+n,m} - W_t\|_2$ tend vers 0 lorsque m et n tendent vers l'infini ce qui conclut la preuve avec (4.20). \square

Théorème 4.2.4 (Existence et unicité des processus ARMA(p, q)).
Soit l'équation récurrente :

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \quad (4.24)$$

où $Z_t \sim \text{BB}(0, \sigma^2)$ et les ϕ_j et les θ_j sont des nombres complexes. On pose $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ et $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$. On suppose que $\phi(z)$ et $\theta(z)$ n'ont pas de zéros communs. Alors l'équation (4.24) admet une solution stationnaire au second ordre si et seulement si le polynôme $\phi(z) \neq 0$ pour $|z| = 1$. Cette solution est unique et a pour expression :

$$X_t = \sum_{k=-\infty}^{\infty} \psi_k Z_{t-k}, \quad (4.25)$$

où les (ψ_k) sont donnés par les coefficients du développement

$$\frac{\theta(z)}{\phi(z)} = \sum_{k \in \mathbb{Z}} \psi_k z^k, \quad (4.26)$$

convergeant dans la couronne $\{z \in \mathbb{C}, \delta_1 < |z| < \delta_2\}$, où $\delta_1 < 1$ et $\delta_2 > 1$ sont définis par $\delta_1 = \max\{z \in \mathbb{C}, |z| < 1, \phi(z) = 0\}$ et $\delta_2 = \min\{z \in \mathbb{C}, |z| > 1, \phi(z) = 0\}$.

Preuve. Nous commençons par énoncer et prouver un lemme utile pour la preuve du théorème 4.2.4.

Lemme 4.2.5. *Soient θ et ϕ deux polynômes à coefficients complexes tels que $\phi(z) \neq 0$ pour $|z| = 1$ et $\phi(0) = 1$ alors la fraction rationnelle $\theta(z)/\phi(z)$ est développable en série de Laurent, c'est-à-dire*

$$\frac{\theta(z)}{\phi(z)} = \sum_{k \in \mathbb{Z}} c_k z^k,$$

où la série $\sum_{k \in \mathbb{Z}} c_k z^k$ est uniformément convergente dans la couronne définie par $\{z \in \mathbb{C}, r_1 < |z| < r_2\}$, où

$$r_1 = \max\{|z| : z \in \mathbb{C}, |z| < 1, \phi(z) = 0\}$$

$$r_2 = \min\{|z| : z \in \mathbb{C}, |z| > 1, \phi(z) = 0\}.$$

avec la convention $\max(\emptyset) = 0$ et $\min(\emptyset) = \infty$.

Le cas $r_1 = 0$ correspond à $\phi(z) \neq 0$ pour tout complexe z tel que $|z| \leq 1$. Dans ce cas on a $c_k = 0$ pour tout $k < 0$. Sinon, pour tout $\eta \in (0, r_1)$, $c_k = O(\eta^{-k})$ quand $k \rightarrow -\infty$.

Le cas $r_2 = \infty$ correspond à $\phi(z) \neq 0$ pour tout complexe z tel que $|z| \geq 1$. Dans ce cas on a $c_k = 0$ pour tout $k > \max(-1, \deg(\theta) - \deg(\phi))$. Sinon, pour tout $\eta \in (0, 1/r_2)$, $c_k = O(\eta^k)$ quand $k \rightarrow \infty$.

Preuve. La décomposition en éléments simples de la fraction rationnelle $\theta(z)/\phi(z)$ s'écrit comme la somme d'un polynôme de degré $\deg(\theta) - \deg(\phi)$ (avec la convention que tout polynôme de degré strictement négatif est le polynôme nul) et de termes de la forme : $a/(z - z_0)^r$, où z_0 est une racine de ϕ de multiplicité supérieure ou égale à r et a est une constante. On écrit :

$$\begin{aligned} \text{si } |z_0| < 1, \quad \frac{1}{(z - z_0)^r} &= \frac{z^{-r}}{(1 - z_0/z)^r}, \text{ lorsque } |z_0| < |z|, \\ \text{si } |z_0| > 1, \quad \frac{1}{(z - z_0)^r} &= \frac{(-z_0)^{-r}}{(1 - z/z_0)^r}, \text{ lorsque } |z| < |z_0|. \end{aligned}$$

On utilise que :

$$\begin{aligned} (1 - u)^{-r} &= \frac{(-1)^{r-1}}{(r-1)!} \sum_{k \geq r-1} \frac{k!}{(k-r+1)!} u^{k-r+1} \\ &= \frac{(-1)^{r-1}}{(r-1)!} \sum_{k \geq 0} \frac{(k+r-1)!}{k!} u^k, \text{ lorsque } |u| < 1, \end{aligned}$$

Ainsi,

$$\text{si } |z_0| < 1, \quad \frac{1}{(z - z_0)^r} = \frac{z^{-r}}{(1 - z_0/z)^r} = z^{-r} \frac{(-1)^{r-1}}{(r-1)!} \sum_{k \geq 0} \frac{(k+r-1)!}{k!} (z_0/z)^k,$$

qui converge si $|z_0| < |z|$,

$$\text{si } |z_0| > 1, \quad \frac{1}{(z - z_0)^r} = \frac{(-z_0)^{-r}}{(1 - z/z_0)^r} = -\frac{z_0^{-r}}{(r-1)!} \sum_{k \geq 0} \frac{(k+r-1)!}{k!} (z/z_0)^k,$$

qui converge si $|z| < |z_0|$.

En majorant $(k+r-1)!/k!$ par k^{r-1} , on en déduit que

$$\begin{aligned} \text{si } |z_0| < 1, \quad \frac{1}{(z - z_0)^r} &= \sum_{k \leq -r} v_k z^k, \text{ qui converge si } |z| > |z_0|, \\ \text{si } |z_0| > 1, \quad \frac{1}{(z - z_0)^r} &= \sum_{k \geq 0} w_k z^k, \text{ qui converge si } |z| < |z_0|, \end{aligned}$$

où $|v_k|$ et $|w_k|$ sont majorés par $C\eta^{|k|}$, C étant une constante strictement positive pour tout η choisi dans $(0, r_1)$ ou $(0, 1/r_2)$, respectivement. \square

Retour à la preuve du théorème 4.2.4

Supposons que $\phi(z) \neq 0$ pour $|z| = 1$, alors d'après le lemme 4.2.5 il existe $r_1 < 1$ et $r_2 > 1$ tels que

$$\psi(z) = \frac{\theta(z)}{\phi(z)} = \sum_{k=-\infty}^{\infty} \psi_k z^k, \quad r_1 < |z| < r_2, \quad (4.27)$$

où la suite $(\psi_k)_{k \in \mathbb{Z}}$ vérifie $\sum_k |\psi_k| < \infty$. Vérifions que le processus (X_t) défini par : $X_t = \sum_{k \in \mathbb{Z}} \psi_k Z_{t-k} = (F_\psi Z)_t$, pour tout $t \in \mathbb{Z}$ est une solution stationnaire de (4.24). D'après la définition 4.1.3, (X_t) est stationnaire. De plus, d'après le lemme 4.2.3, $F_\phi[F_\psi Z] = F_{\phi\psi} Z = F_\theta Z$, ce qui montre l'existence d'une solution stationnaire à (4.24).

D'autre part, si X est un processus stationnaire au second ordre solution de (4.24) alors X vérifie :

$$F_\phi X = F_\theta Z. \quad (4.28)$$

Comme $\phi(z) \neq 0$ pour $|z| = 1$, alors d'après le lemme 4.2.5 il existe $r_1 < 1$ et $r_2 > 1$ tels que :

$$\xi(z) = \frac{1}{\phi(z)} = \sum_{k \in \mathbb{Z}} \xi_k z^k, \quad r_1 < |z| < r_2,$$

où la suite $(\xi_k)_{k \in \mathbb{Z}}$ vérifie $\sum_k |\xi_k| < \infty$. On peut donc appliquer l'opérateur F_ξ aux deux membres de l'équation (4.28) d'où l'on déduit en utilisant le lemme 4.2.3 que $X = F_{\xi\theta} Z = F_\psi Z$ où (ψ_k) est définie dans (4.27). Donc $X_t = \sum_{k \in \mathbb{Z}} \psi_k Z_{t-k} = (F_\psi Z)_t$, pour tout $t \in \mathbb{Z}$, ce qui assure l'unicité de la solution.

Réciproquement, si (X_t) est un processus stationnaire solution de (4.24) de la forme $X_t = \sum_{k \in \mathbb{Z}} \eta_k Z_{t-k}$ où $\sum_k |\eta_k| < \infty$, montrons que $\phi(z) \neq 0$ pour $|z| = 1$. En effet, puisque X est solution de (4.28) alors : $F_\phi X = F_\phi[F_\eta Z] = F_\theta Z$. D'après le lemme 4.2.3, $F_{\phi\eta} Z = F_\theta Z$. Posons $\zeta_k = \sum_{j \in \mathbb{Z}} \phi_j \eta_{k-j}$. On a alors, pour tout $t \in \mathbb{Z}$, $\sum_{k \in \mathbb{Z}} \zeta_k Z_{t-k} = \sum_{j=1}^q \theta_j Z_{t-j}$. En multipliant les deux membres de cette équation par $Z_{t-\ell}$ et en prenant l'espérance, on déduit que $\zeta_\ell = \theta_\ell$, $\ell = 0, \dots, q$ et $\zeta_\ell = 0$, sinon. Ainsi, $\theta(z) = \phi(z)\eta(z)$, $|z| = 1$. Puisque θ et ϕ n'ont pas de racines communes et que $|\eta(z)| \leq \sum_{k \in \mathbb{Z}} |\eta_k| < \infty$, si $|z| = 1$, $\phi(z)$ ne s'annule pas sur le cercle unité : $\{z, |z| = 1\}$, ce qui conclut la preuve du théorème 4.2.4. □

Dans le cas où $\phi(z)$ et $\theta(z)$ ont des zéros communs, deux configurations sont possibles :

- a) Les zéros communs ne sont pas sur le cercle unité. Dans ce cas on se ramène au cas sans zéro commun en annulant les facteurs communs.
- b) Certains des zéros communs se trouvent sur le cercle unité. L'équation (4.24) admet une infinité de solutions stationnaires au second ordre.

Du point de vue de la modélisation, la présence de zéros communs ne présente aucun intérêt puisqu'elle est sans influence sur la densité spectrale de puissance. Elle conduit de plus à une ambiguïté sur l'ordre réel des parties AR et MA.

4.2.3.1 ARMA(p, q) causal

Le théorème 4.2.7 donne une condition nécessaire et suffisante d'existence d'une solution causale à l'équation (4.24).

Définition 4.2.6 (Représentation ARMA causale). Sous les hypothèses du théorème 4.2.4, on dira que l'équation (4.24) fournit une représentation causale de la solution stationnaire au second ordre (X_t) si $(X_t)_{t \in \mathbb{Z}}$ est un processus linéaire causal par rapport à $(Z_t)_{t \in \mathbb{Z}}$.

Théorème 4.2.7 (ARMA(p, q) causal). Soit l'équation récurrente :

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \quad (4.29)$$

où $Z_t \sim \text{BB}(0, \sigma^2)$ et les ϕ_j et les θ_j sont des nombres complexes. On pose $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ et $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$. On suppose que $\phi(z)$ et $\theta(z)$ n'ont pas de zéros communs. Alors l'équation (4.29) fournit une représentation causale de la solution stationnaire au second ordre si et seulement si le polynôme $\phi(z) \neq 0$ pour $|z| \leq 1$. Cette solution a pour expression :

$$X_t = \sum_{k \geq 0} \psi_k Z_{t-k} \quad (4.30)$$

où la suite (ψ_k) est donnée par les coefficients du développement

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

qui converge dans $\{z \in \mathbb{C}, |z| \leq 1\}$.

Preuve. Le théorème 4.2.4 montre l'existence et l'unicité de la solution de l'équation (4.29) et vérifie

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

où la suite (ψ_k) est caractérisée par l'équation

$$\frac{\theta(z)}{\phi(z)} = \sum_{k=0}^{\infty} \psi_k z^k, \quad z \in \mathbb{C}, |z| = 1.$$

Si maintenant $\phi(z) \neq 0$ pour $|z| \leq 1$, alors, d'après le lemme 4.2.5, comme $r_1 = 0$, on a $\psi_k = 0$ pour $k < 0$ et (4.30) suit.

Réciproquement, si X est une solution stationnaire de (4.29) de la forme $X_t = \sum_{k \geq 0} \eta_k Z_{t-k}$ où $\sum_k |\eta_k| < \infty$, montrons que $\phi(z) \neq 0$ pour $|z| \leq 1$. En effet, puisque X est solution de (4.29), alors : $F_\phi X = F_\phi [F_\eta Z] = F_\theta Z$. D'après le lemme 4.2.3, $F_{\phi\eta} Z = F_\theta Z$. Posons, pour tout $k \in \mathbb{N}$, $\zeta_k =$

$\sum_{j \geq 0} \phi_j \eta_{k-j}$, où par convention, $\eta_\ell = 0$, si $\ell < 0$. On a alors, pour tout $t \in \mathbb{Z}$, $\sum_{k \geq 0} \zeta_k Z_{t-k} = \sum_{j=0}^q \theta_j Z_{t-j}$, avec la convention $\theta_0 = 1$. En multipliant les deux membres de cette équation par $Z_{t-\ell}$ et en prenant l'espérance, on déduit que $\zeta_\ell = \theta_\ell$, $\ell = 0, \dots, q$ et $\zeta_\ell = 0$, sinon. Ainsi, $\theta(z) = \phi(z)\eta(z)$, $|z| \leq 1$. Puisque θ et ϕ n'ont pas de racines communes et que $|\eta(z)| \leq \sum_{k \geq 0} |\eta_k| < \infty$, si $|z| \leq 1$, $\phi(z)$ ne s'annule pas sur le disque unité : $\{z, |z| \leq 1\}$, ce qui conclut la preuve du théorème 4.2.7. \square

Définition 4.2.8 (Représentation ARMA inversible). Sous les hypothèses du théorème 4.2.4, on dira que l'équation (4.24) fournit une représentation inversible de la solution stationnaire au second ordre (X_t) si $(X_t)_{t \in \mathbb{Z}}$ est un processus linéaire inversible par rapport à $(Z_t)_{t \in \mathbb{Z}}$.

Théorème 4.2.9 (ARMA(p, q) inversible). Soit l'équation récurrente :

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \quad (4.31)$$

où $Z_t \sim \text{BB}(0, \sigma^2)$ et les ϕ_j et les θ_j sont des nombres complexes. On pose $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ et $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$. On suppose que $\phi(z)$ et $\theta(z)$ n'ont pas de zéros communs. Alors l'équation (4.31) fournit une représentation inversible de la solution stationnaire au second ordre si et seulement si le polynôme $\theta(z) \neq 0$ pour $|z| \leq 1$. Cette solution est unique et a pour expression :

$$Z_t = \sum_{k \geq 0} \pi_k X_{t-k} \quad (4.32)$$

où la suite (π_k) est donnée par les coefficients du développement

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

qui converge dans $\{z \in \mathbb{C}, |z| \leq 1\}$.

La preuve de ce théorème est tout à fait analogue à celle du théorème 4.2.7 et n'est donc pas détaillée ici.

Un modèle ARMA(p, q) est causal et inversible lorsque les racines des polynômes $\phi(z)$ et $\theta(z)$ sont toutes situées à l'extérieur du disque unité. Dans ce cas, X_t et Z_t se déduisent mutuellement l'un de l'autre par des opérations de filtrage causal.

4.2.3.2 Calcul des covariances d'un processus ARMA(p, q) causal

Une première méthode consiste à utiliser l'expression (4.8) où la suite (ψ_k) se détermine de façon récurrente à partir de l'égalité $\psi(z)\theta(z) = \phi(z)$ par identification du terme en z^k . Pour les premiers termes on trouve :

$$\begin{aligned}
\psi_0 &= 1 \\
\psi_1 &= \theta_1 + \psi_0 \phi_1 \\
\psi_2 &= \theta_2 + \psi_0 \phi_2 + \psi_1 \phi_1 \\
&\dots
\end{aligned}$$

La seconde méthode utilise une formule de récurrence, vérifiée par la fonction d'autocovariance d'un processus ARMA(p, q), qui s'obtient en multipliant les deux membres de (4.24) par \bar{X}_{t-k} et en prenant l'espérance. On obtient :

$$\gamma(k) - \phi_1 \gamma(k-1) - \dots - \phi_p \gamma(k-p) = \sigma^2 \sum_{k \leq j \leq q} \theta_j \bar{\psi}_{j-k}, \quad 0 \leq k < \max(p, q+1) \quad (4.33)$$

$$\gamma(k) - \phi_1 \gamma(k-1) - \dots - \phi_p \gamma(k-p) = 0, \quad k \geq \max(p, q+1) \quad (4.34)$$

où nous avons utilisé la causalité du processus pour écrire que $\mathbb{E}[Z_t \bar{X}_{t-k}] = 0$ pour tout $k \geq 1$. Le calcul de la suite $\{\psi_k\}$ pour $k = 1, \dots, p$ se fait comme précédemment. En reportant ces valeurs dans (4.33) pour $0 \leq k \leq p$, on obtient $(p+1)$ équations linéaires aux $(p+1)$ inconnues $(\gamma(0), \dots, \gamma(p))$ que l'on peut résoudre. Pour déterminer les valeurs suivantes on utilise l'expression (4.34).

4.2.3.3 Densité spectrale d'un processus ARMA(p, q)

Théorème 4.2.10 (Densité spectrale d'un processus ARMA(p, q)).
Soit (X_t) un processus ARMA(p, q) (pas nécessairement causal ou inversible) i.e. la solution stationnaire de l'équation (4.24) où les polynômes $\theta(z)$ et $\phi(z)$ sont des polynômes de degré q et p n'ayant pas de zéros communs. Alors (X_t) possède une densité spectrale qui a pour expression :

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{|1 + \sum_{k=1}^q \theta_k e^{-ik\lambda}|^2}{|1 - \sum_{k=1}^p \phi_k e^{-ik\lambda}|^2}, \quad -\pi \leq \lambda \leq \pi. \quad (4.35)$$

Remarque 4.2.11. D'après le théorème 4.2.4, l'expression de f est bien définie puisque ϕ ne s'annule pas sur le cercle unité.

4.3 Filtrage des processus : théorie générale

Dans l'exemple suivant, on explique comment définir un filtrage linéaire d'un processus X uniquement à partir d'un élément de \mathcal{H}_∞^X .

Exemple 4.3.1 (Filtrage dans \mathcal{H}_∞^X). Soit $X = (X_t)_{t \in \mathbb{Z}}$ un processus stationnaire du second ordre de moyenne μ_X et de covariance γ_X . Soit de plus $Y_0 \in \mathcal{H}_\infty^X$. On considère un tableau de nombres complexes $(\alpha_{n,s})_{s \in \mathbb{Z}, n \geq 1}$: pour tout entier $n \in \mathbb{N}$, le cardinal de l'ensemble $\{s \in \mathbb{Z}, \alpha_{n,s} \neq 0\}$ est fini. Supposons que $\sum_{s \in \mathbb{Z}} \alpha_{n,s} X_{-s}$ converge vers Y_0 au sens $L^2(\Omega, \mathcal{A}, \mathbb{P})$, on a aussi, par stationnarité et en utilisant le critère de Cauchy, que pour tout $t \in \mathbb{Z}$, $\sum_{s \in \mathbb{Z}} \alpha_{n,s} X_{t-s}$ converge au sens $L^2(\Omega, \mathcal{A}, \mathbb{P})$ dans \mathcal{H}_∞^X , l'enveloppe linéaire du processus X . On note alors Y_t cette limite et on appelle $Y = (Y_t)$ le processus filtré correspondant. Par continuité de l'espérance et du produit scalaire sur $L^2(\Omega, \mathcal{A}, \mathbb{P})$, on obtient facilement que Y est stationnaire au second ordre de moyenne

$$\mu_Y = \mu_X \times \lim_{n \rightarrow \infty} \sum_{s \in \mathbb{Z}} \alpha_{n,s} ,$$

de fonction d'autocovariance

$$\gamma_Y(\tau) = \lim_{n \rightarrow \infty} \sum_{s \in \mathbb{Z}} \sum_{t \in \mathbb{Z}} \alpha_{n,s} \alpha_{n,t} \gamma_X(\tau - t + s) .$$

Un cas particulier de l'exemple précédent est celui où X est un bruit blanc faible.

Exemple 4.3.2 (Filtrage d'un bruit blanc). Plaçons-nous dans le cas de l'exemple 4.3.1 avec X bruit blanc faible. Dans ce cas, (X_t) forme une base hilbertienne de \mathcal{H}_∞^X et donc

$$\mathcal{H}_\infty^X = \left\{ \sum_{t \in \mathbb{Z}} \alpha_t X_t : (\alpha_t) \in \ell^2(\mathbb{Z}) \right\} ,$$

où $\ell^2(\mathbb{Z})$ est l'ensemble des suites $(x_t) \in \mathbb{C}^{\mathbb{Z}}$ telles que $\sum_t |\alpha_t|^2 < \infty$ et la convergence de la somme $\sum_{t \in \mathbb{Z}}$ est à comprendre au sens $L^2(\Omega, \mathcal{A}, \mathbb{P})$. On peut en particulier prendre $(\alpha_{n,t})_{t \in \mathbb{Z}, n \geq 1}$ de la forme $\alpha_{n,t} = \alpha_t \mathbb{1}(-n \leq t \leq n)$.

Reprenons l'exemple 4.3.1 du filtrage linéaire entièrement défini par un élément $Y_0 \in \mathcal{H}_\infty^X$. Quand, pour définir Y_0 , on peut prendre $(\alpha_{t,n})_{t \in \mathbb{Z}, n \geq 1}$ de la forme $\alpha_{t,n} = \alpha_t \mathbb{1}(-n \leq t \leq n)$ où $(\alpha_t) \in \mathbb{C}^{\mathbb{Z}}$, on notera $\sum_{s \in \mathbb{Z}} \alpha_s B^s$ l'opérateur correspondant,

$$Y_t = \sum_{s \in \mathbb{Z}} \alpha_s B^s(X_t) \Leftrightarrow Y_t = \sum_{s \in \mathbb{Z}} \alpha_s X_{t-s} \quad \text{pour tout } t \in \mathbb{Z} .$$

C'est bien le cas dans l'exemple 4.3.2 qui repose sur des hypothèse particulières pour X . C'était aussi le cas lorsque nous avons défini le filtre $\sum_k \alpha_k B^k$ avec $\sum_k |\alpha_k| < \infty$.

En toute généralité cependant une telle écriture n'est pas valable. Heureusement les représentations spectrales fournissent un outil permettant de traiter efficacement le cas général.

Définition 4.3.3 (Densité d'un champ aléatoire non corrélé). Soit W un champ aléatoire non corrélé sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ de mesure de base μ_W et α une fonction de $L^2(\mathbb{T}, \mu_W)$. Soit V un champ aléatoire non corrélé sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$. Nous dirons que le champ V admet par rapport au champ W la densité α , si pour toute fonction f vérifiant $\int_{\mathbb{T}} |f|^2 |\alpha|^2 d\mu_W$ nous avons

$$\int_{\mathbb{T}} f dV = \int_{\mathbb{T}} f \alpha dW,$$

Nous noterons de façon concise $dV = \alpha dW$.

Le théorème 3.4.5 montre que pour toute fonction f vérifiant $\int_{\mathbb{T}} |f|^2 |\alpha|^2 d\mu_W < \infty$,

$$\mathbb{V} \left(\int_{\mathbb{T}} f dV \right) = \mathbb{V} \left(\int_{\mathbb{T}} f \alpha dW \right) = \int_{\mathbb{T}} |f|^2 |\alpha|^2 d\mu_W.$$

Par conséquent, la mesure de base μ_V du champ V à une densité $|\alpha|^2$ par rapport à la mesure de base μ_W du champ W , *i.e.* $|\alpha|^2 = d\mu_V / d\mu_W$, μ_W -p.p.

Proposition 4.3.4. Soit W un champ aléatoire non corrélé sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ de mesure de base μ_W . Alors, pour toute fonction $\alpha \in L^2(\mathbb{T}, \mu_W)$, il existe une unique champ aléatoire non corrélé V sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ de densité α par rapport au champ W . La base μ_V du champ V a une densité $|\alpha|^2$ par rapport à la base μ_W du champ W .

Preuve. Considérons la mesure (finie) μ_V sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ de densité $|\alpha|^2$ par rapport à la mesure μ_W . L'application $A : f \mapsto A(f) = \alpha f$ de $L^2(\mathbb{T}, \mu_V)$ à valeurs dans $L^2(\mathbb{T}, \mu_W)$ est une isométrie car, pour tout $f, g \in L^2(\mathbb{T}, \mu_V)$,

$$\int_{\mathbb{T}} (\alpha f) \overline{(\alpha g)} d\mu_W = \int_{\mathbb{T}} f \bar{g} |\alpha|^2 d\mu_W = \int_{\mathbb{T}} f \bar{g} d\mu_V.$$

Puisque l'application $W : f \mapsto \int_{\mathbb{T}} f dW$ est une isométrie de $L^2(\mathbb{T}, \mu_W)$ à valeurs dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ (théorème 3.4.5), on en déduit que $W \circ A : f \mapsto \int_{\mathbb{T}} f \alpha dW$ est une isométrie de $L^2(\mathbb{T}, \mu_V)$ dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$. Le théorème 3.4.6 montre qu'il existe un unique champ aléatoire non corrélé sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ de base μ_V tel que, pour tout $f \in L^2(\mathbb{T}, \mu_V)$, $W \circ A(f) = \int_{\mathbb{T}} f dV$. \square

Soit X un processus stationnaire au second ordre. Nous notons \hat{X} le champ spectral de ce processus et μ_X la mesure spectrale associée. Nous allons appliquer le résultat ci-dessus au champ spectral de ce processus.

Théorème 4.3.5. *Soit X un processus centré stationnaire au second-ordre de champ spectral \hat{X} et de mesure spectrale μ_X . Soit $\alpha \in L^2(\mathbb{T}, \mu_X)$. Alors le processus centré*

$$Y_t = \int_{\mathbb{T}} e^{it\lambda} \alpha(\lambda) d\hat{X}(\lambda), \quad t \in \mathbb{Z}, \quad (4.36)$$

est stationnaire au second-ordre. Son champ spectral \hat{Y} admet comme densité α par rapport à \hat{X} , $d\hat{Y} = \alpha d\hat{X}$. Sa mesure spectrale μ_Y admet une densité $|\alpha|^2$ par rapport à μ_X .

Nous dirons que le processus Y est l'image du processus X par le filtre linéaire de fonction de transfert α ou plus simplement que Y est *filtré* de X par la *fonction de transfert* α .

Proposition 4.3.6. *Soit X un processus centré stationnaire au second-ordre de champ spectral \hat{X} et de mesure spectrale μ_X . Soit Y l'image de X par le filtre de fonction de transfert $\alpha \in L^2(\mathbb{T}, \mu_X)$. Alors, les enveloppes linéaires \mathcal{H}_{∞}^X et \mathcal{H}_{∞}^Y des processus X et Y vérifient : $\mathcal{H}_{\infty}^Y \subset \mathcal{H}_{\infty}^X$. Si $\alpha > 0$ μ_X -p.p. alors $\mathcal{H}_{\infty}^X = \mathcal{H}_{\infty}^Y$ et dans ce cas X est l'image de Y par le filtre de fonction de transfert α^{-1} .*

Preuve. Soit $A_{\alpha} : L^2(\mathbb{T}, \mu_Y) \rightarrow L^2(\mathbb{T}, \mu_X)$ l'isométrie

$$A_{\alpha} : f \mapsto \alpha f. \quad (4.37)$$

Les isométries $\hat{X} : L^2(\mathbb{T}, \mu_X) \rightarrow L^2(\Omega, \mathcal{A}, \mathbb{P})$ et $\hat{Y} : L^2(\mathbb{T}, \mu_Y) \rightarrow L^2(\Omega, \mathcal{A}, \mathbb{P})$ vérifient $\hat{Y} = \hat{X} \circ A_{\alpha}$ (théorème 4.3.5). Comme, par construction, les enveloppes linéaires \mathcal{H}_{∞}^X et \mathcal{H}_{∞}^Y sont les images de $L^2(\mathbb{T}, \mu_X)$ et $L^2(\mathbb{T}, \mu_Y)$ par les isométries \hat{X} et \hat{Y} (théorème 3.4.9), cette relation implique que $\mathcal{H}_{\infty}^Y \subset \mathcal{H}_{\infty}^X$.

Puisque toute isométrie est injective, la condition $\mathcal{H}_{\infty}^Y = \mathcal{H}_{\infty}^X$ implique que A_{α} est une surjection. Il est clair que A_{α} est surjective si et seulement si $\alpha > 0$, μ_X -p.p. Dans ce cas, A_{α} est un isomorphisme de $L^2(\mathbb{T}, \mu_Y)$ dans $L^2(\mathbb{T}, \mu_X)$ dont l'inverse est donné par $A_{\alpha}^{-1} : g \rightarrow g\alpha^{-1}$, pour tout $g \in L^2(\mathbb{T}, \mu_Y)$ et $d\hat{X} = \alpha^{-1}d\hat{Y}$. \square

Exemple 4.3.7 (Suite de l'exemple 4.3.1). Soit $(\alpha_{n,k})$ un tableau de coefficients complexes, tels que $\sum_{s \in \mathbb{Z}} \alpha_{n,s} X_{-s}$ converge au sens de $L^2(\Omega, \mathcal{A}, \mathbb{P})$. Notons $Y_0 \in \mathcal{H}_{\infty}^X$ cette limite. Pour tout $n \in \mathbb{N}$,

$$\sum_{s \in \mathbb{Z}} \alpha_{n,s} X_{-s} = \sum_{s \in \mathbb{Z}} \alpha_{n,s} \hat{X}(f_{-s}) = \hat{X} \left(\sum_{s \in \mathbb{Z}} \alpha_{n,s} f_{-s} \right),$$

où, pour $s \in \mathbb{Z}$, $f_s(\lambda) = e^{is\lambda}$ et \hat{X} est le champ spectral de X . Comme \hat{X} définit une isométrie de $L^2(\mathbb{T}, \mu_X)$ dans \mathcal{H}_{∞}^X , la convergence de $\sum_{s \in \mathbb{Z}} \alpha_{n,s} X_{-s}$ au sens de $L^2(\Omega, \mathcal{A}, \mathbb{P})$ est équivalente à la convergence de $\sum_{s \in \mathbb{Z}} \alpha_{n,s} f_{-s}$ au sens de $L^2(\mathbb{T}, \mu_X)$. Notons $\alpha_0 \in L^2(\mathbb{T}, \mu_X)$ cette limite. Nous avons donc $Y_0 = \hat{X}(h_0)$. Le processus (X_t) étant stationnaire, la convergence de

$\sum_{s \in \mathbb{Z}} \alpha_{n,s} X_{-s}$ dans \mathcal{H}_∞^X au sens de $L^2(\Omega, \mathcal{A}, \mathbb{P})$ implique que, pour tout $t \in \mathbb{Z}$, $\sum_{s \in \mathbb{Z}} \alpha_{n,s} X_{t-s}$ converge, dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ vers un élément $Y_t \in \mathcal{H}_\infty^X$. Comme nous l'avons montré dans l'exemple 4.3.1, (Y_t) est un processus stationnaire, de mesure spectrale de densité $|\alpha_0|^2$ par rapport à μ_X . La convergence de $\sum_{s \in \mathbb{Z}} \alpha_{n,s} X_{t-s}$ dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ est équivalente à la convergence de $\sum_{s \in \mathbb{Z}} \alpha_{n,s} f_{t-s}$ dans $L^2(\mathbb{T}, \mu_X)$. Notons α_t cette limite. Comme, pour tout $n \in \mathbb{N}$, $\sum_{s \in \mathbb{Z}} \alpha_{n,s} f_{t-s} = f_t \sum_{s \in \mathbb{Z}} \alpha_{n,s} f_{-s}$, nous avons donc $\alpha_t = f_t \alpha_0$ et

$$Y_t = \int_{\mathbb{T}} e^{it\lambda} \alpha_0(\lambda) d\hat{X}(d\lambda).$$

Par conséquent, $d\hat{Y} = \alpha_0 d\hat{X}$ et Y est le filtrage de X par la fonction de transfert α_0 .

Soit (α_k) une suite que telle que $\sum_s |\alpha_s| < \infty$. Posons $\alpha_{n,s} = \alpha_s \mathbb{1}_{|s| \leq n}$. Pour tout processus X de mesure spectrale μ_X , la suite $(\alpha_{n,s})$ converge dans $L^2(\mathbb{T}, \mu_X)$ et nous pouvons toujours définir un processus Y filtré de X par la fonction de transfert $\sum_{s \in \mathbb{Z}} \alpha_s e^{-i\lambda s}$. Ceci nous permet de retrouver de façon directe les résultats obtenus par une construction explicite dans le théorème 4.1.2.

Notons $\mathcal{S}(\Omega, \mathcal{A}, \mathbb{P})$ l'ensemble des processus indexés par \mathbb{Z} et à valeurs complexes stationnaires au second ordre. A toute fonction mesurable $\alpha : \mathbb{T} \rightarrow \mathbb{C}$, on associe un opérateur F_α de domaine $\mathcal{D}[\alpha] \subset \mathcal{S}(\Omega, \mathcal{A}, \mathbb{P})$ qui à $X \in \mathcal{D}[\alpha]$ associe $Y = F_\alpha X$ qui est le filtré de X par la fonction de transfert α .

Par construction, $\mathcal{D}[\alpha]$ est le sous-ensemble des processus $X \in \mathcal{S}(\Omega, \mathcal{A}, \mathbb{P})$ dont la mesure spectrale μ_X vérifie $\int_{\mathbb{T}} |\alpha|^2 d\mu_X < \infty$. Si la fonction α est mesurable bornée, alors $\mathcal{D}[\alpha] = \mathcal{S}(\Omega, \mathcal{A}, \mathbb{P})$. Remarquons (proposition 4.3.4) que s'il existe une fonction mesurable β sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ telle que $X \in \mathcal{D}[\alpha] \cap \mathcal{D}[\beta]$ et $F_\alpha X = F_\beta X$, alors $\alpha = \beta$, μ_X -p.p. .

Si α et β sont deux fonction mesurables sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$, le domaine de définition du produit de composition des opérateurs F_α et F_β , $F_\alpha \circ F_\beta$ est

$$\mathcal{D}[\alpha, \beta] \stackrel{\text{def}}{=} \{X \in \mathcal{S}(\Omega, \mathcal{A}, \mathbb{P}), X \in \mathcal{D}[\beta], F_\beta X \in \mathcal{D}[\alpha]\}. \quad (4.38)$$

On remarque qu'en général, $\mathcal{D}[\alpha, \beta] \neq \mathcal{D}[\beta, \alpha]$. Prenons par exemple $\alpha(\lambda) = (1 - e^{-i\lambda})^{-1}$ et $\beta(\lambda) = (1 - e^{-i\lambda})$. Nous avons $\mathcal{D}[\alpha, \beta] = \mathcal{S}(\Omega, \mathcal{A}, \mathbb{P})$ alors que $\mathcal{D}[\beta, \alpha]$ est le sous-ensemble des processus $X \in \mathcal{S}(\Omega, \mathcal{A}, \mathbb{P})$ de mesure spectrale μ_X satisfaisant $\int_{\mathbb{T}} |1 - e^{-i\lambda}|^2 \mu_X(d\lambda) < \infty$.

Proposition 4.3.8. *Soient α et β deux fonctions mesurables de $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$. Alors $\mathcal{D}[\alpha, \beta] \subset \mathcal{D}[\alpha\beta]$ et pour tout $X \in \mathcal{D}[\alpha, \beta]$, $F_\alpha \circ F_\beta X = F_{\alpha\beta} X$. En particulier, si $X \in \mathcal{D}[\alpha, \beta] \cap \mathcal{D}[\beta, \alpha]$, $F_\alpha \circ F_\beta X = F_\beta \circ F_\alpha X = F_{\alpha\beta} X$.*

Preuve. La condition $X \in \mathcal{D}[\alpha]$ implique que $\beta \in L^2(\mathbb{T}, \mu_X)$ et la condition $F_\beta X \in \mathcal{D}[\beta]$ implique que $\int_{\mathbb{T}} |\alpha|^2 |\beta|^2 d\mu_X < \infty$; par conséquent $\alpha\beta \in L^2(\mathbb{T}, \mu_X)$, ce qui montre que $X \in \mathcal{D}[\alpha\beta]$.

Posons $U = F_\alpha X$, $V = F_\beta U$, $Y = F_{\alpha\beta} X$. On note \hat{X} le champ spectral de X et μ_X la mesure spectrale associée. Le théorème 4.3.5 montre que \hat{U} a une densité α par rapport à \hat{X} , $d\hat{U} = \alpha d\hat{X}$ et que \hat{V} a une densité β par rapport à \hat{U} . La mesure de base μ_U du champ \hat{U} admet une densité $|\alpha|^2$ par rapport à μ_X ; la mesure de base μ_V du champ \hat{V} admet une densité $|\beta|^2$ par rapport à μ_U , et donc une densité $|\alpha\beta|^2$ par rapport à μ_X . Par construction, \hat{U} est l'isométrie de $L^2(\mathbb{T}, \mu_U)$ dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ telle que $\hat{U} = \hat{X} \circ A_\alpha$, où A_α est définie par (4.37); \hat{V} est l'isométrie de $L^2(\mathbb{T}, \mu_V)$ dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$, $\hat{V} = \hat{V} \circ A_\beta$, où A_β est l'isométrie de $L^2(\mathbb{T}, \mu_V)$ dans $L^2(\mathbb{T}, \mu_U)$, $A_\beta : f \mapsto \beta f$. Donc

$$\hat{V} = \hat{X} \circ A_\alpha \circ A_\beta .$$

et pour tout $f \in L^2(\mathbb{T}, \mu_V)$, $\hat{V}(f) = \hat{X}(\alpha\beta f)$. De même, la mesure de base μ_Y du champ \hat{Y} admet une densité $|\alpha\beta|^2$ par rapport à μ_X et donc $\mu_Y = \mu_V$. le champ spectrale \hat{Y} est l'isométrie de $L^2(\mathbb{T}, \mu_Y)$ telle que

$$\hat{Y} = \hat{X} \circ A_{\alpha\beta} ,$$

où $A_{\alpha\beta}$ est l'isométrie de $L^2(\mathbb{T}, \mu_Y)$ dans $L^2(\mathbb{T}, \mu_X)$, $A_{\alpha\beta} : f \mapsto \alpha\beta f$. On a donc, pour tout $f \in L^2(\mathbb{T}, \mu_Y)$, $\hat{Y}(f) = \hat{X}(\alpha\beta f)$, ce qui montre que $\hat{V} = \hat{Y}$. \square

Définition 4.3.9. Soit α une fonction de transfert (*i.e.* une fonction mesurable de $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$). Nous appelons *inverse* de la fonction de transfert α la fonction $\alpha^{[-1]}(\lambda) = \alpha^{-1}(\lambda) \mathbb{1}_{\{\alpha(\lambda) \neq 0\}}$.

Le domaine de définition de l'inverse $\mathcal{D}[\alpha^{[-1]}]$ est l'ensemble des processus $X \in \mathcal{S}(\Omega, \mathcal{A}, \mathbb{P})$ de mesure spectrale μ_X vérifiant $\int_{\mathbb{T}} \mathbb{1}_{\{\alpha(\lambda) \neq 0\}} |\alpha(\lambda)|^{-2}(\lambda) d\lambda < \infty$. Si la fonction α ne s'annule pas sur \mathbb{T} alors $\mathcal{D}[\alpha^{[-1]}] = \mathcal{D}[\alpha^{-1}]$. Le domaine de définition du produit de composition $F_{\alpha^{[-1]}} \circ F_\alpha$ est $\mathcal{D}[\alpha]$, $F_{\alpha^{[-1]}} \circ F_\alpha = F_{\mathbb{1}_{\{\alpha \neq 0\}}}$. La relation $X = F_{\alpha^{[-1]}} \circ F_\alpha X$ n'est donc valable que pour $X \in \mathcal{D}[\alpha]$ et pour des fonctions $\alpha \neq 0$, μ_X -p.p. .

Problèmes

4.3.10 (Stationnarité, causalité, inversibilité des processus ARMA).

$\{Z_t\}_t$ désigne un bruit blanc. Pour chacune des équations suivantes, existe-t-il un processus stationnaire $\{X_t\}_t$ solution ? Si oui, est-il causal ? est-il inversible ?

- a) $X_t + 0.2X_{t-1} - 0.48X_{t-2} = Z_t$.
- b) $X_t + 1.9X_{t-1} + 0.88X_{t-2} = Z_t + 0.2Z_{t-1} + 0.7Z_{t-2}$.
- c) $X_t + 0.6X_{t-2} = Z_t + 1.2Z_{t-1}$.
- d) $X_t + 1.8X_{t-1} + 0.81X_{t-2} = Z_t$.

4.3.11 (Processus AR(1) et MA(1)). Soit $\{X_t\}$ le processus solution de l'équation $X_t = \phi X_{t-1} + Z_t$, où $\{Z_t\}$ est un bruit blanc de variance σ_Z^2 et $|\phi| > 1$.

1. Cette solution est-elle stationnaire ? causale ?
2. Montrer que $\{X_t\}$ vérifie aussi les équations $X_t = \phi^{-1}X_{t-1} + \tilde{Z}_t$, où $\{\tilde{Z}_t\}$ est un bruit blanc dont on déterminera la variance $\sigma_{\tilde{Z}}^2$.
3. Conclure.
4. Faire de même lorsque $\{X_t\}$ est solution de $X_t = Z_t + \theta Z_{t-1}$ et $|\theta| > 1$.
5. En utilisant ce que l'on a vu dans les questions précédentes, proposer une représentation causale et inversible de $\{X_t\}$ solution de $X_t - 2X_{t-1} = Z_t + 4Z_{t-1}$ où $\{Z_t\}$ est un bruit blanc de variance σ_Z^2 .
6. Lorsque $|\phi| = 1$ existe-t-il une solution stationnaire à l'équation $X_t = \phi X_{t-1} + Z_t$, où $\{Z_t\}$ est un bruit blanc de variance σ_Z^2 ?

4.3.12 (Processus ARMA(1,1)). On considère l'équation

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

où $\{Z_t\}$ est un bruit blanc de moyenne nulle et de variance σ^2 et ϕ et θ sont des réels.

1. A quelles conditions sur ϕ et θ existe-t-il un processus $\{X_t\}$ stationnaire solution de l'équation ci-dessus ?
2. A quelles conditions cette solution est-elle causale ? Dans la suite, on supposera que ces conditions sont vérifiées.
3. Donner une représentation de la solution $\{X_t\}$ sous forme d'une somme infinie $\sum_{k \in \mathbb{Z}} \psi_k Z_{t-k}$. On justifiera la convergence de cette somme, on précisera en quel sens elle converge et on donnera la valeur des coefficients $\{\psi_k\}_{k \in \mathbb{Z}}$.
4. Calculer la fonction d'autocorrélation de $\{X_t\}$.

4.3.13. On considère $\{X_t\}$ et $\{Y_s\}$ deux processus centrés et indépendants *i.e* X_t et Y_s sont indépendants pour tout t, s . On définit le processus produit $\{Z_t\}$ par $Z_t = X_t Y_t$. Soient $\{\varepsilon_t\}$ et $\{\eta_t\}$ deux bruits blancs. On suppose que $\{X_t\}$ et $\{Y_t\}$ sont des ARMA(1,1) inversibles de paramètres respectifs ϕ_1, θ_1 et ϕ_2, θ_2 et de bruits blancs respectifs $\{\varepsilon_t\}$ et $\{\eta_t\}$.

1. Montrer que les processus $\{\varepsilon_t\}$ et $\{\eta_t\}$ sont décorrélés.
2. A quelles conditions peut-on écrire $X_t = \sum_{j \geq 0} \psi_j \varepsilon_{t-j}$ et $Y_t = \sum_{j \geq 0} \tilde{\psi}_j \eta_{t-j}$?
Supposons ces conditions vérifiées dans la suite.
3. Calculer la fonction d'autocorrélation du processus $\{Z_t\}$.

4.3.14 (Calcul de la fonction d'auto-covariance d'un ARMA(2,1)).

On considère le processus $\{X_t\}$ solution de :

$$(1 - B + B^2/4)X_t = (1 + B)Z_t ,$$

où $\{Z_t\}$ est un bruit blanc de variance σ^2 .

1. Montrer que l'on peut écrire X_t sous la forme $X_t = \sum_{k \geq 0} \psi_k Z_{t-k}$.
2. Calculer les coefficients $\{\psi_k\}_{k \geq 0}$.
3. Calculer la fonction d'auto-covariance de $\{X_t\}$.

4.3.15 (Processus ARMA(1,1)). On considère la solution stationnaire de l'équation :

$$X_t + \phi_1 X_{t-1} = Z_t + \theta_1 Z_{t-1} \quad (4.39)$$

où Z_t est centré, blanc de variance σ^2 . On pose $\psi(z) = \theta(z)/\phi(z)$ avec $\phi(z) = 1 + \phi_1 z$ et $\theta(z) = 1 + \theta_1 z$. On suppose $|\phi_1| < 1$. On note $\gamma(h)$ la fonction d'autocovariance (quand elle existe) de X_t .

1. Discuter suivant la valeur de θ_1 si la solution de l'équation (4.39) est causale et/ou inversible. On suppose dans la suite que $\theta_1 \neq \phi_1$.
2. En multipliant chaque membre de l'équation (4.39) par Z_t , puis en prenant l'espérance, déterminer l'expression de $\mathbb{E}[X_t Z_t]$ en fonction de σ^2 .
3. En multipliant chaque membre de l'équation (4.39) par Z_{t-1} , puis en prenant l'espérance, déterminer l'expression de $\mathbb{E}[X_t Z_{t-1}]$ en fonction de σ^2, ϕ_1 et θ_1 .
4. En multipliant chaque membre de l'équation (4.39) successivement par X_t puis X_{t-1} , déduire que

$$A \begin{bmatrix} \gamma(0) \\ \gamma(1) \end{bmatrix} = b$$

où A et b sont une matrice et un vecteur dont on donnera les expressions.

5. Pour $h \geq 2$, en multipliant par X_{t-h} chaque membre de l'équation (4.39) et en prenant l'espérance, déterminer une équation récurrente, dépendant de θ_1, ϕ_1 et σ^2 , que vérifie $\gamma(h)$ et que l'on résoudra.

6. Donner l'expression, en fonction de σ^2 , ϕ_1 et θ_1 , de la densité spectrale de X_t .

Chapitre 5

Prédiction des processus stationnaires au second ordre

5.1 Prédiction linéaire de processus stationnaires

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire au second ordre à valeurs réelles, **d'espérance nulle** et de fonction d'autocovariance $\gamma(h) = \text{cov}(X_h, X_0)$. On cherche à *prédire* la valeur du processus à la date t à partir d'une combinaison linéaire des p derniers échantillons du passé X_{t-1}, \dots, X_{t-p} . La meilleure combinaison linéaire (*i.e.* le prédicteur linéaire optimal) est la projection orthogonale de X_t sur $\mathcal{H}_{t-1,p}$ notée $\text{proj}(X_t | \mathcal{H}_{t-1,p})$, où $\mathcal{H}_{t-1,p}$ est défini par :

$$\mathcal{H}_{t-1,p} = \text{Vect}(X_{t-1}, X_{t-2}, \dots, X_{t-p}) . \quad (5.1)$$

Les indices dans la notation $\mathcal{H}_{t-1,p}$ doivent être compris ainsi : $\mathcal{H}_{t-1,p}$ est le sous-espace vectoriel engendré par les p observations précédant X_{t-1} à savoir X_{t-1}, \dots, X_{t-p} . D'après le théorème 2.4.1,

$$\text{proj}(X_t | \mathcal{H}_{t-1,p}) = \sum_{k=1}^p \phi_{k,p} X_{t-k} , \quad (5.2)$$

où les coefficients $(\phi_{k,p})_{1 \leq k \leq p}$ satisfont

$$\left\langle X_t - \sum_{k=1}^p \phi_{k,p} X_{t-k}, X_{t-j} \right\rangle = 0 , \quad j = 1, \dots, p , \quad (5.3)$$

la notation $\langle \cdot, \cdot \rangle$ correspondant au produit scalaire dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ défini pour X et Y dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ par $\langle X, Y \rangle = \mathbb{E}[XY]$. L'équation (5.3) se réécrit encore sous la forme

$$\langle X_t, X_{t-j} \rangle = \sum_{k=1}^p \phi_{k,p} \langle X_{t-k}, X_{t-j} \rangle , \quad j = 1, \dots, p , \quad (5.4)$$

soit encore

$$\sum_{k=1}^p \phi_{k,p} \gamma(k-j) = \gamma(j) , \quad j = 1, \dots, p . \quad (5.5)$$

En posant Γ_p la matrice de covariance du vecteur $(X_{t-1}, \dots, X_{t-p})$ définie par

$$\Gamma_p = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \vdots \\ \vdots & \ddots & \ddots & \ddots \\ \vdots & & & \gamma(1) \\ \gamma(p-1) & \gamma(p-2) & \cdots & \gamma(1) & \gamma(0) \end{bmatrix} ,$$

on peut réécrire (5.5) comme suit :

$$\Gamma_p \phi_p = \gamma_p, \quad (5.6)$$

où $\phi_p = (\phi_{1,p}, \dots, \phi_{p,p})^T$ et $\gamma_p = (\gamma(1), \gamma(2), \dots, \gamma(p))^T$.

Définition 5.1.1. Nous appellerons dans la suite *erreur de prédiction directe* d'ordre p ou *innovation partielle* d'ordre p le processus :

$$\epsilon_{t,p}^+ = X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p}) = X_t - \sum_{k=1}^p \phi_{k,p} X_{t-k}. \quad (5.7)$$

La variance de l'erreur de prédiction directe d'ordre p est notée σ_p^2 et définie par

$$\sigma_p^2 = \|X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p})\|^2 = \mathbb{E}[|X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p})|^2]. \quad (5.8)$$

D'après (5.2) et la proposition 2.4.3, la variance de l'erreur de prédiction directe d'ordre p a pour expression :

$$\sigma_p^2 = \langle X_t, X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p}) \rangle = \gamma(0) - \sum_{k=1}^p \phi_{k,p} \gamma(k) = \gamma(0) - \phi_p^T \gamma_p. \quad (5.9)$$

Les équations (5.6) et (5.9) sont appelées les *équations de Yule-Walker*.

Notons que (5.6) a une unique solution si et seulement si la matrice Γ_p est inversible auquel cas la solution vaut :

$$\phi_p = \Gamma_p^{-1} \gamma_p. \quad (5.10)$$

La proposition 3.3.9 fournit les conditions suffisantes assurant que Γ_p est inversible pour tout p . Une autre preuve de la proposition 3.3.9 résultat est proposée dans le problème 5.3.3. On a ainsi des conditions sous lesquelles on peut calculer le prédicteur de X_t à partir de X_{t-1}, \dots, X_{t-p} .

Exemple 5.1.2 (Cas d'un processus AR(m) causal). Soit (X_t) le processus AR(m) causal solution de l'équation récurrente :

$$X_t = \phi_1 X_{t-1} + \dots + \phi_m X_{t-m} + Z_t, \quad (5.11)$$

où $Z_t \sim \text{BB}(0, \sigma^2)$ et où $\phi(z) = 1 - \sum_{k=1}^m \phi_k z^k \neq 0$ lorsque $|z| \leq 1$. Dans ce cas, pour tout $p \geq m$:

$$\phi_{k,p} = \begin{cases} \phi_k, & \text{lorsque } 1 \leq k \leq m, \\ 0, & \text{lorsque } m < k \leq p. \end{cases}$$

En effet, (X_t) étant causal on a, pour tout $h \geq 1$, $\mathbb{E}[Z_t X_{t-h}] = 0$ et donc, d'après (5.11), $\mathbb{E}[(X_t - \sum_{k=1}^m \phi_k X_{t-k}) X_{t-h}] = 0$. Ainsi, pour tout $p \geq m$, $\sum_{k=1}^m \phi_k X_{t-k} \in \mathcal{H}_{t-1,p}$ et $(X_t - \sum_{k=1}^m \phi_k X_{t-k}) \perp \mathcal{H}_{t-1,p}$ et donc, d'après le théorème 2.4.1, pour tout $p \geq m$,

$$\sum_{k=1}^m \phi_k X_{t-k} = \text{proj} (X_t | \mathcal{H}_{t-1,p}) .$$

Les coefficients de prédiction d'un processus stationnaire au second ordre fournissent une décomposition particulière de la matrice de covariance Γ_{p+1} sous la forme d'un produit de matrices triangulaires explicitée dans le théorème 5.1.3.

Théorème 5.1.3. *Soit (X_t) un processus stationnaire au second ordre, centré, de fonction d'autocovariance $\gamma(h)$. On note :*

$$A_{p+1} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -\phi_{1,1} & 1 & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & 0 \\ -\phi_{p,p} & -\phi_{p-1,p} & \cdots & -\phi_{1,p} & 1 \end{bmatrix} \quad \text{et} \quad D_{p+1} = \begin{bmatrix} \sigma_0^2 & 0 & \cdots & 0 \\ 0 & \sigma_1^2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \sigma_p^2 \end{bmatrix} ,$$

où les coefficients $(\phi_{k,p})_{1 \leq k \leq p}$ et $(\sigma_k^2)_{1 \leq k \leq p}$ sont respectivement définis dans (5.2) et (5.8). On a alors :

$$\Gamma_{p+1} = A_{p+1}^{-1} D_{p+1} (A_{p+1}^T)^{-1} . \quad (5.12)$$

Preuve. Pour simplifier les notations, posons $\mathcal{H}_k = \mathcal{H}_{k,k} = \text{Vect} (X_k, \dots, X_1)$ et montrons tout d'abord que, pour $k \neq \ell$, nous avons :

$$\langle X_k - \text{proj} (X_k | \mathcal{H}_{k-1}), X_\ell - \text{proj} (X_\ell | \mathcal{H}_{\ell-1}) \rangle = 0 . \quad (5.13)$$

En effet, pour $k < \ell$, on a $X_k - \text{proj} (X_k | \mathcal{H}_{k-1}) \in \mathcal{H}_k \subseteq \mathcal{H}_{\ell-1}$ et $X_\ell - \text{proj} (X_\ell | \mathcal{H}_{\ell-1}) \perp \mathcal{H}_{\ell-1}$. D'autre part, si on note \mathbf{X}_{p+1} le vecteur : $(X_1, \dots, X_{p+1})^T$, alors, par définition des coefficients de prédiction (5.2), on peut écrire :

$$A_{p+1} \mathbf{X}_{p+1} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -\phi_{1,1} & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ -\phi_{p,p} & -\phi_{p-1,p} & \cdots & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{p+1} \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 - \text{proj} (X_2 | \mathcal{H}_1) \\ \vdots \\ X_{p+1} - \text{proj} (X_{p+1} | \mathcal{H}_p) \end{bmatrix} ,$$

qui donne :

$$\mathbb{E} [A_{p+1} \mathbf{X}_{p+1} \mathbf{X}_{p+1}^T A_{p+1}^T] = D_{p+1} ,$$

d'après (5.13) et (5.8). Par ailleurs,

$$\mathbb{E} [A_{p+1} \mathbf{X}_{p+1} \mathbf{X}_{p+1}^T A_{p+1}^T] = A_{p+1} \Gamma_{p+1} A_{p+1}^T ,$$

ce qui démontre (5.12) puisque la matrice A_{p+1} est inversible, son déterminant étant égal à 1. \square

D'après l'équation (5.12) lorsque la matrice Γ_{p+1} est inversible, la variance $\sigma_p^2 = \|\epsilon_{t,p}^+\|^2$ est strictement positive. D'autre part, la suite σ_p^2 est décroissante. En effet, par définition de $\mathcal{H}_{t-1,p}$, $\mathcal{H}_{t-1,p}$ est inclus dans $\mathcal{H}_{t-1,p+1}$ donc $\text{proj}(X_{p+1} | \mathcal{H}_{t-1,p})$ est dans $\mathcal{H}_{t-1,p+1}$. On déduit donc du théorème 2.4.1 que $\sigma_{p+1}^2 \leq \sigma_p^2$. La suite (σ_p^2) étant décroissante et minorée, elle possède donc une limite quand p tend vers l'infini. Cela conduit à la définition suivante, dont nous verrons au paragraphe 6.1 qu'elle joue un rôle fondamental dans la décomposition des processus stationnaires au second ordre.

Définition 5.1.4 (Processus régulier/déterministe). Soit $(X_t)_{t \in \mathbb{Z}}$ un processus aléatoire stationnaire au second ordre. On note $\sigma^2 = \lim_{p \rightarrow \infty} \sigma_p^2$ où σ_p^2 est la variance de l'innovation partielle d'ordre p . On dit que le processus (X_t) est *régulier* si $\sigma^2 > 0$ et *déterministe* si $\sigma^2 = 0$.

Par ailleurs, nous pouvons remarquer que le problème de la recherche des coefficients de prédiction pour un processus stationnaire au second ordre se ramène à celui de la minimisation de l'intégrale :

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |\psi(e^{-i\lambda})|^2 \nu_X(d\lambda)$$

sur l'ensemble \mathcal{P}_p des polynômes à coefficients réels de degré p de la forme $\psi(z) = 1 + \psi_1 z + \dots + \psi_p z^p$. En effet, en utilisant la relation (4.4) de filtrage des mesures spectrales, on peut écrire que la variance de $\|\epsilon_{t,p}^+\|^2$, qui minimise l'erreur de prédiction, a pour expression :

$$\sigma_p^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\phi_p(e^{-i\lambda})|^2 \nu_X(d\lambda) \quad (5.14)$$

où :

$$\phi_p(z) = 1 - \sum_{k=1}^p \phi_{k,p} z^k$$

désigne le *polynôme prédicteur d'ordre p* .

Théorème 5.1.5. Si $\{X_t\}$ est un processus régulier, alors, pour tout p , $\phi_p(z) \neq 0$ pour $|z| \leq 1$. Tous les zéros des polynômes prédicteurs sont à l'extérieur du cercle unité.

Preuve (Preuve du théorème 5.1.5). Nous allons tout d'abord montrer que le prédicteur optimal n'a pas de racines sur le cercle unité. Raisonnons par contradiction. Supposons que le polynôme $\phi_p(z)$ ait deux racines complexes conjuguées, de la forme $\exp(\pm i\theta)$, sur le cercle unité (on traite de façon similaire le cas de racines réelles, $\theta = 0$ ou π). Nous pouvons écrire :

$$\phi_p(z) = \phi_p^*(z)(1 - 2\cos(\theta)z + z^2)$$

On note $\bar{\nu}_X(d\lambda) = \nu_X(d\lambda)|\phi_p^*(e^{-i\lambda})|^2$. $\bar{\nu}_X$ est une mesure positive sur $[-\pi, \pi]$ de masse finie. On note $\bar{\gamma}(\tau)$ la suite des coefficients de Fourier associés à $\bar{\nu}_X$:

$$\bar{\gamma}(\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\tau\lambda} \bar{\nu}_X(d\lambda) .$$

Nous avons donc :

$$\begin{aligned} \sigma_p^2 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} (1 - 2\cos(\theta)e^{-i\lambda} + e^{-2i\lambda}) \bar{\nu}_X(d\lambda) \\ &= \inf_{\psi \in \mathcal{P}_2} \frac{1}{2\pi} \int_{-\pi}^{\pi} |1 + \psi_1 e^{-i\lambda} + \psi_2 e^{-2i\lambda}|^2 \bar{\nu}_X(d\lambda) . \end{aligned}$$

La minimisation de σ_p^2 par rapport à ψ_1 et ψ_2 est équivalente à la résolution des équations de Yule-Walker à l'ordre $p = 2$ pour la suite des covariances $\bar{\gamma}(h)$. Par conséquent la suite des coefficients $\{1, -2\cos(\theta), 1\}$ doit vérifier l'équation :

$$\begin{bmatrix} \bar{\gamma}(0) & \bar{\gamma}(1) & \bar{\gamma}(2) \\ \bar{\gamma}(1) & \bar{\gamma}(0) & \bar{\gamma}(1) \\ \bar{\gamma}(2) & \bar{\gamma}(1) & \bar{\gamma}(0) \end{bmatrix} \begin{bmatrix} 1 \\ -2\cos(\theta) \\ 1 \end{bmatrix} = \begin{bmatrix} \sigma_p^2 \\ 0 \\ 0 \end{bmatrix}$$

De cette équation il s'en suit (les première et troisième lignes sont égales) que $\sigma_p^2 = 0$, ce qui est contraire à l'hypothèse que le processus est régulier.

Démontrons maintenant que les racines des polynômes prédicteurs sont toutes *strictement à l'extérieur du cercle unité*. Raisonnons encore par l'absurde. Supposons que le polynôme prédicteur à l'ordre p ait m racines $\{a_k, |a_k| < 1, 1 \leq k \leq m\}$ à l'intérieur du cercle unité et $(p - m)$ racines $\{b_\ell, |b_\ell| > 1, 1 \leq \ell \leq p - m\}$ à l'extérieur du cercle unité. Le polynôme prédicteur à l'ordre p s'écrit donc :

$$\phi_p(z) = \prod_{k=1}^m (1 - a_k^{-1}z) \prod_{\ell=1}^{p-m} (1 - b_\ell^{-1}z) .$$

Considérons alors le polynôme :

$$\bar{\phi}_p(z) = \prod_{k=1}^m (1 - a_k^* z) \prod_{\ell=1}^{p-m} (1 - b_\ell^{-1}z) .$$

Il a d'une part toutes ses racines strictement à l'extérieur du cercle unité et d'autre part il vérifie $|\bar{\phi}_p(e^{-i\lambda})|^2 < |\phi_p(e^{-i\lambda})|^2$. On a en effet $|1 - a_k^* e^{-i\lambda}| = |1 - a_k e^{i\lambda}| = |a_k| |1 - a_k^{-1} e^{-i\lambda}|$ et donc $|\bar{\phi}_p(e^{-i\lambda})|^2 = (\prod_{k=1}^m |a_k|^2) |\phi_p(e^{-i\lambda})|^2$, ce qui démontre le résultat annoncé puisque $|a_k| < 1$. On en déduit alors que :

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |\bar{\phi}_p(e^{-i\lambda})|^2 \nu_X(d\lambda) < \sigma_p^2,$$

ce qui contredit que $\phi_p(z) = \inf_{\psi \in \mathcal{P}_p} (2\pi)^{-1} \int_{-\pi}^{\pi} |\psi(e^{-i\lambda})|^2 \nu_X(d\lambda)$. \square

Une conséquence directe du théorème 5.1.5 est qu'à toute matrice de covariance de type défini positif, de dimension $(p+1) \times (p+1)$, on peut associer un processus AR(p) causal dont les $(p+1)$ premiers coefficients de covariance sont précisément la première ligne de cette matrice. Ce résultat n'est pas général. Ainsi il existe bien un processus AR(2) causal ayant $\gamma(0) = 1$ et $\gamma(1) = \rho$, comme premiers coefficients de covariance, à condition toutefois que la matrice de covariance soit positive c'est-à-dire que $|\rho| < 1$, tandis qu'il n'existe pas, pour cette même matrice de processus MA(1). Il faut en effet, en plus du caractère positif, que $|\rho| \leq 1/2$ (voir exemple 3.2.10).

5.2 Algorithme de Levinson-Durbin

La solution directe du système des équations de Yule-Walker requiert de l'ordre de p^3 opérations : la résolution classique de ce système implique en effet la décomposition de la matrice Γ_p sous la forme du produit d'une matrice triangulaire inférieure et de sa transposée, $\Gamma_p = L_p L_p^T$ (décomposition de Choleski) et la résolution par substitution de deux systèmes triangulaires. Cette procédure peut s'avérer coûteuse lorsque l'ordre de prédiction est grand (on utilise généralement des ordres de prédiction de l'ordre de quelques dizaines à quelques centaines), ou lorsque, à des fins de modélisation, on est amené à évaluer la qualité de prédiction pour différents horizons de prédiction. L'algorithme de Levinson-Durbin exploite la structure géométrique particulière des processus stationnaires au second ordre pour établir une formule de récurrence donnant les coefficients de prédiction à l'ordre $(p+1)$ à partir des coefficients de prédiction obtenus à l'ordre p . Il fournit également une relation de récurrence entre l'erreur de prédiction directe à l'ordre $p+1$ et l'erreur de prédiction directe à l'ordre p .

On supposera dans toute cette partie que **Γ_p est inversible pour tout $p \geq 1$** .

Supposons que les coefficients de prédiction linéaire et la variance de l'erreur de prédiction directe à l'ordre p , pour $p \geq 0$, sont connus :

$$\text{proj}(X_t | \mathcal{H}_{t-1,p}) = \sum_{k=1}^p \phi_{k,p} X_{t-k} \quad \text{et} \quad \sigma_p^2 = \|X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p})\|^2,$$

et déterminons, à partir de la projection à l'ordre p de X_t , la projection de X_t à l'ordre $p+1$ sur le sous-espace $\mathcal{H}_{t-1,p+1} = \text{Vect}(X_{t-1}, \dots, X_{t-p-1})$.

Pour cela, on décompose cet espace en somme orthogonale de la façon suivante :

$$\begin{aligned}\mathcal{H}_{t-1,p+1} &= \mathcal{H}_{t-1,p} \oplus^{\perp} \text{Vect}(X_{t-p-1} - \text{proj}(X_{t-p-1} | \mathcal{H}_{t-1,p})) \\ &= \mathcal{H}_{t-1,p} \oplus^{\perp} \text{Vect}(\epsilon_{t-p-1,p}^-),\end{aligned}$$

où, de façon générale, $\epsilon_{t,p}^-$ correspond à l'erreur de prédiction rétrograde à l'ordre p définie par :

$$\epsilon_{t,p}^- = X_t - \text{proj}(X_t | \mathcal{H}_{t+p,p}) = X_t - \text{proj}(X_t | \text{Vect}(X_{t+p}, \dots, X_{t+1})).$$

Elle représente la différence entre la valeur à l'instant courant X_t et la projection orthogonale de X_t sur les p échantillons *qui suivent* l'instant courant $\{X_{t+1}, \dots, X_{t+p}\}$. Le qualificatif *rétrograde* est clair : il traduit le fait que l'on cherche à prédire la valeur courante en fonction des valeurs futures.

D'après la proposition 2.4.3,

$$\text{proj}(X_t | \mathcal{H}_{t-1,p+1}) = \text{proj}(X_t | \mathcal{H}_{t-1,p}) + \text{proj}(X_t | \text{Vect}(\epsilon_{t-p-1,p}^-)),$$

où d'après l'exemple 2.4.2 :

$$\text{proj}(X_t | \text{Vect}(\epsilon_{t-p-1,p}^-)) = \alpha \epsilon_{t-p-1,p}^- \quad \text{avec} \quad \alpha = \langle X_t, \epsilon_{t-p-1,p}^- \rangle / \|\epsilon_{t-p-1,p}^-\|^2.$$

On en déduit donc que :

$$\begin{aligned}\text{proj}(X_t | \mathcal{H}_{t-1,p+1}) &= \text{proj}(X_t | \mathcal{H}_{t-1,p}) \\ &\quad + k_{p+1} [X_{t-p-1} - \text{proj}(X_{t-p-1} | \mathcal{H}_{t-1,p})],\end{aligned}\quad (5.15)$$

où

$$k_{p+1} = \frac{\langle X_t, \epsilon_{t-p-1,p}^- \rangle}{\|\epsilon_{t-p-1,p}^-\|^2}. \quad (5.16)$$

Montrons à présent que les coefficients de prédiction rétrograde coïncident avec les coefficients de prédiction directe. Plus précisément, si

$$\text{proj}(X_t | \mathcal{H}_{t-1,p}) = \sum_{k=1}^p \phi_{k,p} X_{t-k}, \quad (5.17)$$

alors

$$\text{proj}(X_{t-p-1} | \mathcal{H}_{t-1,p}) = \sum_{k=1}^p \phi_{k,p} X_{t-p-1+k} = \sum_{k=1}^p \phi_{p+1-k,p} X_{t-k}. \quad (5.18)$$

En effet, les coefficients des deux développements (5.17) et (5.18) sont tous les deux donnés par (5.10). En utilisant (5.17) et (5.18) dans (5.15), on a :

$$\begin{aligned}
\text{proj}(X_t | \mathcal{H}_{t-1,p+1}) &= \sum_{k=1}^{p+1} \phi_{k,p+1} X_{t-k} \\
&= \sum_{k=1}^p (\phi_{k,p} - k_{p+1} \phi_{p+1-k,p}) X_{t-k} + k_{p+1} X_{t-p-1} .
\end{aligned}$$

On en déduit, par unicité, les formules de récurrence donnant les coefficients de prédiction à l'ordre $p+1$ à partir de ceux à l'ordre p :

$$\begin{cases} \phi_{k,p+1} = \phi_{k,p} - k_{p+1} \phi_{p+1-k,p} , & \text{pour } k \in \{1, \dots, p\} , \\ \phi_{p+1,p+1} = k_{p+1} . \end{cases} \quad (5.19)$$

Explicitons à présent la relation (5.16) définissant k_{p+1} . En utilisant (5.18), on a :

$$\begin{aligned}
\langle X_t, \epsilon_{t-p-1,p}^- \rangle &= \langle X_t, X_{t-p-1} - \text{proj}(X_{t-p-1} | \mathcal{H}_{t-1,p}) \rangle \\
&= \gamma(p+1) - \left\langle X_t, \sum_{k=1}^p \phi_{k,p} X_{t-p-1+k} \right\rangle = \gamma(p+1) - \sum_{k=1}^p \phi_{k,p} \gamma(p+1-k) .
\end{aligned}$$

D'autre part,

$$\begin{aligned}
\|\epsilon_{t-p-1,p}^-\|^2 &= \left\langle X_{t-p-1}, X_{t-p-1} - \sum_{k=1}^p \phi_{k,p} X_{t-p-1+k} \right\rangle \\
&= \gamma(0) - \sum_{k=1}^p \phi_{k,p} \gamma(k) = \sigma_p^2 = \|\epsilon_{t,p}^+\|^2 , \quad (5.20)
\end{aligned}$$

ce qui donne

$$k_{p+1} = \frac{\gamma(p+1) - \sum_{k=1}^p \phi_{k,p} \gamma(p+1-k)}{\sigma_p^2} . \quad (5.21)$$

Il nous reste maintenant à déterminer l'erreur de prédiction σ_{p+1}^2 à l'ordre $(p+1)$ en fonction de σ_p^2 . En utilisant l'équation (5.15), on a

$$\begin{aligned}
\epsilon_{t,p+1}^+ &= X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p+1}) \\
&= X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p}) - k_{p+1} [X_{t-p-1} - \text{proj}(X_{t-p-1} | \mathcal{H}_{t-1,p})] \\
&= X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p}) - k_{p+1} \epsilon_{t-p-1,p}^- ,
\end{aligned}$$

dont on déduit d'après (5.20) :

$$\sigma_{p+1}^2 = \|\epsilon_{t,p+1}^+\|^2 = \sigma_p^2 + k_{p+1}^2 \sigma_p^2 - 2k_{p+1} \langle X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p}), \epsilon_{t-p-1,p}^- \rangle .$$

En utilisant que $\text{proj}(X_t | \mathcal{H}_{t-1,p})$ et $\epsilon_{t-p-1,p}^-$ sont orthogonaux, (5.16) et (5.20), on obtient

$$\sigma_{p+1}^2 = \sigma_p^2(1 - k_{p+1}^2). \quad (5.22)$$

A partir de ces récursions, nous allons à présent décrire l'algorithme de Levinson-Durbin.

Pour initialiser l'algorithme, nous nous intéressons au cas $p = 0$. Dans ce cas, la meilleure prédiction de X_t est $\mathbb{E}[X_t] = 0$ et la variance de l'erreur de prédiction est donnée par $\sigma_0^2 = \mathbb{E}[(X_t - 0)^2] = \gamma(0)$. Au pas suivant on a $k_1 = \gamma(1)/\gamma(0)$, en posant $p = 0$ dans (5.21), $\phi_{1,1} = \gamma(1)/\gamma(0)$, en posant $p = 0$ dans (5.19) et $\sigma_1^2 = \gamma(0)(1 - k_1^2)$, en posant $p = 0$ dans (5.22).

L'algorithme de *Levinson-Durbin* qui permet de déterminer les coefficients de prédiction $\{\phi_{m,p}\}_{1 \leq m \leq p, 1 \leq p \leq K}$ à partir de $\gamma(0), \dots, \gamma(K)$ s'écrit alors de la façon suivante :

Algorithme 5.2.1 (Levinson-Durbin)

Initialisation $k_1 = \gamma(1)/\gamma(0)$, $\phi_{1,1} = \gamma(1)/\gamma(0)$ et $\sigma_1^2 = \gamma(0)(1 - k_1^2)$

Récursion Pour $p = \{2, \dots, K\}$ répéter :

a) Calculer

$$\begin{aligned} k_p &= \sigma_{p-1}^{-2} \left(\gamma(p) - \sum_{k=1}^{p-1} \phi_{k,p-1} \gamma(p-k) \right) \\ \phi_{p,p} &= k_p \\ \sigma_p^2 &= \sigma_{p-1}^2(1 - k_p^2) \end{aligned}$$

b) Pour $m \in \{1, \dots, p-1\}$ calculer :

$$\phi_{m,p} = \phi_{m,p-1} - k_p \phi_{p-m,p-1}$$

Proposition 5.2.2. Soit (X_t) un processus stationnaire au second ordre de fonction d'autocovariance $\gamma(h)$. Le coefficient k_{p+1} défini par (5.16) vérifie, pour tout $p \geq 0$:

$$k_{p+1} = \frac{\langle \epsilon_{t,p}^+, \epsilon_{t-p-1,p}^- \rangle}{\|\epsilon_{t,p}^+\| \|\epsilon_{t-p-1,p}^-\|}, \quad (5.23)$$

et

$$|k_{p+1}| \leq 1. \quad (5.24)$$

Preuve. En utilisant que $\text{proj}(X_t | \mathcal{H}_{t-1,p})$ est orthogonal à $\epsilon_{t-p-1,p}^-$, (5.16) et (5.7), on a

$$k_{p+1} = \frac{\langle \epsilon_{t,p}^+, \epsilon_{t-p-1,p}^- \rangle}{\|\epsilon_{t-p-1,p}^-\|^2}.$$

Or, d'après (5.20), $\|\epsilon_{t-p-1,p}^-\|^2 = \sigma_p^2 = \|\epsilon_{t,p}^+\|^2$, la dernière égalité venant de (5.8), d'où l'on déduit (5.23). L'inégalité (5.24) se déduit alors de (5.23) en utilisant l'inégalité de Cauchy-Schwarz. \square

Définition 5.2.3 (Fonction d'autocorrélation partielle). Soit (X_t) un processus stationnaire au second ordre de fonction d'autocovariance $\gamma(h)$. On appelle *fonction d'autocorrélation partielle* la suite des coefficients d'autocorrélation partielle $(k_p)_{p \geq 1}$ définie par :

$$k_p = \text{corr}(X_t, X_{t-1}) = \frac{\langle X_t, X_{t-1} \rangle}{\|X_t\| \|X_{t-1}\|}, \text{ si } p = 1, \quad (5.25)$$

et

$$\begin{aligned} k_p &= \text{corr}(\epsilon_{t,p-1}^+, \epsilon_{t-p,p-1}^-) \\ &= \frac{\langle X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p-1}), X_{t-p} - \text{proj}(X_{t-p} | \mathcal{H}_{t-1,p-1}) \rangle}{\|X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p-1})\| \|X_{t-p} - \text{proj}(X_{t-p} | \mathcal{H}_{t-1,p-1})\|}, \text{ si } p \geq 2. \end{aligned} \quad (5.26)$$

Remarque 5.2.4. Dans (5.25), l'expression pour $p = 1$ est en accord avec celle pour $p \geq 2$ dans la mesure où on peut noter que $\epsilon_{t,0}^+ = X_t$ et que $\epsilon_{t-1,0}^- = X_{t-1}$. Notons aussi que, dans l'expression de k_p , X_t et X_{t-p} sont projetés sur le même sous-espace $\text{Vect}(X_{t-1}, \dots, X_{t-p+1})$. Le résultat remarquable est que la suite des coefficients de corrélation partielle est donnée par :

$$k_p = \phi_{p,p} \quad (5.27)$$

où $\phi_{p,p}$ est défini au moyen des équations de Yule-Walker (5.6).

Dans le cas particulier d'un processus $\text{AR}(m)$ causal, on a alors :

$$k_p = \begin{cases} \phi_{p,p} & \text{pour } 1 \leq p < m, \\ \phi_m & \text{pour } p = m, \\ 0 & \text{pour } p > m. \end{cases}$$

5.3 Algorithme des innovations

L'algorithme des innovations est une application directe de la méthode de Gram-Schmidt et est, à cet égard, plus élémentaire que l'algorithme de Levinson-Durbin. De plus, il ne suppose pas que le processus $(X_t)_{t \in \mathbb{Z}}$ soit stationnaire. L'espérance de X_t étant supposée nulle dans ce chapitre, nous notons

$$\kappa(i, j) = \langle X_i, X_j \rangle = \mathbb{E}[X_i X_j],$$

la fonction d'autocovariance de ce processus. Notons, pour $n \geq 1$,

$$\mathcal{H}_n = \text{Vect}(X_1, \dots, X_n) \text{ et } \sigma_n^2 = \|X_{n+1} - \text{proj}(X_{n+1} | \mathcal{H}_n)\|^2.$$

La procédure d'orthogonalisation de Gram-Schmidt permet alors d'écrire pour tout $n \geq 1$:

$$\mathcal{H}_n = \text{Vect} (X_1, X_2 - \text{proj} (X_2 | X_1), \dots, X_n - \text{proj} (X_n | \mathcal{H}_{n-1})) ,$$

où on utilise la convention suivante : $\text{proj} (X_1 | \mathcal{H}_0) = 0$. On a alors :

$$\text{proj} (X_{n+1} | \mathcal{H}_n) = \sum_{j=1}^n \theta_{n,j} (X_{n+1-j} - \text{proj} (X_{n+1-j} | \mathcal{H}_{n-j})) . \quad (5.28)$$

L'algorithme des innovations décrit dans la proposition suivante fournit une méthode récursive permettant de calculer $(\theta_{n,j})_{1 \leq j \leq n}$ et σ_n^2 pour $n \geq 1$.

Proposition 5.3.1. *Soit (X_t) un processus à moyenne nulle tel que la matrice $[\kappa(i, j)]_{1 \leq i, j \leq n}$ soit inversible pour tout $n \geq 1$ alors*

$$\text{proj} (X_{n+1} | \mathcal{H}_n) = \begin{cases} 0 , & \text{si } n = 0 , \\ \sum_{j=1}^n \theta_{n,j} (X_{n+1-j} - \text{proj} (X_{n+1-j} | \mathcal{H}_{n-j})) , & \text{si } n \geq 1 , \end{cases}$$

où

$$\begin{cases} \sigma_0^2 = \kappa(1, 1) , \\ \theta_{n,n-k} = \sigma_k^{-2} \left[\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} \sigma_j^2 \right] , & 0 \leq k \leq n-1 , \\ \sigma_n^2 = \kappa(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 \sigma_j^2 , & n \geq 1 . \end{cases}$$

Preuve. Remarquons tout d'abord que les vecteurs $(X_i - \text{proj} (X_i | \mathcal{H}_{i-1}))_{i \geq 1}$ sont orthogonaux. En effet, pour $i < j$, $X_i - \text{proj} (X_i | \mathcal{H}_{i-1}) \in \mathcal{H}_{j-1}$ et $X_j - \text{proj} (X_j | \mathcal{H}_{j-1}) \perp \mathcal{H}_{j-1}$. On en déduit, en faisant le produit scalaire de (5.28) par $X_{k+1} - \text{proj} (X_{k+1} | \mathcal{H}_k)$ que, pour $0 \leq k < n$:

$$\langle \text{proj} (X_{n+1} | \mathcal{H}_n), X_{k+1} - \text{proj} (X_{k+1} | \mathcal{H}_k) \rangle = \theta_{n,n-k} \sigma_k^2 .$$

Puisque $\langle X_{n+1} - \text{proj} (X_{n+1} | \mathcal{H}_n), X_{k+1} - \text{proj} (X_{k+1} | \mathcal{H}_k) \rangle = 0$, les coefficients $\theta_{n,n-k}$, $k = 0, \dots, n-1$ sont donnés par

$$\theta_{n,n-k} = \sigma_k^{-2} \langle X_{n+1}, X_{k+1} - \text{proj} (X_{k+1} | \mathcal{H}_k) \rangle . \quad (5.29)$$

En utilisant la représentation (5.28),

$$\begin{aligned} \text{proj} (X_{k+1} | \mathcal{H}_k) &= \sum_{j=1}^k \theta_{k,j} (X_{k+1-j} - \text{proj} (X_{k+1-j} | \mathcal{H}_{k-j})) \\ &= \sum_{j=0}^{k-1} \theta_{k,k-j} (X_{j+1} - \text{proj} (X_{j+1} | \mathcal{H}_j)) , \end{aligned}$$

d'où l'on déduit que

$$\theta_{n,n-k} = \sigma_k^{-2} \left(\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \langle X_{n+1}, X_{j+1} - \text{proj}(X_{j+1} | \mathcal{H}_j) \rangle \right).$$

D'après (5.29), $\langle X_{n+1}, X_{j+1} - \text{proj}(X_{j+1} | \mathcal{H}_j) \rangle = \sigma_j^2 \theta_{n,n-j}$ pour $0 \leq j < n$, nous avons donc pour $k \in \{1, \dots, n-1\}$,

$$\theta_{n,n-k} = \sigma_k^{-2} \left(\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} \sigma_j^2 \right). \quad (5.30)$$

L'équation (5.30) est encore valable lorsque $k = 0$ en utilisant la convention que la somme sur j dans le membre de droite est nulle dans ce cas. Par ailleurs, la proposition 2.4.3 (Pythagore) implique que

$$\begin{aligned} \sigma_n^2 &= \|X_{n+1} - \text{proj}(X_{n+1} | \mathcal{H}_n)\|^2 = \|X_{n+1}\|^2 - \|\text{proj}(X_{n+1} | \mathcal{H}_n)\|^2 \\ &= \kappa(n+1, n+1) - \sum_{k=0}^{n-1} \theta_{n,n-k}^2 \sigma_k^2. \end{aligned} \quad (5.31)$$

□

Alors que l'algorithme de Levinson-Durbin permet de déterminer les coefficients du développement de $\text{proj}(X_{n+1} | \mathcal{H}_n)$ sur X_1, \dots, X_n donnés par $\text{proj}(X_{n+1} | \mathcal{H}_n) = \sum_{j=1}^n \phi_{n,j} X_{n+1-j}$, l'algorithme des innovations calcule les coefficients du développement de $\text{proj}(X_{n+1} | \mathcal{H}_n)$ sur $X_1, X_2 - \text{proj}(X_2 | X_1), \dots, X_n - \text{proj}(X_n | \mathcal{H}_{n-1})$.

Exemple 5.3.2 (Prédiction d'un processus MA(1)). Considérons le processus $X_t = Z_t + \theta Z_{t-1}$ où $(Z_t) \sim \text{BB}(0, \sigma^2)$. Nous avons donc $\kappa(i, j) = 0$ pour $|i - j| > 1$, $\kappa(i, i) = \sigma^2(1 + \theta^2)$ et $\kappa(i, i+1) = \theta\sigma^2$. Dans ce cas, nous avons

$$\begin{cases} \theta_{n,j} = 0, & 2 \leq j \leq n, \\ \theta_{n,1} = \sigma_{n-1}^{-2} \theta \sigma^2, \end{cases}$$

et les variances des innovations qui sont données par

$$\begin{cases} \sigma_0^2 = (1 + \theta^2) \sigma^2, \\ \sigma_n^2 = [1 + \theta^2 - \sigma_{n-1}^{-2} \theta^2 \sigma^2] \sigma^2. \end{cases}$$

Si nous posons $r_n = \sigma_n^2 / \sigma^2$, nous avons

$$\text{proj}(X_{n+1} | \mathcal{H}_n) = \theta (X_n - \text{proj}(X_n | \mathcal{H}_{n-1})) / r_{n-1},$$

avec $r_0 = 1 + \theta^2$, et pour $n \geq 1$, $r_{n+1} = 1 + \theta^2 - \theta^2 / r_n$.

Problèmes

5.3.3 (Inversibilité de la matrice d'autocovariance). On propose dans ce problème une autre démonstration de la proposition 3.3.9.

Soit (X_t) un processus centré stationnaire au second ordre de fonction d'autocovariance γ . Pour tout $t \in \mathbb{N}$, on note Γ_t la matrice $t \times t$ de covariance

$$\Gamma_t = [\gamma(i-j)]_{1 \leq i, j \leq t} .$$

On suppose qu'il existe un indice $r \geq 1$ tel que la matrice Γ_r soit inversible et Γ_{r+1} soit singulière.

1. Montrer que, pour tout $n \geq r + 1$, il existe une suite de coefficients $b^{(n)} \in \mathbb{R}^r$ tel que

$$X_n = \sum_{j=1}^r b_j^{(n)} X_j .$$

2. Montrer que les $b_j^{(n)}$ sont uniformément bornés.
3. Conclure lorsque $\gamma(0) > 0$ et $\lim_{t \rightarrow \infty} \gamma(t) = 0$?

5.3.4 (Prédiction de processus ARMA). Soient $\mathcal{M}_n = \overline{\text{sp}}(X_j, -\infty < j \leq n)$ et (X_t) un processus ARMA(p, q) causal et inversible satisfaisant : $\Phi(B)X_t = \Theta(B)Z_t$, où (Z_t) est un bruit blanc de variance σ^2 . On pose : $\tilde{X}_t = P_{\mathcal{M}_n} X_t$.

1. Montrer que

$$\tilde{X}_{n+h} = - \sum_{j \geq 1} \pi_j \tilde{X}_{n+h-j} \text{ et } \tilde{X}_{n+h} = \sum_{j \geq h} \psi_j Z_{n+h-j} ,$$

où $\sum_{j \geq 0} \pi_j z^j = \Phi(z)/\Theta(z)$ et $\sum_{j \geq 0} \psi_j z^j = \Theta(z)/\Phi(z)$, $|z| \leq 1$. De plus, montrer que

$$\mathbb{E} \left[(X_{n+h} - \tilde{X}_{n+h})^2 \right] = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 .$$

2. Calculer \tilde{X}_{n+1} pour un AR(p) causal et un MA(1) inversible.

Chapitre 6

Décomposition de Wold

6.1 Décomposition de Wold

Un des résultats fondamentaux de la théorie des processus stationnaires au second-ordre est la décomposition de Wold. Cette décomposition permet de décomposer n'importe quel processus stationnaire au second-ordre comme la somme d'un processus résultant du filtrage linéaire d'un bruit blanc et d'un processus déterministe (définition 5.1.4). La preuve de ce résultat est de nature géométrique. L'idée de base est la suivante. Soit $\mathcal{H}_t^X = \overline{\text{Vect}}(X_s, s \leq t)$. \mathcal{H}_t^X est appelé le *passé linéaire* du processus à la date t . Par construction, $\mathcal{H}_t^X \subset \mathcal{H}_{t+1}^X$, et nous disposons ainsi d'une famille de sous-espace emboîtés de $\mathcal{H}_\infty^X = \bigcup_{t \in \mathbb{Z}} \mathcal{H}_t^X$. \mathcal{H}_∞^X est l'enveloppe linéaire du processus. L'espace $\bigcap_{t \in \mathbb{Z}} \mathcal{H}_t^X$, appelé le *passé infini du processus* $\{X_t, t \in \mathbb{Z}\}$ jouera aussi un rôle particulier. Par définition X_t appartient à \mathcal{H}_t^X , mais il n'appartient généralement pas à \mathcal{H}_{t-1}^X . Le théorème de projection dit qu'il existe un unique élément noté $\text{proj}(X_t | \mathcal{H}_{t-1}^X)$ et appartenant à \mathcal{H}_{t-1}^X tel que :

$$\epsilon_t = X_t - \text{proj}(X_t | \mathcal{H}_{t-1}^X) \perp \mathcal{H}_{t-1}^X$$

Dans ce contexte ϵ_t s'appelle l'*innovation* (linéaire) du processus. Il découle de cette construction géométrique que le processus d'innovation est un *processus orthogonal* dans le sens où :

$$\forall s \neq t, \quad \epsilon_s \perp \epsilon_t \quad (6.1)$$

En effet, pour $s < t$, nous pouvons écrire $\epsilon_s \in \mathcal{H}_s^X \subset \mathcal{H}_{t-1}^X$ et $\epsilon_t \perp \mathcal{H}_{t-1}^X$. Et donc $\epsilon_s \perp \epsilon_t$.

La proposition qui suit montre que le processus d'innovation est la limite des processus d'innovations partielles à l'ordre p .

Proposition 6.1.1. *Pour tout $Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ et tout $t \in \mathbb{Z}$ nous avons :*

$$\lim_{p \rightarrow \infty} \text{proj}(Y | \mathcal{H}_{t,p}^X) = \text{proj}(Y | \mathcal{H}_t^X)$$

où $\mathcal{H}_{t,p}^X = \text{Vect}(X_t, X_{t-1}, \dots, X_{t-p+1})$.

Exemple 6.1.2 (Bruit blanc). Supposons que $\{X_t\}$ soit un bruit blanc. Nous avons $\text{proj}(X_t | \mathcal{H}_{t-1,p}^X) = 0$ pour tout p et donc $\text{proj}(X_t | \mathcal{H}_{t-1}^X) = 0$. Nous avons donc $\epsilon_t = X_t - \text{proj}(X_t | \mathcal{H}_{t-1}^X) = X_t$: le processus X_t coïncide avec son innovation. Ceci signifie qu'un bruit blanc ne peut être prédit de façon linéaire à partir de son passé.

Exemple 6.1.3 (Prédiction d'un processus AR(p) causal). On considère le processus AR(p) causal défini par l'équation récurrente $X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t$ où $Z_t \sim \text{BB}(0, \sigma^2)$. Dans le cas causal, on a $\mathcal{H}_t^X = \mathcal{H}_t^Z$ et, pour tout $k \geq 1$, on avait $\mathbb{E}[X_{t-k} Z_t] = 0$. Par conséquent $Z_t \perp \mathcal{H}_{t-1}^X$ et $\mathcal{H}_t^X = \mathcal{H}_{t-1}^X \oplus \text{Vect}(Z_t)$. On en déduit que :

$$\text{proj} (X_t | \mathcal{H}_{t-1}^X) = \sum_{k=1}^p \phi_k \text{proj} (X_{t-k} | \mathcal{H}_{t-1}^X) + \text{proj} (Z_t | \mathcal{H}_{t-1}^X) = \sum_{k=1}^p \phi_k X_{t-k}$$

et donc $X_t - \text{proj} (X_t | \mathcal{H}_{t-1}^X) = X_t - \sum_{k=1}^p \phi_k X_{t-k} = Z_t$. Par conséquent le bruit blanc Z_t , qui intervient dans l'équation récurrente d'un AR causal, est précisément l'innovation du processus AR. Ce résultat montre que $\sum_{k=1}^p \phi_k X_{t-k}$ est la projection de $X(t)$ sur tout le passé \mathcal{H}_{t-1} et qu'elle coïncide avec la projection orthogonale sur le passé $\mathcal{H}_{t-1,p}$ de durée p . Par conséquent, pour tout $m \geq p$, la suite des coefficients de prédiction est $\{\phi_1, \dots, \phi_p, \underbrace{0, \dots, 0}_{m-p}\}$.

Ce résultat est faux pour un AR non causal.

Exemple 6.1.4 (Processus harmonique). Soit le processus harmonique $X_t = A \cos(\lambda_0 t + \Phi)$ où A est une variable aléatoire, centrée, de variance σ_A^2 et Φ une variable aléatoire, indépendante de A et distribuée suivant une loi uniforme sur $[-\pi, \pi]$. Le processus X_t est stationnaire au second-ordre, centré, de fonction d'autocovariance $\gamma(\tau) = (\sigma_A^2/2) \cos(\lambda_0 \tau)$. Les coefficients du prédicteur linéaire optimal à l'ordre 2 sont donnés par :

$$\begin{bmatrix} \phi_{1,2} \\ \phi_{2,2} \end{bmatrix} = \begin{bmatrix} 1 & \cos(\lambda_0) \\ \cos(\lambda_0) & 1 \end{bmatrix}^{-1} \begin{bmatrix} \cos(\lambda_0) \\ \cos(2\lambda_0) \end{bmatrix} = \begin{bmatrix} 2 \cos(\lambda_0) \\ -1 \end{bmatrix}$$

On vérifie facilement que $\sigma_2^2 = \|X_t - \text{proj} (X_t | \mathcal{H}_{t-1,2}^X)\|^2 = 0$. Par conséquent, on a :

$$X_t = \text{proj} (X_t | \mathcal{H}_{t-1,2}^X) = 2 \cos(\lambda_0) X_{t-1} - X_{t-2} \in \mathcal{H}_{t-1}^X$$

et donc la projection $\text{proj} (X_t | \mathcal{H}_{t-1}^X) = X_t$, ce qui implique que $\epsilon_t = 0$. À l'inverse du bruit blanc, le processus est entièrement prédictible à partir de son passé.

En appliquant la proposition 6.1.1 à X_t , nous pouvons écrire :

$$\lim_{p \rightarrow \infty} \text{proj} (X_t | \mathcal{H}_{t-1,p}^X) = \text{proj} (X_t | \mathcal{H}_{t-1}^X) \quad \text{et} \quad \lim_{p \rightarrow \infty} \epsilon_{t,p}^+ = \epsilon_t \quad (6.2)$$

Le processus d'innovation ϵ_t est donc la limite en moyenne quadratique de la suite des innovations partielles $\epsilon_{t,p}^+ = X_t - \text{proj} (X_t | \mathcal{H}_{t-1,p}^X)$. Une conséquence immédiate est que le processus d'innovation est un processus stationnaire au second ordre. En utilisant, en effet, la continuité du produit scalaire et la stationnarité au second ordre de l'innovation partielle d'ordre p , on peut écrire :

$$\langle \epsilon_{t+\tau}, \epsilon_t \rangle = \lim_{p \rightarrow \infty} \langle \epsilon_{t+\tau,p}^+, \epsilon_{t,p}^+ \rangle = \lim_{p \rightarrow \infty} \langle \epsilon_{\tau,p}^+, \epsilon_{0,p}^+ \rangle \quad (6.3)$$

qui ne dépend que de τ . En particulier nous avons :

$$\sigma^2 = \|\epsilon_t\|^2 = \lim_{p \rightarrow \infty} \|X_t - \text{proj} (X_t | \mathcal{H}_{t,p}^X)\|^2 = \lim_{p \rightarrow \infty} \sigma_p^2$$

Dans le cas du bruit blanc on obtient $\sigma^2 = \mathbb{E}[X_t^2] \neq 0$ et donc, d'après la définition 5.1.4, le bruit blanc est un processus régulier. D'un autre côté, le processus harmonique, pour lequel $\sigma^2 = 0$, est déterministe. Nous remarquons aussi que la somme d'un bruit blanc et d'un processus harmonique est un processus régulier.

La structure géométrique emboîtée des espaces $\{\mathcal{H}_t^X\}$ et l'orthogonalité des innovations fournissent, pour tout $s < t$, la formule suivante de décomposition en somme orthogonale :

$$\mathcal{H}_t^X = \mathcal{H}_s^X \oplus^\perp \text{Vect}(\epsilon_{s+1}, \dots, \epsilon_t) \quad (6.4)$$

Notons, tout d'abord, que $\epsilon_t = X_t - \text{proj}(X_t | \mathcal{H}_{t-1}^X) \in \mathcal{H}_t^X$ et que $\epsilon_t \perp \mathcal{H}_{t-1}^X$, ce qui implique que $\mathcal{H}_{t-1}^X \oplus^\perp \text{Vect}(\epsilon_t) \subseteq \mathcal{H}_t^X$. D'un autre côté, puisque $X_t = \epsilon_t + \text{proj}(X_t | \mathcal{H}_{t-1}^X)$,

$$\mathcal{H}_t^X = \overline{\text{Vect}(\epsilon_t + \text{proj}(X_t | \mathcal{H}_{t-1}^X), \{X_s, s \leq t-1\})} = \overline{\text{Vect}(\epsilon_t, \{X_s, s \leq t-1\})},$$

ce qui entraîne que $\mathcal{H}_t^X \subseteq \mathcal{H}_{t-1}^X \oplus^\perp \text{Vect}(\epsilon_t)$. En conclusion $\mathcal{H}_t^X = \mathcal{H}_{t-1}^X \oplus^\perp \text{Vect}(\epsilon_t)$. En réitérant ce raisonnement, on en déduit la décomposition (6.4). Cette décomposition orthogonale de l'espace \mathcal{H}_t^X n'est pas sans rappeler la décomposition de Gram-Schmidt. Notons qu'à l'inverse de la décomposition de Gram-Schmidt classique, nous procédons ici dans le sens rétrograde. Définissons pour tout $s \geq 0$:

$$\psi_s = \frac{\langle X_t, \epsilon_{t-s} \rangle}{\sigma^2} \quad (6.5)$$

Remarquons que ψ_s ne dépend pas de t . En effet, la continuité du produit scalaire et la stationnarité conjointe du processus X_t et de l'innovation partielle impliquent que :

$$\langle X_t, \epsilon_{t-s} \rangle = \lim_{p \rightarrow \infty} \langle X_t, \epsilon_{t-s,p}^+ \rangle = \lim_{p \rightarrow \infty} \langle X_0, \epsilon_{-s,p}^+ \rangle$$

Lemme 6.1.5. *La suite $\{\psi_s, s \in \mathbb{Z}\}$ est de carré sommable et $\psi_0 = 1$.*

Preuve. Remarquons, tout d'abord, que la relation $\langle \text{proj}(X_t | \mathcal{H}_{t-1}^X), \epsilon_t \rangle = 0$ entraîne que :

$$\begin{aligned} \psi_0 &= \frac{\langle X_t, \epsilon_t \rangle}{\sigma^2} \\ &= \frac{\langle X_t - \text{proj}(X_t | \mathcal{H}_{t-1}^X), \epsilon_t \rangle}{\sigma^2} \\ &= 1. \end{aligned}$$

D'autre part, pour tout $s \geq 0$, la projection orthogonale de X_t sur $\mathcal{H}_{t,s}^\epsilon = \text{Vect}(\epsilon_t, \epsilon_{t-1}, \dots, \epsilon_{t-s+1})$ s'écrit, du fait de l'orthogonalité du processus d'innovation, $\text{proj}(X_t | \mathcal{H}_{t,s}^\epsilon) = \sum_{k=0}^{s-1} \psi_k \epsilon_{t-k}$. On en déduit que $\|\text{proj}(X_t | \mathcal{H}_{t,s}^\epsilon)\|^2 = \sigma^2 \sum_{k=0}^{s-1} \psi_k^2$. On a alors d'après l'égalité de Pythagore (proposition 2.4.3) :

$$\|\text{proj}(X_t | \mathcal{H}_{t,s}^\epsilon)\|^2 = \sigma^2 \sum_{k=0}^{s-1} \psi_k^2 = \|X_t\|^2 - \|X_t - \text{proj}(X_t | \mathcal{H}_{t,s}^\epsilon)\|^2 \leq \|X_t\|^2$$

ce qui conclut la preuve. \square

La suite $(\psi_s)_{s \geq 0}$ étant de carré sommable, la suite $s \rightarrow X_{t,s} = \sum_{k=0}^s \psi_k \epsilon_{t-k}$ est, pour t fixé, une suite de Cauchy dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Elle admet donc, quand $s \rightarrow \infty$, une limite que nous notons :

$$U_t = \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}$$

et qui est un processus stationnaire au second-ordre. On a, en effet :

$$\mathbb{E}[U_t] = \langle U_t, 1 \rangle = \lim_{s \rightarrow \infty} \sum_{k=0}^s \psi_k \langle \epsilon_{t-k}, 1 \rangle = 0$$

et

$$\mathbb{E}[U_{t+\tau} U_t] = \langle U_{t+\tau}, U_t \rangle = \lim_{s \rightarrow \infty} \left(\sum_{k=0}^s \psi_k \epsilon_{t+\tau-k}, \sum_{k=0}^s \psi_k \epsilon_{t-k} \right) = \lim_{s \rightarrow \infty} \left(\sum_{k=0}^s \psi_k \epsilon_{\tau-k}, \sum_{k=0}^s \psi_k \epsilon_{-k} \right)$$

qui est indépendant de t .

Le théorème suivant, connu sous le nom de *décomposition de Wold*, est vraisemblablement le résultat le plus important de la théorie des processus stationnaires au second-ordre.

Théorème 6.1.6 (Décomposition de Wold). *Soit X_t un processus stationnaire au second ordre et ϵ_t son processus d'innovation. On suppose que X_t est un processus régulier ($\sigma^2 = \|\epsilon_t\|^2 \neq 0$). On note $U_t = \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}$ où $\psi_k = \langle X_t, \epsilon_{t-k} \rangle / \sigma^2$. Alors il existe un processus V_t tel que :*

$$X_t = U_t + V_t, \quad (6.6)$$

et tel que :

- a) pour tout (t, s) , $\langle V_t, \epsilon_s \rangle = 0$, qui implique que $\langle V_t, U_s \rangle = 0$,
- b) $V_t = \text{proj}(X_t | \mathcal{H}_{-\infty}^X)$ est la projection orthogonale de X_t sur $\mathcal{H}_{-\infty}^X = \bigcap_{t=-\infty}^{\infty} \mathcal{H}_t^X$,
- c) U_t est un processus régulier et $\epsilon_t = U_t - \text{proj}(U_t | \mathcal{H}_{t-1}^U)$ est l'innovation du processus $\{U_t, t \in \mathbb{Z}\}$. De plus, $\mathcal{H}_t^\epsilon = \mathcal{H}_t^U$.

d) V_t est un processus déterministe et $\mathcal{H}_t^V = \mathcal{H}_{-\infty}^X$.

Preuve. Elle est donnée en fin de chapitre. \square

Un processus $\{X_t\}$ tel que $\mathcal{H}_{-\infty}^X = \{0\}$ est dit *purement non déterministe*. Pour un tel processus la partie déterministe de la décomposition de Wold est identiquement nulle. Par exemple, le processus régulier U_t de la décomposition de Wold est purement non déterministe. En effet, en appliquant la décomposition de Wold au processus U_t on a, pour tout t , $U_t = U_t + V_t$ avec $V_t = 0$ et donc, d'après le point d), $\mathcal{H}_{-\infty}^U = \{0\}$. Le théorème de Wold permet donc de décomposer tout processus stationnaire au second-ordre sous la forme d'une somme de deux processus orthogonaux, le premier étant *purement non déterministe* et le second étant *déterministe*. La partie purement non-déterministe s'exprime comme le filtrage d'un bruit blanc par un filtre linéaire invariant dans le temps de réponse impulsionnelle $\{\psi_k\}$ *causale* ($\psi_k = 0$ pour $k < 0$) et de carré sommable (pas nécessairement de module sommable).

Exemple 6.1.7 (Processus MA(1)). Soit $\{Z_t\}$ un bruit blanc et soit le processus $X_t = Z_t + \theta_1 Z_{t-1}$. Remarquons que, par construction, $\mathcal{H}_t^X \subseteq \mathcal{H}_t^Z$ mais que l'inclusion réciproque n'est pas nécessairement vérifiée. Montrons par contre que, pour $|\theta_1| < 1$, nous avons effectivement $\mathcal{H}_t^X = \mathcal{H}_t^Z$. En effet, en réitérant p fois l'équation $X_t = Z_t + \theta_1 Z_{t-1}$ et en résolvant par rapport à Z_t , nous obtenons :

$$Z_t = X_t - \theta_1 X_{t-1} + \theta_1^2 X_{t-2} - \cdots + (-1)^p \theta_1^p X_{t-p} - (-1)^p \theta_1^{p+1} Z_{t-p}$$

En prenant la limite en p , nous en déduisons que, si $|\theta_1| < 1$, alors :

$$Z_t = \sum_{k=0}^{\infty} (-\theta_1)^k X_{t-k}$$

ce qui montre que $\mathcal{H}_t^Z \subset \mathcal{H}_t^X$ et donc que $\mathcal{H}_t^X = \mathcal{H}_t^Z$. Dans ce cas, nous pouvons écrire :

$$\begin{aligned} \text{proj}(X_t | \mathcal{H}_{t-1}^X) &= \text{proj}(Z_t | \mathcal{H}_{t-1}^X) + \theta_1 \text{proj}(Z_{t-1} | \mathcal{H}_{t-1}^X) = \text{proj}(Z_t | \mathcal{H}_{t-1}^Z) + \theta_1 \text{proj}(Z_{t-1} | \mathcal{H}_{t-1}^Z) \\ &= 0 + \theta_1 Z_{t-1}, \end{aligned}$$

en remarquant que $\text{proj}(Z_t | \mathcal{H}_{t-1}^Z) = 0$ car $\{Z_t, t \in \mathbb{Z}\}$ est un bruit blanc. On en déduit que $X_t - \text{proj}(X_t | \mathcal{H}_{t-1}^X) = X_t - \theta_1 Z_{t-1} = Z_t$. Par conséquent, lorsque $|\theta_1| < 1$, le processus $\{Z_t, t \in \mathbb{Z}\}$ est l'innovation du processus $\{X_t, t \in \mathbb{Z}\}$. Notons que le processus $\{X_t, t \in \mathbb{Z}\}$ est *purement non déterministe* et que les coefficients de la décomposition de Wold sont simplement donnés par $\psi_0 = 1$, $\psi_1 = \theta$, et $\psi_k = 0$ pour $k > 1$.

6.2 Preuve du théorème 6.1.6

Preuve (Preuve du théorème 6.1.6).

1. Par définition, $V_t = X_t - \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k} \in \mathcal{H}_t^X$. Pour $s > t$, $\epsilon_s \perp \mathcal{H}_t^X$, et donc $\langle V_t, \epsilon_s \rangle = 0$. Pour $s \leq t$, $\langle V_t, \epsilon_s \rangle = \langle X_t, \epsilon_s \rangle - \psi_{t-s} \sigma^2$ qui est égal à 0 par définition de ψ_k .
2. Montrons tout d'abord que $V_t \in \mathcal{H}_{-\infty}^X$. La preuve se fait par récurrence. Nous avons $V_t \in \mathcal{H}_t^X$ et $V_t \perp \epsilon_t$ (d'après la propriété précédente). Comme $\mathcal{H}_t^X = \mathcal{H}_{t-1}^X \oplus^{\perp} \text{Vect}(\epsilon_t)$, on en déduit que $V_t \in \mathcal{H}_{t-1}^X$. Supposons à présent que $V_t \in \mathcal{H}_{t-s}^X$, pour $s \geq 0$. Comme $V_t \perp \epsilon_{t-s}$ et que $\mathcal{H}_{t-s}^X = \mathcal{H}_{t-s-1}^X \oplus^{\perp} \text{Vect}(\epsilon_{t-s})$, nous avons $V_t \in \mathcal{H}_{t-s-1}^X$. On a donc $V_t \in \mathcal{H}_{-\infty}^X = \bigcap_{s=-\infty}^{\infty} \mathcal{H}_s^X$. Il reste à montrer que $X_t - V_t = \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}$ est orthogonal à $\mathcal{H}_{-\infty}^X$. Pour cela considérons $Y \in \mathcal{H}_{-\infty}^X$. Nous avons :

$$\langle X_t - V_t, Y \rangle = \left\langle \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}, Y \right\rangle = \lim_{s \rightarrow +\infty} \sum_{k=0}^s \psi_k \langle \epsilon_{t-k}, Y \rangle$$

Mais, par définition, $Y \in \mathcal{H}_{-\infty}^X$ implique que, pour tout t , $Y \in \mathcal{H}_t^X$. Comme $\epsilon_{t-k} \perp \mathcal{H}_{t-s-1}^X$ pour $0 \leq k \leq s$, nous avons $\sum_{k=0}^s \psi_k \langle \epsilon_{t-k}, Y \rangle = 0$. Et donc, pour tout $Y \in \mathcal{H}_{-\infty}^X$, on a :

$$\langle X_t - V_t, Y \rangle = \langle U_t, Y \rangle = 0. \quad (6.7)$$

3. Notons que (6.7) implique que, pour tout t , $U_t \perp \mathcal{H}_{-\infty}^X$ et donc $\mathcal{H}_t^U = \overline{\text{Vect}(U_s, s \leq t)} \perp \mathcal{H}_{-\infty}^X$. On peut alors poser $\mathcal{L}_t = \mathcal{H}_t^U \oplus^{\perp} \mathcal{H}_{-\infty}^X$. La décomposition $X_t = U_t + V_t$ et la propriété précédente ($V_t = \text{proj}(X_t | \mathcal{H}_{-\infty}^X)$) impliquent que, pour tout t , $\mathcal{H}_t^X \subset \mathcal{L}_t$, et donc $\epsilon_t \in \mathcal{L}_t$. Comme, pour tout t , $\epsilon_t \perp \mathcal{H}_{t-u}^X$ pour tout $u \geq 0$, $\epsilon_t \perp Y$ pour tout $Y \in \mathcal{H}_{-\infty}^X$, puisque, en particulier, $Y \in \mathcal{H}_{t-u}^X$. Nous avons $\epsilon_t \perp \mathcal{H}_{-\infty}^X$. Et donc $\epsilon_t \in \mathcal{H}_t^U$. Cela entraîne que $\sum_{k=1}^{\infty} \psi_k \epsilon_{t-k} \in \mathcal{H}_{t-1}^U$. Notons que $\sum_{k=1}^{\infty} \psi_k \epsilon_{t-k} = U_t - \epsilon_t$ ($\psi_0 = 1$). Par conséquent, pour tout $Y \in \mathcal{H}_{t-1}^U$ on a :

$$\left\langle U_t - \sum_{k=1}^{\infty} \psi_k \epsilon_{t-k}, Y \right\rangle = \langle \epsilon_t, Y \rangle = 0.$$

Cela implique que $\sum_{k=1}^{\infty} \psi_k \epsilon_{t-k}$ est la projection orthogonale de U_t sur \mathcal{H}_{t-1}^U et donc que :

$$\epsilon_t = U_t - \text{proj}(U_t | \mathcal{H}_{t-1}^U).$$

Cela signifie que $\{\epsilon_t, t \in \mathbb{Z}\}$ est le processus d'innovation du processus $\{U_t, t \in \mathbb{Z}\}$. Comme, par hypothèse, $\sigma^2 = \|\epsilon_t\|^2 \neq 0$, le processus $\{U_t, t \in \mathbb{Z}\}$

$\mathbb{Z}\}$ est donc régulier. Remarquons que, comme $\epsilon_t \in \mathcal{H}_t^U$, nous avons $\mathcal{H}_t^\epsilon \subset \mathcal{H}_t^U$. Comme, par construction, $\mathcal{H}_t^U \subset \mathcal{H}_t^\epsilon$, nous avons $\mathcal{H}_t^U = \mathcal{H}_t^\epsilon$.

4. Montrons tout d'abord que, pour tout t , on a :

$$\mathcal{H}_t^V = \overline{\text{Vect}(V_s, s \leq t)} = \mathcal{H}_{-\infty}^X \quad (6.8)$$

Pour tout t , $V_t \in \mathcal{H}_{-\infty}^X$ et donc $\mathcal{H}_t^V \subseteq \mathcal{H}_{-\infty}^X$. D'un autre côté, d'après les points précédents, $\mathcal{H}_t^X = \mathcal{H}_t^\epsilon \stackrel{\perp}{\oplus} \mathcal{H}_t^V$. Et donc, quel que soit $Y \in \mathcal{H}_{-\infty}^X$, alors $Y \in \mathcal{H}_{s-1}^X$ pour tout s , de telle sorte que $(Y, \epsilon_s) = 0$ et donc $Y \in \mathcal{H}_t^V$, ce qui implique que $\mathcal{H}_{-\infty}^X \subseteq \mathcal{H}_t^V$. Ce qui démontre (6.8). Partant de (6.8), on déduit que $\text{proj}(V_t | \mathcal{H}_{t-1}^V) = \text{proj}(V_t | \mathcal{H}_{-\infty}^X) = \text{proj}(V_t | \mathcal{H}_t^V) = V_t$ et que $\|V_t - \text{proj}(V_t | \mathcal{H}_{t-1}^V)\|^2 = 0$: le processus $\{V_t, t \in \mathbb{Z}\}$ est donc déterministe. □

Chapitre 7

Modèles d'états

7.1 Définition et exemples

La forme générique d'un modèle d'état est donnée par les équations

$$X_{t+1} = A_t X_t + R_t U_t, \quad (7.1)$$

$$Y_t = B_t X_t + S_t V_t, \quad (7.2)$$

où

- a) $\{U_k\}_{k \geq 0}$ et $\{V_k\}_{k \geq 0}$ sont les bruits d'état et de mesure ; nous supposons que les vecteurs $\{(U_t, V_t)\}_{t \in \mathbb{Z}}$ sont des bruits blancs vectoriels décorrélés et de covariance identité ;
- b) La condition initiale X_0 est un vecteur aléatoire de moyenne μ_ν et de covariance Γ_ν et est décorrélée des bruits de d'état et de mesures $\{U_k\}$ and $\{V_k\}$;
- c) La *matrice de transition d'états* A_t , la *matrice de mesures* B , et les matrices R_t et S_t sont des matrices connues.

Ces modèles ont été introduits dans la littérature à la fin des années 1950 [2] ; ces modèles sont aujourd'hui couramment utilisés en traitement des séries temporelles scalaires ou multivariées, avec des applications en économétrie, finance, mais aussi dans différents domaines du traitement du signal et du contrôle.

Exemple 7.1.1 (tendance aléatoire en présence de bruit). Soit β une constante et (U_t) un bruit blanc vectoriel. Soit Z_1 une variable aléatoire décorrélée de (U_t) et considérons le processus

$$Z_{t+1} = Z_t + \beta + R U_t = Z_1 + \beta t + R(U_1 + \dots + U_t), t \geq 0.$$

Lorsque l'écart type R est faible Z_t est approximativement une fonction affine du temps ; le bruit d'état (U_t) introduit une "fluctuation" sur cette fonction. Pour construire une représentation d'état de (Z_t) , posons $X_t = [Z_t, \beta]^T$, et considérons l'équation d'état

$$X_{t+1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} X_t + U_t \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Le processus Z_t est relié à l'état par l'équation $Z_t = [10]X_t$. L'observation donnée par $Y_t = [10]X_t + S V_t$ est donc une fonction affine une tendance linéaire perturbée, observée dans un bruit additif.

Exemple 7.1.2 (tendance aléatoire et composante saisonnière).

7.2 Modèles d'états stationnaires

Dans cette partie, nous supposons que les matrices A_t , R_t , B_t et S_t ne dépendent pas du temps et nous considérons le modèle d'état :

$$X_{t+1} = AX_t + RU_t, \quad (7.3)$$

$$Y_t = BX_t + SV_t. \quad (7.4)$$

L'équation d'état est dite *stable* si toutes les valeurs propres de la matrice A sont de modules strictement inférieurs à 1, où de façon équivalente si $\det(I - Az) \neq 0$ pour $|z| \leq 1$. Dans ce cas, l'équation (7.3) admet une unique solution stationnaire donnée par

$$X_t = \sum_{j=0}^{\infty} A^j U_{t-j-1},$$

le processus d'observations étant alors donné par

$$Y_t = V_t + \sum_{j=0}^{\infty} BA^j U_{t-j-1}.$$

Exemple 7.2.1 (Processus autorégressif bruité). Un processus autorégressif d'ordre p , $\{Z_t\}_{t \geq 0}$, est un processus stationnaire au second-ordre qui satisfait l'équation aux différences

$$Z_{t+1} = \phi_1 Z_t + \dots + \phi_p Z_{t-p+1} + U_t, \quad (7.5)$$

où $\{U_k\}_{k \geq 0}$ est un bruit blanc. Définissons le vecteur

$$X_t = (Z_t, \dots, Z_{t-p+1})^T, \quad (7.6)$$

en "empilant" les valeurs retardées. Appelons A la *matrice compagnon* associée au polynôme de prédiction $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$,

$$A = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_p \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}. \quad (7.7)$$

En utilisant les notations précédentes, l'équation aux différences (7.5) peut être réécrite sous forme d'états :

$$X_t = AX_{t-1} + (1 \ 0 \dots 0)^t U_{t-1}, \quad (7.8)$$

$$Y_t = (1 \ 0 \dots 0) X_t. \quad (7.9)$$

Si le processus autorégressif n'est pas directement observable mais que nous disposons de mesures bruitées de celui-ci, l'équation de mesures (7.9) est remplacée par

$$Y_t = (1 \ 0 \dots 0) X_t + V_t, \quad (7.10)$$

où $\{V_k\}_{k \geq 0}$ est le bruit de mesure. Le polynôme caractéristique de la matrice compagnon A est donné par :

$$\varphi_A(\lambda) \stackrel{\text{def}}{=} \det(\lambda I - A) = \lambda^p - \phi_1 \lambda^{p-1} - \dots - \phi_p \quad (7.11)$$

La matrice compagnon A est dite *stable* si toutes ses valeurs propres sont de modules strictement inférieurs à 1. La matrice est stable si et seulement si $\phi(z) \neq 0$ pour $|z| \leq 1$ (et donc le processus AR est stationnaire causal).

Exemple 7.2.2 (Représentation d'état d'un modèle ARMA(p, q)). Considérons un modèle ARMA(p, q) donné par la relation

$$X_t - \sum_{j=1}^p \phi_j X_{t-j} = Z_t + \sum_{j=1}^q \theta_j Z_{t-j}, \quad (7.12)$$

où $(Z_t) \sim \text{BB}(0, \sigma^2)$ et $\phi(z) \neq 0$ pour $|z| \leq 1$. On peut donner différentes représentations d'état de ce modèle. Posons $r = \max(p, q+1)$ et $\phi_j = 0$ pour $j > p$ et $\theta_j = 0$ pour $j > q$. Nous avons

$$Y_t = [\theta_{r-1} \theta_{r-2} \dots \theta_0] X_t$$

avec $X_t = [X_t, X_{t-1} \dots X_{t-r+1}]^T$ et

$$X_t = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_r \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} + Z_t \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Il est possible de donner une représentation d'état de dimension plus réduite. Posons $m = \max(p, q)$ et $\phi_j = 0$ pour $j > p$. Alors

$$Y_t = BX_t + Z_t, \quad (7.13)$$

avec $B = [1 \ 0 \dots 0]$ et

$$X_{t+1} = AX_t + HZ_t \quad (7.14)$$

avec

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ \phi_m & \phi_{m-1} & \phi_{m-2} & \dots & \phi_1 \end{pmatrix} \quad \text{et} \quad H = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \dots \\ \psi_{m-1} \\ \psi_m \end{pmatrix}$$

où ψ_1, \dots, ψ_m sont les m premiers coefficients du développement de $\theta(z)/\phi(z)$ convergeant sur le disque $|z| \leq 1$. Posons $H = [\psi_1, \psi_2, \dots, \psi_m]^T$. Nous avons, pour $i \in \{1, \dots, m\}$, $\psi_i = BA^{i-1}H$ et comme $\det(zI - A) = z^m - \sum_{j=1}^m \phi_j z^{m-j}$, le théorème de Cayley-Hamilton montre que

$$A^m - \phi_1 A^{m-1} - \dots - \phi_m I = 0.$$

En itérant les équations (7.13) et (7.14), nous obtenons

$$\begin{aligned} Y_t &= BX_t + Z_t, \\ Y_{t+1} &= BAX_t + BHZ_t + Z_{t+1}, \\ &\vdots \\ Y_{t+m} &= BA^m X_t + BA^{m-1} H Z_t + \dots + Z_{t+m}, \end{aligned}$$

ce qui implique

$$\begin{aligned} &Y_{t+m} - \phi_1 Y_{t+m-1} - \dots - \phi_m Y_t \\ &= [-\phi_m \dots - \phi_1 1] \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \psi_1 & 1 & 0 & \dots & 0 \\ \psi_2 & \psi_1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \psi_m & \psi_{m-1} & \psi_{m-2} & \dots & 1 \end{pmatrix} \begin{pmatrix} Z_t \\ Z_{t+1} \\ Z_{t+2} \\ \dots \\ Z_{t+m} \end{pmatrix} \quad (7.15) \end{aligned}$$

Comme ψ_1, \dots, ψ_m sont les m premiers coefficients du développement de $\theta(z)/\phi(z)$, nous avons, pour $j \in \{1, \dots, m\}$, $\psi_j = \sum_{i=1}^j \phi_i \psi_{j-i} + \theta_j$ (en posant $\theta_0 = 1$). Eq. (7.15) implique $Y_{t+m} - \sum_{i=1}^m \phi_i Y_{t+m-i} = \sum_{i=0}^m \theta_i Z_{t+m-i}$, ce qui montre que (Y_t) vérifie (7.12)

7.3 Le filtre de Kalman

Définition 7.3.1 (Prédicteur linéaire optimal). Soit $X = [X(1), \dots, X(d_x)]^t$ un vecteur aléatoire de dimension d_x et Y_0, \dots, Y_t une famille de vecteurs aléatoires de dimensions d_y , éléments de $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Nous supposons de plus que $\mathbb{E}[X] = 0$ et $\mathbb{E}[Y_i] = 0$ pour $i = 0, \dots, k$. Le prédicteur linéaire optimal du vecteur aléatoire X étant donné les observations Y_0, \dots, Y_t est

le vecteur $[\hat{X}(1), \dots, \hat{X}(d_x)]^t$ dont les composantes $\hat{X}(j)$, $j = 1, \dots, d_x$, sont égales aux projections au sens de L^2 des variables aléatoires $X(j)$ sur les espaces linéaires

$$\text{Vect}(Y_1, \dots, Y_{d_y}) \stackrel{\text{def}}{=} \text{Vect}(\{Y_i(j)\}_{0 \leq i \leq k, 1 \leq j \leq d_y}) .$$

Nous notons

$$\hat{X} = \text{proj}(X | Y_0, \dots, Y_t) = \text{proj}(X | \text{Vect}(Y_0, \dots, Y_t)) .$$

La définition 7.3.1 montre que la j ème composante $X(j)$ du vecteur X est projeté sur l'espace engendré par les composantes des vecteurs Y_1, \dots, Y_t ,

$$\left\{ Y : Y = \sum_{i=0}^k \sum_{j=1}^{d_y} \alpha_{i,j} Y_i(j), \quad \alpha_{i,j} \in \mathbb{R} \right\} .$$

En procédant comme pour les variables scalaires, la projection \hat{X} peut être écrite

$$\hat{X} = \sum_{i=0}^k \Phi_i Y_i ,$$

où Φ_0, \dots, Φ_t sont des matrices $d_x \times d_y$. Les relations d'orthogonalité qui caractérisent la projection de \hat{X} se réduisent à

$$\sum_{i=0}^k \Phi_i \mathbb{E}[Y_i Y_j^t] = \mathbb{E}[X Y_j^t] \quad \text{for } j = 0, \dots, k, \quad (7.16)$$

où $\mathbb{E}[Y_i Y_j^t]$ et $\mathbb{E}[X Y_j^t]$ sont des matrices de dimension $d_y \times d_y$ et $d_x \times d_y$ définies par

$$\begin{aligned} [\mathbb{E}[Y_i Y_j^t]]_{l_1 l_2} &= \mathbb{E}[Y_i(l_1) Y_j(l_2)] , \\ [\mathbb{E}[X Y_j^t]]_{l_1 l_2} &= \mathbb{E}[X(l_1) Y_j(l_2)] . \end{aligned}$$

Le théorème de projection assure qu'il existe au moins une solution à ce système d'équation. Cette solution est unique si la matrice $d_y(k+1) \times d_y(k+1)$

$$\Gamma_t = \begin{pmatrix} \mathbb{E}[Y_0 Y_0^t] & \dots & \mathbb{E}[Y_0 Y_t^t] \\ \vdots & & \vdots \\ \mathbb{E}[Y_n Y_0^t] & \dots & \mathbb{E}[Y_n Y_n^t] \end{pmatrix}$$

est inversible. Comme dans le cas scalaire, la matrice de covariance de l'erreur de prédiction s'exprime de la façon suivante

$$\text{Cov}(X - \hat{X}) = \mathbb{E}[X(X - \hat{X})^t] = \mathbb{E}[X X^t] - \mathbb{E}[\hat{X} \hat{X}^t] . \quad (7.17)$$

Notons que, pour toute matrice A (déterministe) de dimension appropriée

$$\text{proj}(AX|Y_0, \dots, Y_t) = A \text{proj}(X|Y_0, \dots, Y_t) . \quad (7.18)$$

7.3.1 Innovation

Considérons l'espace $\text{Vect}(Y_0, \dots, Y_j)$ engendré par les observations jusqu'à l'instant j . Par analogie avec la procédure d'orthogonalisation de Gram-Schmidt, nous pouvons remplacer les vecteurs aléatoires $\{Y_0, \dots, Y_j\}$ par un ensemble de vecteurs aléatoires $\{\epsilon_0, \dots, \epsilon_j\}$ décorrelés dont les composantes engendrent le même sous-espace de L^2 :

$$\text{Vect}(Y_0, \dots, Y_j) = \text{Vect}(\epsilon_0, \dots, \epsilon_j) \quad \text{pour tout } j = 0, \dots, k. \quad (7.19)$$

Posons $\epsilon_0 = Y_0$ puis définissons par récurrence la suite de vecteurs aléatoires, pour $j \geq 0$

$$\epsilon_{j+1} = Y_{j+1} - \text{proj}(Y_{j+1} | \text{Vect}(Y_0, \dots, Y_j)) \quad (7.20)$$

La projection de Y_{j+1} sur $\text{Vect}(Y_0, \dots, Y_j) = \text{Vect}(\epsilon_0, \dots, \epsilon_j)$ a une forme explicite car, par construction, les vecteurs aléatoires $\epsilon_0, \dots, \epsilon_j$ sont décorrelés :

$$\text{proj}(Y_{j+1} | \text{Vect}(\epsilon_0, \dots, \epsilon_j)) = \sum_{i=0}^j \mathbb{E}[Y_{j+1} \epsilon_i^t] [\mathbb{E}[\epsilon_i \epsilon_i^t]]^{-1} \epsilon_i, \quad (7.21)$$

Cette relation conduit à la formule récursive :

$$\epsilon_{j+1} = Y_{j+1} - \sum_{i=0}^j \mathbb{E}[Y_{j+1} \epsilon_i^t] [\mathbb{E}[\epsilon_i \epsilon_i^t]]^{-1} \epsilon_i. \quad (7.22)$$

Le processus $\{\epsilon_j, j \geq 0\}$ est appelée *innovation* (ou innovation linéaire) du processus $\{Y_j, j \geq 0\}$. La suite d'innovation $\{\epsilon_j\}_{j \geq 0}$, définie récursivement par l'équation (7.22) est décorrelée et vérifie, pour tout $j \geq 0$,

$$\epsilon_j \in \text{Vect}(Y_0, \dots, Y_j) \quad \text{and} \quad Y_j \in \text{Vect}(\epsilon_0, \dots, \epsilon_j). \quad (7.23)$$

Pour éviter la dégénérescence de (7.21) et (7.22), nous devons supposer que la matrice de covariance $\mathbb{E}[\epsilon_j \epsilon_j^t]$ est définie positive.

Définition 7.3.2 (Processus non déterministe). Le processus (Y_k) est *non-déterministe* si pour tout $j \geq 0$ la matrice

$$\text{Cov}[Y_{j+1} - \text{proj}(Y_{j+1} | Y_0, \dots, Y_j)]$$

est définie positive.

Soit $Z \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ et soit $\hat{Z}_{|t}$ le prédicteur linéaire optimal de Z étant donné les observations jusqu'à l'instant k . $\hat{Z}_{|t}$ vérifie la récursion

$$\begin{aligned}\hat{Z}_{|t} &= \sum_{i=0}^k \mathbb{E}[Z\epsilon_i^t] [\mathbb{E}[\epsilon_i\epsilon_i^t]]^{-1} \epsilon_i \\ &= \hat{Z}_{|k-1} + \mathbb{E}[Z\epsilon_t^t] [\mathbb{E}[\epsilon_t\epsilon_t^t]]^{-1} \epsilon_t.\end{aligned}\tag{7.24}$$

La matrice de covariance de l'erreur de prédiction vérifie

$$\begin{aligned}\text{Cov}(Z - \hat{Z}_{|t}) &= \text{Cov}(Z) - \text{Cov}(\hat{Z}_{|t}) \\ &= \text{Cov}(Z) - \sum_{i=0}^k \mathbb{E}[Z\epsilon_i^t] [\mathbb{E}[\epsilon_i\epsilon_i^t]]^{-1} \mathbb{E}[\epsilon_i Z^t] \\ &= \text{Cov}(Z) - \text{Cov}(\hat{Z}_{|k-1}) - \mathbb{E}[Z\epsilon_t^t] [\mathbb{E}[\epsilon_t\epsilon_t^t]]^{-1} \mathbb{E}[\epsilon_t Z^t].\end{aligned}\tag{7.25}$$

7.3.2 Equations de prédiction et de filtrage

Considérons le modèle d'état

$$X_{t+1} = A_t X_t + R_t U_t,\tag{7.26}$$

$$Y_t = B_t X_t + S_t V_t,\tag{7.27}$$

où le bruit d'état $\{U_k\}_{k \geq 0}$ et de mesure $\{V_k\}_{k \geq 0}$ sont des bruits blancs faibles (moyenne nulle et covariance identité), décorrélés. L'état initial X_0 est supposé décorrélé des bruits d'états et de mesure $\{U_k\}_{k \geq 0}$ and $\{V_k\}_{k \geq 0}$ et est tel que $\mathbb{E}[X_0] = 0$ et $\text{Cov}(X_0) = \Sigma_0$. Nous supposons aussi que le processus (Y_k) est non déterministe (définition 7.3.2). L'équation (7.27) montre qu'une condition suffisante pour que les observations soient non déterministes est que la matrice de covariance du bruit $S_k S_k^t$ soit définie positive pour tout $k \geq 0$.

Pour tout processus scalaire ou vectoriel $\{Z_t\}_{k \geq 0}$, nous notons $\hat{Z}_{t|n}$ la projection de Z_t sur l'espace linéaire engendré par les vecteurs d'observations Y_0, \dots, Y_n . Avec cette convention, $\hat{X}_{t|t-1}$ correspond au prédicteur linéaire optimal (à un pas) de l'état et $\hat{Y}_{t|t-1}$ à la prédiction linéaire optimale de l'observation étant donné l'état. Le processus d'innovation ϵ_t introduit au paragraphe précédent est par définition égal à la différence $Y_t - \hat{Y}_{t|t-1}$ entre l'observation à la date k et sa prédiction linéaire optimale. Nous utiliserons de plus les notations suivantes :

$$\Gamma_t \stackrel{\text{def}}{=} \text{Cov}(\epsilon_t) \quad \text{and} \quad \Sigma_{t|n} \stackrel{\text{def}}{=} \text{Cov}(X_t - \hat{X}_{t|n}).$$

En projetant (7.27) sur $\text{Vect}(Y_0, \dots, Y_{t-1})$ nous obtenons

$$\hat{Y}_{t|t-1} = B_t \hat{X}_{t|t-1} + S_t \hat{V}_{t|t-1}. \quad (7.28)$$

Sous les hypothèses que nous avons faites sur le modèle d'état $\mathbb{E}[V_t Y_j^t] = 0$ pour $j = 0, \dots, k-1$, de telle sorte que $\hat{V}_{t|t-1} = 0$. Par conséquent

$$\epsilon_t = Y_t - \hat{Y}_{t|t-1} = Y_t - B_t \hat{X}_{t|t-1}. \quad (7.29)$$

Nous appliquons ensuite la décomposition (7.24) au vecteur aléatoire X_{t+1} pour obtenir l'équation de mise à jour du prédictor. Equation (7.24) appliquée à $Z = X_{t+1}$ implique

$$\hat{X}_{t+1|t} = \hat{X}_{t+1|k-1} + \mathbb{E}[X_{t+1} \epsilon_t^t] [\mathbb{E}[\epsilon_t \epsilon_t^t]]^{-1} \epsilon_t. \quad (7.30)$$

En projetant l'équation d'état (7.26) sur l'espace engendré par Y_0, \dots, Y_{t-1} , nous obtenons

$$\hat{X}_{t+1|k-1} = A_t \hat{X}_{t|t-1} + R_t \hat{U}_{t|t-1} = A_t \hat{X}_{t|t-1}, \quad (7.31)$$

où nous avons utilisé que $\mathbb{E}[U_t Y_j^t] = 0$ pour les indices $j = 0, \dots, k-1$. Par conséquent, (7.30) s'exprime

$$\hat{X}_{t+1|t} = A_t \hat{X}_{t|t-1} + H_t \epsilon_t, \quad (7.32)$$

où H_t , est appelé le *gain de Kalman* donné par

$$H_t \stackrel{\text{def}}{=} \mathbb{E}[X_{t+1} \epsilon_t^t] \Gamma_t^{-1}. \quad (7.33)$$

Pour calculer le gain de Kalman, notons que

$$\epsilon_t = Y_t - B_t \hat{X}_{t|t-1} = B_t (X_t - \hat{X}_{t|t-1}) + S_t V_t. \quad (7.34)$$

En utilisant la relation $\mathbb{E}[V_t (X_t - \hat{X}_{t|t-1})^t] = 0$, (7.34) implique que

$$\Gamma_t = B_t \Sigma_{t|t-1} B_t^t + S_t S_t^t, \quad (7.35)$$

où $\Sigma_{t|t-1}$ est la covariance de l'erreur de prédiction $X_t - \hat{X}_{t|t-1}$. En utilisant le même principe,

$$\begin{aligned} \mathbb{E}[X_{t+1} \epsilon_t^t] &= A_t \mathbb{E}[X_t \epsilon_t^t] + R_t \mathbb{E}[U_t \epsilon_t^t] \\ &= A_t \Sigma_{t|t-1} B_t^t + R_t \mathbb{E}[U_t (X_t - \hat{X}_{t|t-1})^t] B_t^t \\ &= A_t \Sigma_{t|t-1} B_t^t, \end{aligned} \quad (7.36)$$

où nous avons utilisé

$$U_t \perp \text{Vect}(X_0, U_0, \dots, U_{t-1}, V_0, \dots, V_{t-1}) \supseteq \text{Vect}(X_t, Y_0, \dots, Y_{t-1}).$$

En combinant (7.35) et (7.36) conduit à l'expression suivante du gain de Kalman :

$$H_t = A_t \Sigma_{t|t-1} B_t^t \{B_t \Sigma_{t|t-1} B_t^t + S_k S_k^t\}^{-1}. \quad (7.37)$$

Finalement, nous devons évaluer la covariance de l'erreur de prédiction $\Sigma_{t+1|t}$. En utilisant l'équation d'état $X_{t+1} = A_t X_t + R_t U_t$ et $\mathbb{E}[X_t U_t^t] = 0$, nous obtenons

$$\text{Cov}(X_{t+1}) = A_t \text{Cov}(X_t) A_t^t + R_t R_t^t. \quad (7.38)$$

De façon similaire, le prédicteur linéaire optimal à la date k $\hat{X}_{t|t-1}$ et l'innovation à la date k ϵ_t sont aussi décorréliées. Par conséquent,

$$\text{Cov}(\hat{X}_{t+1|t}) = A_t \text{Cov}(\hat{X}_{t|t-1}) A_t^t + H_t \Gamma_t H_t^t. \quad (7.39)$$

En utilisant (7.17), nous obtenons

$$\begin{aligned} \Sigma_{t+1|t} &= \text{Cov}(X_{t+1}) - \text{Cov}(\hat{X}_{t+1|t}) \\ &= A_t \Sigma_{t|t-1} A_t^t + R_t R_t^t - H_t \Gamma_t H_t^t, \end{aligned} \quad (7.40)$$

en soustrayant (7.39) de (7.38). L'équation (7.40) est appelée *équation de Riccati*. En rassemblant (7.29), (7.32), (7.35), (7.37), et (7.40), nous obtenons la forme classique que *filtre de Kalman*

Algorithme 7.3.3 (Récursion de Kalman pour la prédiction)

Initialisation : $\hat{X}_{0|-1} = 0$ et $\Sigma_{0|-1} = \Sigma_\nu$.

Récursion : Pour $k = 0, \dots, n$,

$$\epsilon_t = Y_t - B_t \hat{X}_{t|t-1}, \quad \text{innovation} \quad (7.41)$$

$$\Gamma_t = B_t \Sigma_{t|t-1} B_t^t + S_k S_k^t, \quad \text{cov. innovation} \quad (7.42)$$

$$H_t = A_t \Sigma_{t|t-1} B_t^t \Gamma_t^{-1}, \quad \text{Gain de Kalman} \quad (7.43)$$

$$\hat{X}_{t+1|t} = A_t \hat{X}_{t|t-1} + H_t \epsilon_t, \quad \text{predict. état} \quad (7.44)$$

$$\Sigma_{t+1|t} = (A_t - H_t B_t) \Sigma_{t|t-1} A_t^t + R_k R_k^t. \quad \text{cov. erreur de prédiction} \quad (7.45)$$

Remarque 7.3.4. L'évaluation de la vraisemblance pour un modèle d'état général est difficile. Pour un modèle linéaire gaussien, ϵ_t et Γ_t déterminent complètement la densité de probabilité de l'observation Y_t conditionnellement aux observations Y_0, \dots, Y_{t-1} ,

$$(2\pi)^{-d_y/2} |\Gamma_t|^{-1/2} \exp \left\{ -\frac{1}{2} \epsilon_t^t \Gamma_t^{-1} \epsilon_t \right\} \quad (7.46)$$

où d_y est la dimension du vecteur d'observations. Par conséquent, la log-vraisemblance du vecteur d'observations jusqu'à la date n peut être obtenu

de la façon suivante

$$\ell_n = -\frac{(n+1)d_y}{2} \log(2\pi) - \frac{1}{2} \sum_{k=0}^n \{ \log |\Gamma_t| + \epsilon_t^t \Gamma_t^{-1} \epsilon_t \} , \quad (7.47)$$

qui peut être calculé (récursivement en n) en utilisant l'algorithme 7.3.3. L'équation (7.47) s'avère très importante pour l'estimation des paramètres dans un modèle d'état.

Exemple 7.3.5 (Filtrage d'une marche aléatoire observée en présence de bruit). Afin d'illustrer l'utilisation de l'algorithme 7.3.3 sur un exemple élémentaire, considérons une marche aléatoire discrète observée en présence de bruit

$$\begin{aligned} X_{t+1} &= X_t + \sigma_u U_t , \\ Y_t &= X_t + \sigma_v V_t . \end{aligned}$$

Dans cet exemple, toutes les variables sont scalaires. En appliquant les équations de Kalman (pour la prédiction), nous obtenons, pour $k \geq 1$,

$$\hat{X}_{t+1|t} = \hat{X}_{t|t-1} + \frac{\Sigma_{t|t-1}}{\Sigma_{t|t-1} + \sigma_v^2} (Y_t - \hat{X}_{t|t-1}) \quad (7.48)$$

$$= (1 - a_t) \hat{X}_{t|t-1} + a_t Y_t ,$$

$$\begin{aligned} \Sigma_{t+1|t} &= \Sigma_{t|t-1} + \sigma_u^2 - \frac{\Sigma_{t|t-1}^2}{\Sigma_{t|t-1} + \sigma_v^2} \\ &= \frac{\Sigma_{t|t-1} \sigma_v^2}{\Sigma_{t|t-1} + \sigma_v^2} + \sigma_u^2 \stackrel{\text{def}}{=} f(\Sigma_{t|t-1}) , \end{aligned} \quad (7.49)$$

où nous avons noté $a_t = \Sigma_{t|t-1} / (\Sigma_{t|t-1} + \sigma_v^2)$. La récursion est initialisée par $\hat{X}_{0|-1} = 0$ et $\Sigma_{0|-1} = \Sigma_\nu$. Dans ce cas particulier, les matrices de transition d'états et d'observations sont indépendantes du temps et il est intéressant de considérer le comportement en temps long du filtre de Kalman. Il est facile de montrer que la matrice de covariance de prédiction converge vers une limite Σ , qui est une solution de l'équation

$$\Sigma = f(\Sigma) = \frac{\Sigma \sigma_v^2}{\Sigma + \sigma_v^2} + \sigma_u^2 .$$

En résolvant cette solution (en prenant la solution positive) nous obtenons

$$\Sigma = \frac{1}{2} \left(\sigma_u^2 + \sqrt{\sigma_u^4 + 4\sigma_u^2 \sigma_v^2} \right) .$$

Pour $M < \infty$, $\sup_{0 \leq \Sigma \leq M} |\dot{f}(\Sigma)| < 1$. De plus, pour $k \geq 1$, $(\Sigma_{t+1|t} - \Sigma_\infty)(\Sigma_{t|t-1} - \Sigma_\infty) \geq 0$. Ces remarques montrent que $\Sigma_{t+1|t}$ appartient pour tout k à l'intervalle $\Sigma_{t|t-1}$ et Σ_∞ , et en particulier $\Sigma_{t+1|t} \leq \max(\Sigma_{1|0}, \Sigma_\infty)$.

Comme le coefficient de Lipshitz de f sur tout ensemble borné de \mathbb{R}^+ est strictement inférieur à 1, indépendamment de la valeur de Σ_ν , et donc les coefficients $a_t = \Sigma_{t|t-1}/(\Sigma_{t|t-1} + \sigma_v^2)$ converge vers

$$a_\infty = \frac{\Sigma_\infty}{\Sigma_\infty + \sigma_v^2},$$

et l'erreur quadratique $(Y_{t+1} - \hat{Y}_{t+1|t})$ converge vers $\Sigma_\infty + \sigma_v^2$.

7.3.2.1 Equations de Kalman pour le filtre

L'algorithme 7.3.3 permet d'évaluer de façon récursive le prédicteur linéaire optimal de l'état $\hat{X}_{t|t-1}$ et la covariance de l'erreur de prédiction $\Sigma_{t|t-1}$. Il est bien entendu possible d'obtenir des équations similaires pour la moyenne de la distribution de filtrage $\hat{X}_{t|t}$ et la covariance associée $\Sigma_{t|t}$.

Nous utilisons cette fois encore l'équation (7.24), que nous appliquons avec $Z = X_t$, pour obtenir

$$\hat{X}_{t|t} = \hat{X}_{t|t-1} + \mathbb{E}[X_t \epsilon_t^t] \Gamma_t^{-1} \epsilon_t = \hat{X}_{t|t-1} + K_t \epsilon_t \quad (7.50)$$

où $K_t \stackrel{\text{def}}{=} \text{Cov}(X_t, \epsilon_t) \Gamma_t^{-1}$ est le gain de Kalman (pour le filtrage). Le membre de droite de (7.50) peut être réécrit de la façon suivante

$$\hat{X}_{t|t-1} = A_{t-1} \hat{X}_{k-1|k-1} + R_{t-1} \hat{U}_{k-1|k-1} = A_{t-1} \hat{X}_{k-1|k-1}, \quad (7.51)$$

où nous avons utilisé

$$U_{t-1} \perp \text{Vect}(X_0, U_0, \dots, U_{k-2}) \supseteq \text{Vect}(Y_0, \dots, Y_{t-1}).$$

De façon similaire, le second terme du membre de droite de (7.50) peut être écrit de façon équivalente

$$K_t = \Sigma_{t|t-1} B_t^t \Gamma_t^{-1}, \quad (7.52)$$

car $\epsilon_t = B_t(X_t - \hat{X}_{t|t-1}) + S_t V_t$ et $\mathbb{E}[X_t V_t^t] = 0$.

Pour obtenir une récursion complète, il reste à relier $\Sigma_{t|t}$ et $\Sigma_{t|t-1}$. L'équation d'état $X_t = A_{t-1} X_{t-1} + R_{t-1} U_{t-1}$ et l'équation de prédiction d'état $\hat{X}_{t|t-1} = A_{t-1} \hat{X}_{k-1|k-1}$ implique que

$$\begin{aligned} \text{Cov}(X_t) &= A_{t-1} \text{Cov}(X_{t-1}) A_{t-1}^t + R_{t-1} R_{t-1}^t, \\ \text{Cov}(\hat{X}_{t|t-1}) &= A_{t-1} \text{Cov}(\hat{X}_{k-1|k-1}) A_{t-1}^t, \end{aligned}$$

ce qui, combiné avec (7.17), produit

$$\Sigma_{t|t-1} = A_{t-1} \Sigma_{k-1|k-1} A_{t-1}^t + R_{t-1} R_{t-1}^t. \quad (7.53)$$

De façon similaire, la récursion d'état $X_t = A_{t-1}X_{t-1} + R_{t-1}U_{t-1}$ et l'équation de mise à jour du filtre $\hat{X}_{t|t} = A_{t-1}\hat{X}_{t-1|k-1} + K_t\epsilon_t$ implique que

$$\Sigma_{t|t} = A_{t-1}\Sigma_{t-1|k-1}A_{t-1}^t + R_{t-1}R_{t-1}^t - K_t\Gamma_tK_t^t. \quad (7.54)$$

En rassemblant ces différentes récursions, nous obtenons les équations de Kalman pour le filtre

Algorithme 7.3.6 (Filtrage de Kalman) *Pour $k = 0, \dots, n$*

a) *Si $k = 0$, posons $\hat{X}_{t|t-1} = 0$ et $\Sigma_{t|t-1} = \Sigma_\nu$; autrement*

$$\begin{aligned} \hat{X}_{t|t-1} &= A_{t-1}\hat{X}_{t-1|k-1}, \\ \Sigma_{t|t-1} &= A_{t-1}\Sigma_{t-1|k-1}A_{t-1}^t + R_{t-1}R_{t-1}^t. \end{aligned}$$

b) *Mise à jour*

$$\epsilon_t = Y_t - B_t\hat{X}_{t|t-1}, \quad \text{innovation} \quad (7.55)$$

$$\Gamma_t = B_t\Sigma_{t|t-1}B_t^t + S_kS_k^t, \quad \text{cov. innovation} \quad (7.56)$$

$$K_t = \Sigma_{t|t-1}B_t^t\Gamma_t^{-1}, \quad \text{gain de Kalman (filtre)} \quad (7.57)$$

$$\hat{X}_{t|t} = \hat{X}_{t|t-1} + K_t\epsilon_t, \quad \text{estim. moy. du filtre} \quad (7.58)$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - K_tB_t\Sigma_{t|t-1}. \quad \text{cov. filtre} \quad (7.59)$$

Remarque 7.3.7. Les adaptations nécessaires pour transformer les récursions de prédiction et de filtrage au cas où les bruits d'états et de mesures ne sont pas centrés sont directs. L'idée est de définir un modèle d'état sur les variables centrées $X_t^* = X_t - \mathbb{E}[X_t]$, $U_t^* = U_t - \mathbb{E}[U_t]$, $Y_t^* = Y_t - \mathbb{E}[Y_t]$, and $V_t^* = V_t - \mathbb{E}[V_t]$; les moyennes des variables d'états et des mesures peuvent être calculés directement à partir des équations d'états et de mesure

$$\begin{aligned} \mathbb{E}[X_{t+1}] &= A_t\mathbb{E}[X_t] + R_t\mathbb{E}[U_t], \\ \mathbb{E}[Y_t] &= B_t\mathbb{E}[X_t] + S_t\mathbb{E}[V_t]. \end{aligned}$$

Il est clair que

$$\begin{aligned} X_{t+1}^* &= X_{t+1} - \mathbb{E}[X_{t+1}] = A_t(X_t - \mathbb{E}[X_t]) + R_t(U_t - \mathbb{E}[U_t]) \\ &= A_tX_t^* + R_tU_t^* \end{aligned}$$

et de façon similaire

$$Y_t^* = Y_t - \mathbb{E}[Y_t] = B_tX_t^* + S_tV_t^*.$$

Par conséquent $\{X_t^*, Y_t^*\}_{k \geq 0}$ satisfont le modèle (7.26)–(7.27) avec $X_0^* = 0$, $\mathbb{E}[U_t^*] = 0$ and $\mathbb{E}[V_t^*] = 0$. Les récursions de Kalman peuvent être utilisées directement pour calculer $\hat{X}_{t|t-1}^*$, le prédicteur linéaire optimal de l'état X_t^*

étant donné Y_0^*, \dots, Y_{t-1}^* . Le prédicteur linéaire optimal de X_t étant donnés Y_0, \dots, Y_{t-1} est alors

$$\hat{X}_{t|t-1} = \hat{X}_{t|t-1}^* + \mathbb{E}[X_t] .$$

Les équations de filtrage peuvent être obtenus de façon similaire.

7.3.3 Equations de lissage

Pour $k = 0, \dots, n-1$ nous notons $\hat{U}_{t|n}$ la moyenne de lissage du bruit d'état, *i.e.*, le prédicteur linéaire optimal du bruit d'état U_t en fonction des observations Y_0, \dots, Y_n . Nous notons $\Xi_{t|n}$ la matrice de covariance

$$\Xi_{t|n} \stackrel{\text{def}}{=} \text{Cov}(U_t - \hat{U}_{t|n}) .$$

Nous allons tout d'abord donner la forme des récursions ; nous justifierons ensuite ces récursions

Algorithme 7.3.8 (Lissage du bruit d'état) *Filtrage direct : Appliquer le filtre de Kalman (Algorithme 7.3.3) et mémoriser pour $k = 0, \dots, n$ les innovations ϵ_t , l'inverse de la covariance de l'innovation Γ_t^{-1} , et la covariance de prédiction $\Sigma_{t|t-1}$, et*

$$A_t \stackrel{\text{def}}{=} A_t - H_t B_t ,$$

où H_t est le gain de Kalman (forme prédiction).

Lissage arrière : Pour $k = n-1, \dots, 0$, calculer

$$p_t = \begin{cases} B_n^t \Gamma_n^{-1} \epsilon_n & \text{for } k = n-1, \\ B_{t+1}^t \Gamma_{t+1}^{-1} \epsilon_{t+1} + A_{t+1}^t p_{t+1} & \text{autrement,} \end{cases} \quad (7.60)$$

$$C_t = \begin{cases} B_n^t \Gamma_n^{-1} B_n & \text{for } k = n-1, \\ B_{t+1}^t \Gamma_{t+1}^{-1} B_{t+1} + A_{t+1}^t C_{t+1} A_{t+1} & \text{autrement,} \end{cases} \quad (7.61)$$

$$\hat{U}_{t|n} = R_t^t p_t , \quad (7.62)$$

$$\Xi_{t|n} = I - R_t^t C_t R_t . \quad (7.63)$$

Initialisation du lissage de l'état : Calculer

$$\hat{X}_{0|n} = \Sigma_\nu (B_0^t \Gamma_0^{-1} \epsilon_0 + A_0^t p_0) , \quad (7.64)$$

$$\Sigma_{0|n} = \Sigma_\nu - \Sigma_\nu [B_0^t \Gamma_0^{-1} B_0 + A_0^t C_0 A_0] \Sigma_\nu . \quad (7.65)$$

Lissage de l'état : Pour $k = 0, \dots, n-1$,

$$\hat{X}_{t+1|n} = A_t \hat{X}_{t|n} + R_t \hat{U}_{t|n}, \quad (7.66)$$

$$\begin{aligned} \Sigma_{t+1|n} &= A_t \Sigma_{t|n} A_t^t + R_t \Xi_{t|n} R_t^t \\ &\quad - A_t \Sigma_{t|t-1} A_t^t C_t R_k R_k^t - R_k R_k^t C_t A_t \Sigma_{t|t-1} A_t^t. \end{aligned} \quad (7.67)$$

L'algorithme 7.3.8 est assez complexe, comportant trois étapes. La première consiste à appliquer les équations de Kalman (forme prédiction). La seconde est une passe arrière pour obtenir les estimateurs lissés du bruit de mesure. La troisième étape consiste à calculer dans une passe avant les estimateurs de lissage des états en utilisant les estimateurs lissés du bruit d'état. La preuve de la correction des récursions ci-dessus est divisée en deux parties, correspondant aux deux dernières étapes de l'algorithme.

Preuve (Lissage arrière). Nous commençons par justifier les équations requises pour calculer le lissage du bruit d'état $\hat{U}_{t|n}$ pour $k = n - 1$ à 0 (en décroissant). Comme précédemment, nous utiliserons la suite des innovations $\{\epsilon_0, \dots, \epsilon_n\}$ plutôt que directement les observations $\{Y_0, \dots, Y_n\}$. En utilisant (7.24), nous obtenons :

$$\hat{U}_{t|n} = \sum_{i=0}^n \mathbb{E} [U_t \epsilon_i^t] \Gamma_i^{-1} \epsilon_i = \sum_{i=k+1}^n \mathbb{E} [U_t \epsilon_i^t] \Gamma_i^{-1} \epsilon_i, \quad (7.68)$$

où nous avons utilisé la propriété

$$U_t \perp \text{Vect}(Y_0, \dots, Y_t) = \text{Vect}(\epsilon_0, \dots, \epsilon_t),$$

pour obtenir la deuxième expression. Nous allons prouver par récurrence que pour $i = k + 1, \dots, n$,

$$\mathbb{E} [U_t (X_i - \hat{X}_{i|i-1})^t] = \begin{cases} R_t^t, & i = k + 1, \\ R_t^t A_{t+1}^t A_{k+2}^t \dots A_{i-1}^t, & i \geq k + 2, \end{cases} \quad (7.69)$$

$$\mathbb{E} [U_t \epsilon_i^t] = \begin{cases} R_t^t B_{t+1}^t, & i = k + 1, \\ R_t^t A_{t+1}^t A_{k+2}^t \dots A_{i-1}^t B_i^t, & i \geq k + 2. \end{cases} \quad (7.70)$$

Notons tout d'abord que

$$\begin{aligned} \mathbb{E} [U_t \epsilon_{t+1}^t] &= \mathbb{E} [U_t (X_{t+1} - \hat{X}_{t+1|t})^t] B_{t+1}^t \\ &= \mathbb{E} [U_t X_{t+1}^t] B_{t+1}^t = R_t^t B_{t+1}^t, \end{aligned}$$

en utilisant (7.29) et les relations d'orthogonalité $U_t \perp V_{t+1}$, $U_t \perp \text{Vect}(Y_0, \dots, Y_t)$ et $U_t \perp X_t$. Supposons maintenant que les relations (7.69)–(7.70) sont satisfaites pour $i \geq k + 1$. En combinant l'équation d'état (7.26) et l'équation de mise à jour du prédictor (7.32), nous obtenons

$$X_{i+1} - \hat{X}_{i+1|i} = A_i (X_i - \hat{X}_{i|i-1}) + R_i U_i - H_i S_i V_i. \quad (7.71)$$

En utilisant $\mathbb{E}[U_t U_i^t] = 0$ et $\mathbb{E}[U_t V_i^t] = 0$, l'hypothèse de récurrence implique que

$$\mathbb{E}\left[U_t(X_{i+1} - \hat{X}_{i+1|i})^t\right] = \mathbb{E}\left[U_t(X_i - \hat{X}_{i|i-1})^t\right] A_i^t = R_t^t A_{t+1}^t A_{k+2}^t \dots A_i^t. \quad (7.72)$$

En procédant comme dans le cas $i = k$ ci-dessus, nous écrivons

$$\mathbb{E}[U_t \epsilon_{i+1}^t] = \mathbb{E}\left[U_t(X_{i+1} - \hat{X}_{i+1|i})^t\right] B_{i+1}^t = R_t^t A_{t+1}^t A_{k+2}^t \dots A_i^t B_{i+1}^t, \quad (7.73)$$

qui, par récurrence, montre que (7.69)–(7.70) est vérifié pour tous les indices $i \geq k+1$. En combinant (7.70) et (7.68), nous obtenons

$$\hat{U}_{t|n} = R_t^t \left(B_{t+1}^t \Gamma_{t+1}^{-1} \epsilon_{t+1} + \sum_{i=k+2}^n A_{t+1}^t \dots A_{i-1}^t B_i^t \Gamma_i^{-1} \epsilon_i \right), \quad (7.74)$$

où le terme entre parenthèse correspond à p_t défini récursivement par la relation (7.60) : ceci montre (7.62).

Pour calculer la covariance $\Xi_{t|n}$ de l'erreur $U_t - \hat{U}_{t|n}$, nous appliquons Eq. (7.25) qui implique

$$\Xi_{t|n} = \text{Cov}(U_t) - \text{Cov}(\hat{U}_{t|n}) \quad (7.75)$$

$$\begin{aligned} &= I - \sum_{i=k+1}^n \mathbb{E}[U_t \epsilon_i^t] \Gamma_i^{-1} \mathbb{E}[\epsilon_i U_t^t] \\ &= I - R_t^t \left(B_{t+1}^t \Gamma_{t+1}^{-1} B_{t+1} + \sum_{i=k+2}^n A_{t+1}^t \dots A_{i-1}^t B_i^t \Gamma_i^{-1} B_i A_{i-1} \dots A_{t+1} \right) R_t, \end{aligned}$$

où I est la matrice identité. \square

Preuve (Lissage de l'état). En utilisant une preuve par induction similaire à (7.69)–(7.70), nous obtenons

$$\mathbb{E}\left[X_t(X_i - \hat{X}_{i|i-1})^t\right] = \begin{cases} \Sigma_{t|t-1}, & i = k, \\ \Sigma_{t|t-1} A_t^t A_{t+1}^t \dots A_{i-1}^t, & i \geq k+1, \end{cases} \quad (7.76)$$

$$\mathbb{E}[X_t \epsilon_i^t] = \begin{cases} \Sigma_{t|t-1} B_t^t, & i = k, \\ \Sigma_{t|t-1} A_t^t A_{t+1}^t \dots A_{i-1}^t B_i^t, & i \geq k+1, \end{cases} \quad (7.77)$$

En utilisant (7.24), le prédicteur linéaire optimal de l'état initial X_0 en fonction des observations Y_0, \dots, Y_n est donné par

$$\hat{X}_{0|n} = \sum_{i=0}^n \mathbb{E}[X_0 \epsilon_i^t] \Gamma_i^{-1} \epsilon_i. \quad (7.78)$$

Une application directe de (7.77),

$$\hat{X}_{0|n} = \Sigma_\nu \left(B_0^t \Gamma_0^{-1} \epsilon_0 + \sum_{i=1}^n A_0^t \dots A_{i-1}^t B_i^t \Gamma_i^{-1} \epsilon_i \right), \quad (7.79)$$

montre (7.64). En procédant comme pour (7.75), l'expression de la covariance de l'erreur d'état (7.65) se déduit de (7.25).

L'équation de mise à jour (7.66) est une conséquence directe de la linéarité de la projection combinée avec l'équation (7.26). Finalement, pour prouver (7.67), nous combinons l'équation (7.26) avec (7.66) pour obtenir

$$\begin{aligned} \text{Cov}(X_{t+1} - \hat{X}_{t+1|n}) &= \text{Cov}[A_t(X_t - \hat{X}_{t|n}) + R_t(U_t - \hat{U}_{t|n})] = \\ &A_t \Sigma_{t|n} A_t^t + R_t \Xi_{t|n} R_t^t - A_t \mathbb{E} \left[X_t \hat{U}_{t|n}^t \right] R_t^t - R_t \mathbb{E} \left[\hat{U}_{t|n} X_t^t \right] A_t^t, \end{aligned} \quad (7.80)$$

où nous avons utilisé $\mathbb{E} \left[\hat{X}_{t|n} (U_t - \hat{U}_{t|n})^t \right] = 0$ pour obtenir la seconde relation. Pour calculer $\mathbb{E} \left[X_t \hat{U}_{t|n}^t \right]$, nous utilisons (7.74), en écrivant

$$\mathbb{E} \left[X_t \hat{U}_{t|n}^t \right] = \mathbb{E} \left[X_t \epsilon_{t+1}^t \right] \Gamma_{t+1}^{-1} B_{t+1} R_t + \sum_{i=k+2}^n \mathbb{E} \left[X_t \epsilon_i^t \right] \Gamma_i^{-1} B_i A_{i-1} \dots A_{t+1} R_t. \quad (7.81)$$

Finalement, en invoquant (7.77), nous obtenons

$$\begin{aligned} \mathbb{E} \left[X_t \hat{U}_{t|n}^t \right] &= \Sigma_{t|t-1} A_t^t B_{t+1}^t \Gamma_{t+1}^{-1} B_{t+1} R_t \\ &+ \sum_{i=k+2}^n \Sigma_{t|t-1} A_t^t A_{t+1}^t \dots A_{i-1}^t B_i^t \Gamma_i^{-1} B_i A_{i-1} \dots A_{t+1} R_t, \end{aligned}$$

qui peut être réécrit

$$\mathbb{E} \left[X_t \hat{U}_{t|n}^t \right] = \Sigma_{t|t-1} A_t^t C_t R_t. \quad (7.82)$$

L'équation (7.67) découle de (7.80). \square

7.4 Le modèle linéaire Gaussien

Dans un modèle linéaire Gaussien, les bruits d'états (U_t) et de mesure (V_t) sont gaussiens. Une propriété remarquable des modèles linéaires d'états gaussiens définis par (7.1)–(7.2) est que la suite des vecteurs d'états $X_{0:n}$ et des observations $Y_{0:n}$ sont conjointement gaussiens (pour tout n); par conséquent, les distributions conditionnelles des états étant données les ob-

servations sont elles aussi gaussiennes. Ces distributions conditionnelles sont donc déterminées par la donnée de leurs moyennes et de leurs matrices de covariance. Nous étudierons plus spécifiquement dans la suite de ce chapitre :

- a) la distribution conditionnelle de l'état X_t étant donnée les observations jusqu'à la date $k-1$ et la distribution initiale ν , que nous notons $\phi_{\nu,t|t-1}$; nous appelons cette loi la distribution de *prédiction* (ou distribution *prédictive*),
- b) la distribution de l'état à la date k étant donnée les observations jusqu'à l'instant k et la distribution initiale ν , que nous notons $\phi_{\nu,t}$; nous appelons cette loi la distribution de *filtrage*.

Les distributions $\phi_{t|t-1}$ et ϕ_t sont des fonctions des observations; pour alléger les notations, cette dépendance n'est pas indiquée explicitement. Nous notons $L_{\nu,n}$ la vraisemblance des observations, définie comme la distribution du vecteur des observations $Y_{0:n}$. Notons q_t la distribution de X_{t+1} conditionnelle à $X_t = x$: dans le modèle linéaire gaussien considéré ici, cette loi admet une densité gaussienne, de moyenne $A_t x$, et de covariance $R_k R_k^t$. De façon plus précise, pour toute fonction $f : \mathbb{R}^{d_x} \rightarrow \mathbb{R}_+$:

$$\mathbb{E}[f(X_{t+1}) | X_t = x_t] = \int f(A_t x_t + R_t u) \phi(u) du = \int f(x') q_t(x, x') dx',$$

où $\phi(u)$ est la densité d'une loi $N(0, I)$ et

$$q(x, \cdot) = N(A_t x, R_k R_k^t). \quad (7.83)$$

Nous notons de la même façon g_t la loi de l'observation Y_t conditionnelle à l'état $X_t = x$: dans le modèle linéaire gaussien, cette loi admet une densité gaussienne de moyenne $B_t x$ et de covariance $S_k S_k^t$. Plus précisément, pour toute fonction $f : \mathbb{R}^{d_y} \rightarrow \mathbb{R}_+$,

$$\mathbb{E}[f(Y_t) | X_t = x_t] = \int f(B_t x_t + S_t v) \phi(v) dv = \int f(y) g_t(x, y) dy,$$

où

$$g_t(x, y) = N(B_t x, S_k S_k^t).$$

En utilisant ces notations et les relations d'indépendance conditionnelles du modèle linéaire gaussien, la loi de prédiction et de filtrage sont données pour $k \geq 1$ et toute fonction $f : \mathbb{R}^{d_x} \rightarrow \mathbb{R}^+$,

$$\phi_{\nu,t|t-1}(f) \stackrel{\text{def}}{=} (L_{\nu,k-1})^{-1} \iint \alpha_{\nu,k-1}(x_{t-1}) q_{t-1}(x_{t-1}, x_t) f(x_t) \quad (7.84)$$

$$\phi_{\nu,t}(f) \stackrel{\text{def}}{=} (L_{\nu,t})^{-1} \iint \alpha_{\nu,t}(x_t) f(x_t) \quad (7.85)$$

où α_t est la fonction donnée par

$$\alpha_{\nu,t}(x_t) = \int \cdots \int \nu(dx_0) g_0(x_0, Y_0) \prod_{i=1}^k q_{i-1}(x_{i-1}, x_i) g_i(x_i, Y_i) dx_{0:k-1} \quad (7.86)$$

et $L_{\nu,t}$ est la vraisemblance des k premières observations définie par

$$L_{\nu,t} = \int \alpha_{\nu,t}(x_t) dx_t. \quad (7.87)$$

En utilisant les définitions précédentes, il est possible de calculer de façon récursive les lois prédictive et de filtrage : pour tout $k \geq 1$,

$$\begin{aligned} \phi_{\nu,t}(f) &= (L_{\nu,t})^{-1} \int \alpha_{\nu,t}(x_t) f(x_t) dx_t \\ &= (L_{\nu,t})^{-1} \iint \alpha_{\nu,k-1}(x_{t-1}) q_{t-1}(x_{t-1}, x_t) g(x_t, Y_t) f(x_t) dx_{k-1:t}, \\ &= \left(\frac{L_{\nu,k-1}}{L_{\nu,t}} \right) \int \phi_{\nu,t|t-1}(x_t) g(x_t, Y_t) f(x_t) dx_t. \end{aligned} \quad (7.88)$$

Cette relation montre qu'il est possible de calculer de déduire la loi de filtrage à la date k de la loi de prédiction. De la même façon,

$$\begin{aligned} \phi_{\nu,t+1|t}(f) &= (L_{\nu,t})^{-1} \iint \alpha_{\nu,t}(x_t) q(x_t, x_{t+1}) f(x_{t+1}) dx_{t:t+1} \\ &= \iint \phi_{\nu,t}(x_t) q_t(x_t, x_{t+1}) f(x_{t+1}) dx_{t:t+1}. \end{aligned} \quad (7.89)$$

En appliquant successivement les équations (7.88) et (7.89), on obtient une formule récursive permettant de calculer les lois de prédiction et de filtrage pour tous les indices temporels. Ces formules sont valables de façon tout à fait générale (et s'étendent par exemple au cas où les bruits d'états et de mesure sont des suites de variables aléatoires indépendantes dont la distribution n'est pas nécessairement gaussienne).

Dans le modèle linéaire gaussien, les distributions de prédiction et de filtrage sont gaussiennes et nous notons $\hat{X}_{\nu,t|t-1}$ et $\hat{X}_{\nu,t|t}$ les moyennes des distributions de prédiction et de filtrage et $\Sigma_{\nu,t|t-1}$ et $\Sigma_{\nu,t|t}$ les covariances associées :

$$\phi_{\nu,t|t-1} = N\left(\hat{X}_{\nu,t|t-1}, \Sigma_{\nu,t|t-1}\right), \quad (7.90)$$

$$\phi_{\nu,t} = N\left(\hat{X}_{\nu,t|t}, \Sigma_{\nu,t|t}\right), \quad (7.91)$$

Le lemme élémentaire suivant est très utile pour calculer la distribution des lois de prédiction et de filtrage dans le modèle linéaire Gaussien.

Proposition 7.4.1 (Conditionnement dans le modèle linéaire gaussien).

Soit X et V de vecteurs gaussiens indépendants vérifiant $\mathbb{E}[X] = \mu_X$,

$\text{Cov}(X) = \Sigma_X$, et $\text{Cov}(V) = \Sigma_V$. Supposons que $\mathbb{E}[V] = 0$. Considérons le modèle

$$Y = BX + V, \quad (7.92)$$

où B est une matrice (déterministe) de dimension appropriée. Supposons de plus que $B\Sigma_X B^t + \Sigma_V$ est une matrice de rang complet (invertible). Alors

$$\begin{aligned} \mathbb{E}[X|Y] &= \mathbb{E}[X] + \text{Cov}(X, Y) \{\text{Cov}(Y)\}^{-1} (Y - \mathbb{E}[Y]) \\ &= \mu_X + \Sigma_X B^t \{B\Sigma_X B^t + \Sigma_V\}^{-1} (Y - B\mu_X) \end{aligned} \quad (7.93)$$

et

$$\begin{aligned} \text{Cov}(X | Y) &= \text{Cov}(X - \mathbb{E}[X|Y]) = \mathbb{E}[(X - \mathbb{E}[X|Y])X^t] \\ &= \Sigma_X - \Sigma_X B^t \{B\Sigma_X B^t + \Sigma_V\}^{-1} B\Sigma_X. \end{aligned} \quad (7.94)$$

Preuve. Notons par \hat{X} le membre de droite de l'équation (7.93). Nous avons

$$X - \hat{X} = X - \mathbb{E}[X] - \text{Cov}(X, Y) \{\text{Cov}(Y)\}^{-1} (Y - \mathbb{E}[Y]),$$

ce qui montre que

$$\text{Cov}(X - \hat{X}, Y) = \text{Cov}(X, Y) - \text{Cov}(X, Y) \{\text{Cov}(Y)\}^{-1} \text{Cov}(Y) = 0. \quad (7.95)$$

Les vecteurs aléatoires Y et $X - \hat{X}$ sont donc conjointement Gaussiens (ce sont des transformations linéaires de vecteurs gaussiens) et ils sont décorrélés. Par conséquent, Y et $X - \hat{X}$ sont aussi indépendants. En écrivant

$$X = \hat{X} + (X - \hat{X}),$$

où \hat{X} est $\sigma(Y)$ -mesurable (combinaison linéaire des composantes du vecteur Y) et $X - \hat{X}$ est indépendant de \hat{X} , on vérifie alors aisément que $\hat{X} = \mathbb{E}[X|Y]$ et que, de plus,

$$\text{Cov}(X | Y) \stackrel{\text{def}}{=} \text{Cov}[(X - \hat{X})(X - \hat{X})' | Y] = \text{Cov}(X - \hat{X}).$$

Finalement, nous obtenons (7.94) en remarquant que

$$\text{Cov}(X - \hat{X}) = \mathbb{E}[(X - \hat{X})(X - \hat{X})^t] = \mathbb{E}[(X - \hat{X})X^t],$$

en utilisant (7.95) et le fait que \hat{X} est une transformation linéaire de Y . \square

Pour le modèle linéaire Gaussien, la proposition 7.4.1 implique en particulier que les moyennes a posteriori $\hat{X}_{\nu, t|t-1}$ et $\hat{X}_{\nu, t|t}$ dépendent des observations (sont des combinaisons linéaires des observations passées), mais que les matrices de covariance $\Sigma_{\nu, t|t-1}$ et $\Sigma_{\nu, t|t}$ sont entièrement déterminées par la

donnée du modèle (elles ne dépendent pas des observations et pourraient donc être calculées *hors ligne*).

Proposition 7.4.2 (Loi de filtrage dans le modèle linéaire gaussien).

Les moyennes et covariance et de prédiction peuvent être mises à jour récursivement de la façon suivante : pour $k \geq 0$.

Filtrage :

$$\hat{X}_{t|t} = \hat{X}_{t|t-1} + \Sigma_{t|t-1} B_t^t (B_t \Sigma_{t|t-1} B_t^t + S_k S_k^t)^{-1} (Y_t - B_t \hat{X}_{t|t-1}), \quad (7.96)$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1} B_t^t (B_t \Sigma_{t|t-1} B_t^t + S_k S_k^t)^{-1} B_t \Sigma_{t|t-1}, \quad (7.97)$$

avec les conventions $\hat{X}_{0|-1} = 0$ et $\Sigma_{0|-1} = \Sigma_\nu$.

Prédiction :

$$\hat{X}_{t+1|t} = A_t \hat{X}_{t|t}, \quad (7.98)$$

$$\Sigma_{t+1|t} = A_t \Sigma_{t|t} A_t^t + R_k R_k^t, \quad (7.99)$$

Preuve. Considérons tout d'abord (7.96) et (7.97). L'état est distribué suivant $X_t \sim N(\hat{X}_{t|t-1}, \Sigma_{t|t-1})$ et l'observation est donnée par :

$$Y_t = B_t X_t + V_t,$$

où $V_t \sim N(0, S_k S_k^t)$ est indépendant de X_t . Les équations (7.96) et (7.97) découlent directement de la proposition Proposition 7.4.1.

Les équations (7.98) et (7.99) sont obtenues en calculant la moyenne et la covariance de

$$X_{t+1} = A_t X_t + R_t U_t$$

où X_t et U_t sont indépendants et distribués respectivement suivant les lois $N(\hat{X}_{t|t}, \Sigma_{t|t})$ et $N(0, I)$. \square

7.4.1 Lissage dans le modèle linéaire Gaussien

Nous notons par $\hat{X}_{t|n}$ et $\Sigma_{t|n}$ la moyenne et la covariance de la distribution de lissage $\phi_{\nu,t|n}$. Pour toute fonction $f : \mathbb{R}^{d_x} \rightarrow \mathbb{R}_+$ et tout $k \in \{0, \dots, n\}$, la loi de filtrage est donnée par

$$\phi_{\nu,t|n}(f) = (L_{\nu,n})^{-1} \int \alpha_{\nu,t}(x_t) f(x_t) \beta_{t|n}(x_t) dx_t, \quad (7.100)$$

$$= \frac{\int \alpha_{\nu,t}(x_t) f(x_t) \beta_{t|n}(x_t) dx_t}{\int \alpha_{\nu,t}(x_t) \beta_{t|n}(x_t) dx_t}, \quad (7.101)$$

où $\beta_{t|n}$ est la fonction *rétrograde* définie, pour $k \in \{0, n-1\}$ par

$$\beta_{t|n}(x_t) = \int \cdots \int q_t(x_t, x_{t+1}) g_{t+1}(x_{t+1}, Y_{t+1}) \prod_{i=k+2}^n q_{i-1}(x_{i-1}, x_i) g_i(x_i, Y_i) . \quad (7.102)$$

La fonction $\beta_{t|n}$ peut être calculée récursivement dans le sens rétrograde, *i.e.* pour tout $k \in \{0, \dots, n-1\}$

$$\beta_{t|n}(x_t) = \int q_t(x_t, x_{t+1}) g_{t+1}(x_{t+1}, Y_{t+1}) \beta_{t+1|n}(x_{t+1}) dx_{t+1} , \quad (7.103)$$

où, par convention, nous avons posé $\beta_{n|n} \equiv 1$. En utilisant cette relation dans l'expression précédente de la loi de lissage, nous obtenons une équation récursive elle aussi rétrograde de mise à jour de la loi de lissage

$$\begin{aligned} \phi_{\nu,t|n}(f) &\stackrel{\text{def}}{=} \frac{1}{L_{\nu,n}} \int \alpha_{\nu,t}(x_t) f(x_t) \beta_{t|n}(x_t) dx_t \\ &= \frac{1}{L_{\nu,n}} \int \frac{\alpha_{\nu,t}(x_t)}{\alpha_{\nu,t+1}(x_{t+1})} f(x_t) q_t(x_t, x_{t+1}) g_{t+1}(x_{t+1}, Y_{t+1}) \beta_{t+1|n}(x_{t+1}) dx_{t:t+1} \end{aligned} \quad (7.104)$$

$$= \iint B_{\nu,t}(x_{t+1}, x_t) \phi_{\nu,t+1|n}(x_{t+1}) f(x_t) dx_{t:t+1} , \quad (7.105)$$

où $B_{\nu,t}$ est le noyau rétrograde défini par

$$B_{\nu,t}(x_{t+1}, x_t) = \frac{\alpha_{\nu,t}(x_t) q_t(x_t, x_{t+1}) g_{t+1}(x_{t+1}, Y_{t+1})}{\alpha_{\nu,t+1}(x_{t+1})} \quad (7.106)$$

$$= \frac{\phi_{\nu,t|n}(x_t) q_t(x_t, x_{t+1}) g_{t+1}(x_{t+1}, Y_{t+1})}{\int \phi_{\nu,t|n}(x_t) q_t(x_t, x_{t+1}) g_{t+1}(x_{t+1}, Y_{t+1}) q_t(x_t, x_{t+1}) g_{t+1}(x_{t+1}, Y_{t+1}) dx_t} . \quad (7.107)$$

En utilisant la formule élémentaire des espérances conditionnelles, nous avons pour toute fonction f non-négative et tout entier $k \in \{0, \dots, n-1\}$,

$$\begin{aligned} \mathbb{E}_{\nu} [f(X_t) | Y_{0:n}] &= \mathbb{E} [\mathbb{E} [f(X_t) | X_{t+1}, Y_{0:n}] | Y_{0:n}] \\ &= \int \phi_{t+1|n}(dx_{t+1}) \mathbb{E} [f(X_t) | X_{t+1} = x_{t+1}, Y_{0:n}] . \end{aligned} \quad (7.108)$$

Par conséquent, en identifiant (7.108) et (7.105), nous obtenons

$$\mathbb{E}_{\nu} [f(X_t) | X_{t+1} = x_{t+1}, Y_{0:n}] = \int f(x_t) B_{\nu,t}(x_{t+1}, x_t) dx_t \quad (7.109)$$

qui est donc un noyau *rétrograde*.

L'algorithme de lissage permet de calculer ces quantités récursivement, en procédant dans le sens inverse du temps (ou sens rétrograde). La loi de lissage pour $k = n$ coïncide avec la loi de filtrage. L'idée est de déduire, pour $k = 0, \dots, n-1$ récursivement la moyenne $\hat{X}_{t|n}$ et la covariance $\Sigma_{t|n}$ de la loi de lissage à partir de $\hat{X}_{t+1|n}$ et $\Sigma_{t+1|n}$.

$B_{\nu,t}(x_{t+1}, \cdot)$ est égal à la loi de X_t conditionnelle à $X_{t+1} = x_{t+1}$ dans le modèle

$$X_{t+1} = A_t X_t + R_t U_t,$$

où $X_t \sim N(\hat{X}_{t|t}, \Sigma_{t|t})$ et $U_t \sim N(0, I)$ et est indépendant de X_t . En utilisant la Proposition 7.4.1, $B_{\nu,t}(x_{t+1}, \cdot)$ est une loi gaussienne de moyenne et de covariance données respectivement par

$$\hat{X}_{t|t} + \Sigma_{t|t} A_t^t (A_t \Sigma_{t|t} A_t^t + R_t R_t^t)^{-1} (x_{t+1} - A_t \hat{X}_{t|t}), \quad (7.110)$$

et

$$\Sigma_{t|t} - \Sigma_{t|t} A_t^t (A_t \Sigma_{t|t} A_t^t + R_t R_t^t)^{-1} A_t \Sigma_{t|t}. \quad (7.111)$$

La formule de mise à jour rétrograde des moyennes et des covariances de la loi de lissage est donc donnée par :

$$\hat{X}_{t|n} = \hat{X}_{t|t} + \Sigma_{t|t} A_t^t M_t (\hat{X}_{t+1|n} - A_t \hat{X}_{t|t}), \quad (7.112)$$

$$\Sigma_{t|n} = \Sigma_{t|t} - \Sigma_{t|t} A_t^t M_t A_t \Sigma_{t|t} + \Sigma_{t|t} A_t^t M_t \Sigma_{t+1|n} M_t A_t \Sigma_{t|t}, \quad (7.113)$$

où

$$M_t = (A_t \Sigma_{t|t} A_t^t + R_t R_t^t)^{-1}.$$

Les récursions données ci-dessus correspondent à l'algorithme de lissage de Rauch, Tung et Striabel [3].

7.5 La méthode à deux filtres

7.5.1 Paramétrisation par la matrice d'information

En supposant que Σ_V et $\text{Cov}(Y) = B^t \Sigma_X B + \Sigma_V$ sont des matrices de covariance de rang complet, la densité de X conditionnelle à Y , que nous notons dans ce paragraphe $p(x|y)$ est, en appliquant la règle de Bayes, proportionnelle à la densité a priori $p(x)$ de X et la densité conditionnelle $p(y|x)$ de l'observation Y conditionnelle à l'état X ,

$$p(x|y) \propto \exp \left\{ -\frac{1}{2} [(y - Bx)^t \Sigma_V^{-1} (y - Bx) + (x - \mu_X)^t \Sigma_X^{-1} (x - \mu_X)] \right\}, \quad (7.114)$$

où le symbole \propto indique que les deux quantités apparaissant à droite et à gauche de l'expression précédente diffèrent d'une constante multiplicative qui ne dépend pas de x . Notons que dans l'équation précédente, cette constante est facile à calculer car $p(x|y)$ est la densité d'une loi Gaussienne multidimensionnelle densité de probabilité. Nous pouvons calculer la densité conditionnelle $p(x|y)$, en écrivant le terme apparaissant dans l'exponentielle dans (7.114) comme une forme quadratique de la variable x :

$$p(x|y) \propto \exp \left\{ -\frac{1}{2} \left[x^t (B^t \Sigma_V^{-1} B + \Sigma_X^{-1}) x - x^t (B^t \Sigma_V^{-1} y + \Sigma_X^{-1} \mu_X) - (B^t \Sigma_V^{-1} y + \Sigma_X^{-1} \mu_X)^t x \right] \right\}, \quad (7.115)$$

ou de façon équivalente

$$p(x|y) \propto \exp \left\{ -\frac{1}{2} [(x - \mu_{X|Y})^t \Sigma_{X|Y}^{-1} (x - \mu_{X|Y})] \right\}, \quad (7.116)$$

avec

$$\mu_{X|Y} = \Sigma_{X|Y} (B^t \Sigma_V^{-1} y + \Sigma_X^{-1} \mu_X), \quad (7.117)$$

$$\Sigma_{X|Y} = (B^t \Sigma_V^{-1} B + \Sigma_X^{-1})^{-1}. \quad (7.118)$$

Les équations (7.118) et (7.117) sont équivalentes aux équations (7.94) et (7.93). Le fait que (7.94) et (7.118) coïncide est une conséquence directe du *lemme d'inversion matriciel*.

Il est intéressant d'utiliser une paramétrisation alternative utilisant l'inverse de la matrice de covariance $\Pi = \Sigma^{-1}$ et le vecteur $\kappa = \Pi \mu$. La matrice Π est appelée *matrice d'information* et (κ, Π) définissent la *paramétrisation information* de la densité.

7.5.2 Le modèle linéaire gaussien

Proposition 7.5.1. *Considérons le modèle*

$$Y = BX + V, \quad (7.119)$$

où B est une matrice déterministe et X et V sont des vecteurs Gaussiens indépendants vérifiant $\kappa_X = \text{Cov}(X)^{-1} \mathbb{E}[X]$, $\Pi_X = \text{Cov}(X)^{-1}$, $\Pi_V = \text{Cov}(V)^{-1}$ et $\kappa_V = \mathbb{E}[V] = 0$, Then

$$\kappa_{X|Y} = \kappa_X + B^t \Pi_V Y, \quad (7.120)$$

$$\Pi_{X|Y} = \Pi_X + B^t \Pi_V B, \quad (7.121)$$

où $\kappa_{X|Y} = \text{Cov}(X|Y)^{-1} \mathbb{E}[X|Y]$ and $\Pi_{X|Y} = \text{Cov}(X|Y)^{-1}$.

Les équations (7.120) et (7.121) sont des réécritures de (7.117) et (7.118), respectivement.

Si les matrices Π_X , Π_V , ou $\Pi_{X|Y}$ ne sont pas des matrices de rang complet, (7.120) et (7.121) restent encore valides en utilisant le concept de distribution a priori impropre. Considérons l'équation Eq. (7.114) et supposons que la fonction $p(x)$, par exemple, est constante. Dans ce cas, (7.115) a pour expression

$$p(x|y) \propto \exp \left\{ -\frac{1}{2} [x^t (B^t \Sigma_V^{-1} B) x - x^t (B^t \Sigma_V^{-1} y) - (B^t \Sigma_V^{-1} y)^t x] \right\}, \quad (7.122)$$

qui est (en tant que fonction de x) une densité gaussienne lorsque $B^t \Sigma_V^{-1} B$ est de rang complet. Bien entendu, il faut interpréter cette formule avec une certaine précaution car il n'existe pas de densité de probabilité $p(x)$ qui soit constante sur \mathbf{X} . Il s'agit, dans la terminologie des statistiques bayésiennes, d'une loi *impropre*. L'utilisation de telle loi est commune dans le cadre de la théorie de l'inférence bayésienne.

L'interprétation de (7.122) est que, en prenant une a priori impropre constant pour X , la moyenne de X conditionnelle à Y est donnée par

$$(B^t \Sigma_V^{-1} B)^{-1} B^t \Sigma_V^{-1} Y. \quad (7.123)$$

Le message important est que (7.123) découle de (7.120) en supposant que Π_X est la matrice nulle et que κ_X est le vecteur nul. Donc la Proposition 7.5.1 couvre aussi le cas où la loi a priori sur X est impropre, ce qui correspond à supposer que Π_X et κ_X sont égaux à 0. L'exemple suivant illustre une situation un peu plus complexe :

Exemple 7.5.2. Considérons le modèle (7.119) où X est un vecteur aléatoire de dimension 2, Y est une variable scalaire et

$$B = \begin{pmatrix} 1 & 0 \end{pmatrix} \quad \text{et} \quad \text{Cov}(V) = \sigma^2.$$

En utilisation la proposition 7.5.1 les paramètres de la loi a posteriori sont donnés par

$$\kappa_{X|Y} = \kappa_X + \begin{pmatrix} \sigma^{-2} Y \\ 0 \end{pmatrix}, \quad (7.124)$$

$$\Pi_{X|Y} = \Pi_X + \begin{pmatrix} \sigma^{-2} & 0 \\ 0 & 0 \end{pmatrix}. \quad (7.125)$$

En particulier, si la loi a priori sur X est impropre constante, alors (7.124) et (7.125) montre que la distribution a posteriori de la première composante de X conditionnelle à Y est gaussienne de moyenne Y et de variance σ^2 ,

mais que la loi a posteriori de la seconde composante est elle aussi impropre et constante.

L'exemple précédent illustre le fait important que les formules données dans la Proposition 7.5.1 restent valables même lorsque κ_X et Π_X est associé à une loi a priori impropre. La loi de X conditionnelle à X peut elle-même être impropre, mais les paramètres "information" de cette loi impropre sont donnés par (7.120) et (7.121).

Nous utiliserons aussi le résultat suivant

Lemme 7.5.3.

$$\begin{aligned} & \int \exp \left\{ -\frac{1}{2} [(y - Bx)^t \Sigma^{-1} (y - Bx)] \right\} \exp \left\{ -\frac{1}{2} [(y^t \Pi y - 2y^t \kappa)] \right\} dy \\ & \propto \exp \left\{ -\frac{1}{2} [x^t B^t (I + \Pi \Sigma)^{-1} \Pi Bx - 2x^t B^t (I + \Pi \Sigma)^{-1} \kappa] \right\}, \quad (7.126) \end{aligned}$$

Preuve. Notons $p(x)$ le terme apparaissant dans le terme de gauche de l'expression (7.126). Nous avons

$$p(x) = \exp \left\{ -\frac{1}{2} x^t B^t \Sigma^{-1} Bx \right\} \int \exp -\frac{1}{2} [y^t (\Pi + \Sigma^{-1}) y - 2y^t (\kappa + \Sigma^{-1} Bx)] dy. \quad (7.127)$$

En complétant le carré, nous avons

$$\begin{aligned} & \{y - (\Pi + \Sigma^{-1})^{-1} (\kappa + \Sigma^{-1} Bx)\}^t (\Pi + \Sigma^{-1}) \{y - (\Pi + \Sigma^{-1})^{-1} (\kappa + \Sigma^{-1} Bx)\} \\ & - (\kappa + \Sigma^{-1} Bx)^t (\Pi + \Sigma^{-1})^{-1} (\kappa + \Sigma^{-1} Bx). \quad (7.128) \end{aligned}$$

Par conséquent

$$p(x) \propto \exp -\frac{1}{2} \left\{ [-2x^t B^t \Sigma^{-1} (\Pi + \Sigma^{-1})^{-1} \kappa + x^t B^t (\Sigma^{-1} - \Sigma^{-1} (\Pi + \Sigma^{-1})^{-1} \Sigma^{-1}) Bx] \right\}, \quad (7.129)$$

où nous n'avons fait figurer que les termes qui dépendent de x . L'équation (7.126) découle des identités matricielles $\Sigma^{-1} (\Pi + \Sigma^{-1})^{-1} = (I + \Pi \Sigma)^{-1}$ et

$$\begin{aligned} \Sigma^{-1} - \Sigma^{-1} (\Pi + \Sigma^{-1})^{-1} \Sigma^{-1} &= \Sigma^{-1} (\Pi + \Sigma^{-1})^{-1} [(\Pi + \Sigma^{-1}) - \Sigma^{-1}] \\ &= (I + \Pi \Sigma)^{-1} \Pi. \end{aligned}$$

□

7.5.3 Récursion rétrograde

Nous allons maintenant réécrire la récursion rétrograde en utilisant les paramètres d'information. Les fonctions rétrogrades définies par (7.102) ne sont pas des distributions de probabilités. En effet, $\beta_{t|n}(x)$ est la densité des observations futures Y_{t+1}, \dots, Y_n conditionnelles à $X_t = x$. Dans les modèles linéaires gaussiens, la Proposition 7.5.1 montre que $\beta_{t|n}(x)$ est de la forme

$$p(y|x) \propto \exp -\frac{1}{2} [(y - Mx)^t \Sigma^{-1} (y - Mx)] ,$$

où les matrices M et Σ sont données par (7.93) et (7.94). En procédant comme au paragraphe précédent, cette densité peut être mise sous la forme (7.122), en remplaçant B et Σ_V par M et Σ , respectivement.

En utilisant cette réécriture, nous pouvons réinterpréter $\beta_{t|n}(x)$ comme la distribution de X_t conditionnelle à Y_{t+1}, \dots, Y_n dans un *pseudo-modèle* où l'état X_t est supposé être distribué suivant une loi a priori impropre constante. La fonction $\beta_{t|n}(x)$ n'est une distribution de probabilité *propre* que si $M^t \Sigma^{-1} M$ est de rang complet. En particulier, rappelons que la récursion rétrograde Eq. (7.103) est initialisée en posant $\beta_{n|n}(x) = 1$: par conséquent, $\beta_{n|n}$ n'est jamais une distribution gaussienne propre.

Nous notons par $\kappa_{t|n}$ et $\Pi_{t|n}$ les paramètres d'information (produit de la matrice de précision et de la moyenne et matrice de précision) correspondant à la distribution (propre ou impropre) $\beta_{t|n}$ pour $k = n$. Par définition, $\kappa_{n|n} = 0$ et $\Pi_{n|n} = 0$. Remarquons que $\kappa_{t|n}$ et $\Pi_{t|n}$ permettent de spécifier $\beta_{t|n}$ à une *constante multiplicative inconnue*.

Proposition 7.5.4 (Récursion rétrograde pour les paramètres d'information). *Considérons le modèle linéaire gaussien donné par (7.1)–(7.2) et supposons que $S_k S_k^t$ est de rang complet pour tout $k \geq 0$. Les paramètres d'information $\kappa_{t|n}$ and $\Pi_{t|n}$ qui déterminent la fonction rétrograde $\beta_{t|n}$ (à une constante multiplicative près), sont donnés récursivement par :*

Initialisation : Posons $\kappa_{n|n} = 0$ et $\Pi_{n|n} = 0$.

Récursion rétrograde : Pour $k = n - 1$ à 0 ,

$$\tilde{\kappa}_{t+1|n} = B_{t+1}^t (S_{t+1} S_{t+1}^t)^{-1} Y_{t+1} + \kappa_{t+1|n} , \quad (7.130)$$

$$\tilde{\Pi}_{t+1|n} = B_{t+1}^t (S_{t+1} S_{t+1}^t)^{-1} B_{t+1} + \Pi_{t+1|n} , \quad (7.131)$$

$$\kappa_{t|n} = A_t^t \left(I + \tilde{\Pi}_{t+1|n} R_t R_t^t \right)^{-1} \tilde{\kappa}_{t+1|n} , \quad (7.132)$$

$$\Pi_{t|n} = A_t^t \left(I + \tilde{\Pi}_{t+1|n} R_t R_t^t \right)^{-1} \tilde{\Pi}_{t+1|n} A_t . \quad (7.133)$$

Preuve. Considérons tout d'abord le calcul de la fonction

$$\tilde{\beta}_{t+1|n}(x) \propto g_{t+1}(x) \beta_{t+1|n}(x) \quad (7.134)$$

à partir de $\beta_{t+1|n}$. Cette fonction peut être interprétée comme la distribution a posteriori de X dans un pseudo-modèle dans lequel X serait distribué suivant la densité (potentiellement impropre) $\beta_{t+1|n}$ (spécifiée par les paramètres d'information $\kappa_{t+1|n}$ et $\Pi_{t+1|n}$) et l'observation Y serait donnée par

$$Y = B_{t+1}X + S_{t+1}V,$$

où V est indépendant de X . Les équations (7.130)–(7.131) correspondent aux paramètres d'information de la loi (potentiellement impropre) $\tilde{\beta}_{t+1|n}$ par application de la Proposition 7.5.1.

La fonction rétrograde (7.103) est définie par

$$\beta_{t|n}(x) \propto \int Q_t(x, dx') \tilde{\beta}_{t+1|n}(x'). \quad (7.135)$$

Comme Q_t est un noyau de transition gaussien (correspondant à l'équation d'état (7.1)), (7.135) se déduit directement du Lemme 7.5.3 qui donne (7.132) et (7.133). \square

Il est possible de calculer récursivement la loi de lissage dans la récursion permettant de calculer la fonction rétrograde

Algorithme 7.5.5 (Lissage avant-arrière) *Récursion directe : Mettre en oeuvre l'algorithme du filtrage de Kalman (Algorithme 7.3.6) et mémoriser la moyenne $\hat{X}_{t|t}$ et la covariance $\Sigma_{t|t}$ de filtrage.*

Récursion rétrograde : Mettre en oeuvre la récursion rétrograde puis calculer pour tout k

$$\hat{X}_{t|n} = \hat{X}_{t|t} + \Sigma_{t|t} (I + \Pi_{t|n} \Sigma_{t|t})^{-1} (\kappa_{t|n} - \Pi_{t|n} \hat{X}_{t|t}), \quad (7.136)$$

$$\Sigma_{t|n} = \Sigma_{t|t} - \Sigma_{t|t} (I + \Pi_{t|n} \Sigma_{t|t})^{-1} \Pi_{t|n} \Sigma_{t|t}. \quad (7.137)$$

Preuve. Les équations (7.136) and (7.137) s'obtiennent exactement comme dans le Lemme 7.5.3, en remplaçant $(y - Bx)^t \Sigma^{-1} (y - Bx)$ by $(x - \mu)^t \Sigma^{-1} (x - \mu)$ et en appliquant le résultat avec $\mu = \hat{X}_{t|t}$, $\Sigma = \Sigma_{t|t}$, $\kappa = \kappa_{t|n}$ et $\Pi = \Pi_{t|n}$. Si la matrice $\Pi_{t|n}$ est inversible, (7.136) et (7.137) découlent de la Proposition 7.4.1. \square

Partie III

Estimation statistique

Chapitre 8
Introduction aux statistiques
asymptotiques

8.1 Notions de convergence

Soit $\{X_n\}_{n \geq 0}$ une suite de vecteurs aléatoires à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ et définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Nous notons, pour $x, y \in \mathbb{R}^d$, $d(x, y) = \|x - y\|$ la distance euclidienne.

Nous donnons les trois notions de convergence pour une suite de v.a. $\{X_n\}_{n \geq 0}$ dans l'ordre *croissant*, c'est-à-dire de la notion la plus faible à la plus forte (l'implication d'une convergence à une autre est établie au théorème 8.5.1).

Définition 8.1.1. Nous dirons que la suite $\{X_n\}_{n \geq 0}$ *converge en loi* vers X et nous noterons

$$X_n \xrightarrow{\mathcal{L}} X ,$$

si pour toute fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$ continue bornée,

$$\lim_n \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)] .$$

Il serait plus approprié de définir la convergence en loi uniquement à partir des lois :

Définition 8.1.2. Nous dirons que la suite de probabilités $\{\mu_n\}$ définies sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *converge étroitement* vers la probabilité μ définie sur $(X, \mathcal{B}(X))$ et nous noterons

$$\mu_n \rightsquigarrow \mu ,$$

si pour toute fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$ continue bornée,

$$\lim_{n \rightarrow \infty} \mu_n(f) = \mu(f) ,$$

où pour une fonction f intégrable et une mesure ν , $\nu(f) = \int f(x)\nu(dx)$.

Il est donc clair que $X_n \xrightarrow{\mathcal{L}} X$ est équivalent à $P_{X_n} \rightsquigarrow P_X$, où l'on a noté P_{X_n} et P_X les probabilités images de X_n et X . On voit à cette occasion qu'il n'est pas nécessaire de définir les v.a. $\{X_n\}_{n \geq 0}$ et X sur le même espace de probabilité pour définir la convergence en loi.

Il existe divers caractérisations de cette convergence très utiles en pratique qui seront abordées dans la section 8.5 (voir le Lemme 8.3.1 et le théorème 8.3.3).

Définition 8.1.3. Nous dirons que la suite $\{X_n\}_{n \geq 0}$ *converge en probabilité* vers X et nous noterons

$$X_n \xrightarrow{\mathbb{P}} X ,$$

si pour tout $\epsilon > 0$, nous avons :

$$\lim_{n \rightarrow \infty} \mathbb{P}(d(X_n, X) > \epsilon) = 0 .$$

Au contraire de la convergence en loi, cette définition utilise que X_n et X sont définies sur le même espace de probabilité ; autrement, on ne serait calculer la loi de $d(X_n, X)$. Nous verrons les nombreuses relations qui existent entre convergence en loi et convergence en probabilité dans la section 8.5. Notons cependant qu'il est un cas où ces définitions coïncident. Si X est déterministe, i.e. s'il existe $c \in \mathbb{R}^d$ tel que $\mathbb{P}(X = c) = 1$, alors il est facile de voir que $X_n \xrightarrow{\mathbb{P}} X$ (on notera plutôt $X_n \xrightarrow{\mathbb{P}} c$ dans ce cas très particulier) est équivalent à $X_n \xrightarrow{\mathcal{L}} X$ (voir théorème 8.5.1-(b)). Il n'y a pas de contradiction avec la remarque précédente, les loi v.a. déterministes sont définies sans ambiguïté indépendamment de l'espace de probabilité sur lesquelles elles sont définies ; il s'en suit que pour X constante la loi de $d(X_n, X)$ est connue uniquement à partir de la loi de X_n .

Définition 8.1.4. Nous dirons que la suite $\{X_n\}_{n \geq 0}$ *converge presque-sûrement* vers X et nous noterons

$$X_n \xrightarrow{\mathbb{P}\text{-p.s.}} X,$$

si :

$$\mathbb{P}(\limsup_n d(X_n, X) = 0) = 1.$$

Cette fois-ci, il est nécessaire de définir $\{X_n\}_{n \geq 0}$ et X sur le même espace de probabilité sinon la loi de $\limsup_{n \rightarrow \infty} d(X_n, X)$ n'a pas de sens.

Une méthode essentielle permettant de prouver une telle convergence est le Lemme de Borel-Cantelli :

Lemme 8.1.5 (Lemme de Borel-Cantelli). *Soit $\{A_n, n \in \mathbb{N}\}$ une suite d'événements tels que $\sum_{n \geq 0} \mathbb{P}(A_n) < \infty$. Alors*

$$\mathbb{P}\left(\sum_{n \geq 0} \mathbb{1}(A_n) < \infty\right) = 1. \quad (8.1)$$

Preuve. $\left(\bigcup_{n \geq m} A_n\right)_m$ est une suite décroissante d'événements dont l'intersection coïncide avec l'événement $\left\{\sum_{n \geq 0} \mathbb{1}(A_n) = \infty\right\}$. Or $\mathbb{P}\left(\bigcup_{n \geq m} A_n\right) \leq \sum_{n \geq m} \mathbb{P}(A_n)$ qui tend vers zéro quand $m \rightarrow \infty$ par hypothèse, ce qui implique (8.1). \square

8.2 Suites tendues

Introduisons la notion de *tension* pour une variable aléatoire.

Définition 8.2.1. Une v.a. X à valeur dans un espace topologique quelconque est dite *tendue* si pour tout $\epsilon > 0$, il existe un compact K tel que

$\mathbb{P}(X \notin K) \leq \epsilon$. Une famille $\{X_\alpha, \alpha \in A\}$ est dite *uniformément tendue* (ou *bornée en probabilité*) si, pour tout $\epsilon > 0$, il existe un compact K tel que $\sup_{\alpha \in A} \mathbb{P}(X_\alpha \notin K) \leq \epsilon$.

Toute v.a. à valeurs dans \mathbb{R}^d équipé de la topologie associée à la distance euclidienne est *tendue* puisque $\bigcap_{n \geq 1} \{\|x\| \geq n\} = \emptyset$. De façon générale toute famille finie de variables aléatoires réelles est tendue. Il s'en suit qu'une suite $(X_n)_{n \in \mathbb{N}}$ de v.a. à valeurs dans \mathbb{R}^d est bornée en probabilité si, pour tout $\epsilon > 0$, il existe $M > 0$ tel que

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\|X_n\| > M) \leq \epsilon.$$

Un critère simple de tension uniforme pour une famille $\{X_\alpha, \alpha \in A\}$ de v.a. à valeurs dans \mathbb{R}^d est obtenu comme suit. Si, pour $p > 0$, nous avons $\sup_\alpha \mathbb{E}[\|X_\alpha\|^p] < \infty$, l'inégalité de Markov montre que :

$$\mathbb{P}(\|X_\alpha\| \geq M) \leq M^{-1} \mathbb{E}[\|X_\alpha\|^p],$$

et donc que $\{X_\alpha\}$ est uniformément tendue. Plus généralement la notion de tension est intimement liée à la notion de convergence en loi. Le résultat suivant montre en effet que toute suite convergeant en loi est tendue et qu'il existe une réciproque (partielle) à ce résultat.

Théorème 8.2.2 (Prohorov dans \mathbb{R}^d). *Soit $\{X_n\}$ une suite de v.a. de \mathbb{R}^d .*

- a) *Si $X_n \xrightarrow{\mathcal{L}} X$, alors la famille $\{X_n, n \in \mathbb{N}\}$ est uniformément tendue,*
- b) *Si la famille $\{X_n, n \in \mathbb{N}\}$ est uniformément tendue, alors il existe une sous suite telle que $X_{n_j} \xrightarrow{\mathcal{L}} X$ pour une v.a X .*

Preuve. (a)) : Si $X_n \xrightarrow{\mathcal{L}} X$, on, pour tout M , $\limsup_n \mathbb{P}(\|X_n\| \geq M) \leq \limsup_n \mathbb{E}[\phi_M(\|X_n\|)] = \mathbb{E}[\phi_M(\|X\|)] \leq \mathbb{P}(\|X\| \geq M - 1)$, où ϕ_M est une fonction continue bornée sur \mathbb{R} telle que $\mathbb{1}_{[M, \infty)} \leq \phi_M \leq \mathbb{1}_{[M-1, \infty)}$. Donc pour tout $\epsilon > 0$, on peut trouver $M > 0$ et $N \in \mathbb{N}$ tels que $\mathbb{P}(\|X_n\| \geq M) \leq \epsilon$ pour tout $n \geq N$. En augmentant M suffisamment, on obtient que $\mathbb{P}(\|X_n\| \geq M) \leq \epsilon$ pour tout $n \in \{0, \dots, N\}$.

(b)) : Voir Theorem 18.6 dans [1]. □

Corollaire 8.2.3. *Soit $\{X_n\}$ une suite de v.a. de \mathbb{R}^d uniformément tendue et X une v.a. de \mathbb{R}^d vérifiant : pour toute suite croissante divergente d'entiers (k_n) , si $X_{k_n} \xrightarrow{\mathcal{L}} Y$, alors $Y \sim X$. Alors $X_n \xrightarrow{\mathcal{L}} X$.*

Preuve. D'après le théorème de Prohorov, de toute sous-suite de X_n , on peut extraire une sous-suite qui converge en loi vers X . Un raisonnement par contradiction montre que $X_n \xrightarrow{\mathcal{L}} X$: en effet, si tel n'était pas le cas, il existerait une fonction continue bornée $f : \mathbb{R}^d \rightarrow \mathbb{R}$ telle que $\mathbb{E}[f(X_n)]$ ne converge pas vers $\mathbb{E}[f(X)]$. On aurait donc une sous-suite $\{X_{n_j}\}$ et $\epsilon > 0$ tels que $|\mathbb{E}[f(X_{n_j})] - \mathbb{E}[f(X)]| > \epsilon$ pour tout $n \in \mathbb{N}$. Par suite, aucune sous-suite de $\{X_{n_j}\}$ ne convergerait vers X , d'où la contradiction. □

8.3 Caractérisations de la convergence en loi

Le lemme de Portmanteau donne des conditions équivalentes à la convergence en loi.

Lemme 8.3.1 (Portmanteau). *Soit $\{X_n\}_{n \geq 0}$ et X une suite de vecteurs aléatoires à valeurs dans \mathbb{R}^d et soit \mathbf{P}_n et \mathbf{P} leurs loi respectives. Les assertions suivantes sont équivalentes.*

- a) $X_n \xrightarrow{\mathcal{L}} X$,
- b) Pour toute fonction f bornée et lipschitzienne ($|f(x) - f(y)| \leq K_f \|x - y\|$ pour tout $x, y \in \mathbb{R}^d$), $\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$,
- c) $\liminf \mathbf{P}_n(G) \geq \mathbf{P}(G)$ pour tout ensemble ouvert,
- d) $\limsup \mathbf{P}_n(F) \leq \mathbf{P}(F)$ pour tout ensemble fermé,
- e) Pour tout ensemble borélien A tel que $\mathbf{P}(A) = \mathbf{P}(\bar{A}) = \mathbf{P}(\text{int} A)$,

$$\lim_{n \rightarrow \infty} \mathbf{P}_n(A) = \mathbf{P}(A)$$

où \bar{A} est la fermeture de A (A complété par l'ensemble des valeurs d'adhérence des suites d'éléments de A).

Preuve. On montre une succession d'implications qui démontre l'équivalence des propositions :

- 1. (a) \Rightarrow (b) est trivial puisque toute fonction lipschitzienne est continue.
- 2. (b) \Rightarrow (c) : soit G un ouvert. Pour tout m , définissons $f_m(x) = \max(1, m d(x, G^c))$. La suite $\{f_m\}$ est une suite croissante de fonctions positives bornées lipschitziennes $\{f_m\}$ qui convergent simplement vers $\mathbb{1}_G$. Par construction, $\liminf \mathbb{E}[\mathbb{1}_G(X_n)] \geq \liminf \mathbb{E}[f_m(X_n)] = \mathbb{E}[f_m(X)]$ par (b)). Le théorème de Beppo-Levi montre d'autre part $\mathbb{E}[f_m(X)]$ converge vers en croissant vers $\mathbb{E}[\mathbb{1}_G(x)]$, ce qui établit le résultat.
- 3. (c) \Leftrightarrow (d) est évident en prenant les complémentaires
- 4. (c) \Leftrightarrow (d) \Rightarrow (e) : pour tout $A \in \mathcal{B}(\mathbb{R}^d)$ vérifiant $\mathbf{P}(A) = \mathbf{P}(\bar{A}) = \mathbf{P}(\text{int} A)$, nous avons

$$\begin{aligned} \mathbf{P}(A) = \mathbf{P}(\text{int} A) &\leq \liminf \mathbf{P}_n(\text{int} A) \leq \liminf \mathbf{P}_n(A) \leq \\ &\limsup \mathbf{P}_n(A) \leq \limsup \mathbf{P}_n(\bar{A}) \leq \mathbf{P}(\bar{A}) = \mathbf{P}(A). \end{aligned}$$

D'où (e) en coinçant $\liminf \mathbf{P}_n(A)$ et $\limsup \mathbf{P}_n(A)$ entre les deux mêmes bornes.

- 5. (e) \Rightarrow (a) : Soit $A_M = [-M, M]^d$. Alors $\phi : M \mapsto \mathbf{P}(A_M)$ est une fonction croissante $\mathbb{R}_+ \rightarrow [0, 1]$, continue à droite, qui a un nombre fini de discontinuités plus grande qu'un $\epsilon > 0$ donné. L'ensemble de ses discontinuités de ϕ sur \mathbb{R}_+ est donc au plus dénombrable. Soit (M_n) une suite croissante divergente de points de continuité de ϕ : pour tout

$\epsilon > 0$ il existe n tel que $\phi(M) = \mathbf{P}([-M, M]^d) > 1 - \epsilon$. Comme, en tout point de continuité M de ϕ , $\mathbf{P}(A_M) = \mathbf{P}(\bar{A}_M) = \mathbf{P}(\text{int} A_M)$, en utilisant (e)), on obtient que $\limsup \mathbf{P}_n(A_{M_n}) > 1 - \epsilon$, puis par suite que (\mathbf{P}_n) est une suite uniformément tendue. Utilisons le corollaire 8.2.3. Soit (k_n) une suite croissante divergente d'entiers telle que $X_{k_n} \xrightarrow{\mathcal{L}} Y$. Alors Pour tout $A \in \mathcal{B}(\mathbb{R}^d)$ vérifiant $\mathbf{P}(A) = \mathbf{P}(\bar{A}) = \mathbf{P}(\text{int} A)$, on a $\mathbb{P}(Y \in A) = \mathbf{P}(A)$. On définit maintenant, pour $\mathbf{x} = (x_1, \dots, x_d)$, le borélien $B_{\mathbf{x}} = (-\infty, x_1] \times \dots \times (-\infty, x_d]$. En raisonnant comme pour les A_M , on montre aisément qu'il existe un ensemble $\mathcal{C} \subseteq \mathbb{R}^d$ dense dans \mathbb{R}^d tel que, pour tout $\mathbf{x} \in \mathcal{C}$, $\mathbf{P}(B_{\mathbf{x}}) = \mathbf{P}(\text{int} B_{\mathbf{x}})$. Comme \mathcal{C} est dense, on a $\sigma(B_{\mathbf{x}}, \mathbf{x} \in \mathcal{C}) = \mathcal{B}(\mathbb{R}^d)$. On note

$$\mathcal{C}' = \{\mathbf{x} : \mathbf{P}(B_{\mathbf{x}}) = \mathbf{P}(\text{int} B_{\mathbf{x}})\}.$$

Comme \mathcal{C}' contient \mathcal{C} , on a $\sigma(B_{\mathbf{x}}, \mathbf{x} \in \mathcal{C}') = \mathcal{B}(\mathbb{R}^d)$. Comme $B_{\mathbf{x}} \cap B_{\mathbf{y}}$ a sa frontière incluse dans l'union des des frontières de $B_{\mathbf{x}}$ et $B_{\mathbf{y}}$, on voit que $\{B_{\mathbf{x}}, \mathbf{x} \in \mathcal{C}'\}$ est un π -système (*i.e.* est stable par intersection finie). Le corollaire 6.1 dans [1] permet donc de conclure que, la loi de Y coïncidant avec \mathbf{P} sur ce π -système, elle coïncide avec sur tous les boréliens. \square

La proposition (e)) implique que $\lim_n F_n(\mathbf{x}) = F(\mathbf{x})$ en tout point de continuité de F , où l'on a noté $F_n(\mathbf{x}) = \mathbb{P}(X_n \leq \mathbf{x})$ et $F(\mathbf{x}) = \mathbb{P}(X \leq \mathbf{x})$ les fonctions de répartition de X_n et X . D'après la preuve de (e)) \Rightarrow (a)), on voit que la réciproque est en fait vraie : si $\lim_n F_n(\mathbf{x}) = F(\mathbf{x})$ en tout point de continuité de F , alors $X_n \xrightarrow{\mathcal{L}} X$. Une conséquence immédiate est que, si la fonction de répartition de X est continue, alors $X_n \xrightarrow{\mathcal{L}} X$ est équivalent à $\mathbb{P}(X_n \leq \mathbf{x}) \rightarrow \mathbb{P}(X \leq \mathbf{x})$ pour tout $\mathbf{x} \in \mathbb{R}^d$. Pour $d = 1$, la convergence est en fait uniforme en \mathbf{x} comme l'indique le résultat suivant.

Lemme 8.3.2. *Supposons, pour $d = 1$, que $X_n \xrightarrow{\mathcal{L}} X$ et que la fonction de répartition de X est continue. Alors,*

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |\mathbb{P}(X_n \leq x) - \mathbb{P}(X \leq x)| = 0.$$

Preuve. Soit $F_n(x) = \mathbb{P}(X_n \leq x)$ et $F(x) = \mathbb{P}(X \leq x)$. F étant continue, il existe des points $-\infty^d = x_0 < x_1 < \dots < x_k = \infty^d$ tels que $F(x_i) = i/k$. F_n et F étant croissantes, nous avons, pour tout $x \in \mathbb{R}^d$, en choisissant i tel que $x_{i-1} \leq x \leq x_i$:

$$\begin{aligned} F_n(x) - F(x) &\leq F_n(x_i) - F(x_{i-1}) = F_n(x_i) - F(x_i) + 1/k \\ F_n(x) - F(x) &\geq F_n(x_{i-1}) - F(x_i) = F_n(x_{i-1}) - F(x_{i-1}) - 1/k. \end{aligned}$$

Donc $|F_n(x) - F(x)|$ est borné par $\sup_i |F_n(x_i) - F(x_i)| + 1/k$ pour tout x . Par conséquent,

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq \lim_{n \rightarrow \infty} \sup_{i \in \{0, \dots, k\}} |F_n(x_i) - F(x_i)| + 1/k = 1/k,$$

ce qui permet de conclure, en choisissant k arbitrairement grand. \square

Le raisonnement s'étend sans difficulté au cas multidimensionnel mais nous omettons cet énoncé, la fonction de répartition étant mieux adaptée à la dimension un.

Comme nous l'avons déjà vu dans le lemme de Portmanteau, pour démontrer la convergence en loi, il suffit de s'intéresser à un sous-ensemble des fonctions continues bornées, par exemple, les fonctions lipschitziennes bornées, mais cette classe peut encore être réduite. Nous allons en fait démontrer dans cette partie qu'il suffit de s'intéresser à une seule fonction, la *fonction caractéristique*,

$$t \rightarrow \mathbb{E} \left[e^{it^T X} \right], \quad t \in \mathbb{R}^d.$$

Pour tout t , la fonction $x \rightarrow e^{it^T X}$ est continue et bornée. Par conséquent, si $X_n \xrightarrow{\mathcal{L}} X$, $\mathbb{E} \left[e^{it^T X_n} \right] \rightarrow \mathbb{E} \left[e^{it^T X} \right]$. Le théorème de continuité de Levy montre que la réciproque est vraie.

Théorème 8.3.3 (Lévy). *Soit X_n et X des v.a. à valeurs dans \mathbb{R}^d . Alors $X_n \xrightarrow{\mathcal{L}} X$ si et seulement si $\mathbb{E} \left[e^{it^T X_n} \right] \rightarrow \mathbb{E} \left[e^{it^T X} \right]$. De plus si, pour tout t , $\mathbb{E} \left[e^{it^T X_n} \right]$ converge vers une fonction $\phi(t)$ qui est continue en 0, alors $\phi(t)$ est la fonction caractéristique d'une v.a. X et $X_n \xrightarrow{\mathcal{L}} X$.*

Preuve. L'implication directe est évidente. Considérons la réciproque. Supposons tout d'abord que la famille $\{X_n\}$ est bornée en probabilité et utilisons le corollaire 8.2.3. Soit (k_n) une suite croissante divergente d'entiers telle que $X_{k_n} \xrightarrow{\mathcal{L}} Y$. La fonction caractéristique de $\mathbb{E} \left[e^{it^T Y} \right]$ est alors la limite de $\mathbb{E} \left[e^{it^T X_{k_j}} \right]$ et donc, pour tout t , $\mathbb{E} \left[e^{it^T Y} \right] = \mathbb{E} \left[e^{it^T X} \right]$ (ou $\phi(t)$ dans le cas où l'on n'a pas supposé l'existence de X). Comme la fonction caractéristique caractérise la loi (voir le théorème 14.1 dans [1]), X et Y ont la même loi.

Il suffit maintenant de démontrer que l'hypothèse $\mathbb{E} \left[e^{it^T X_n} \right] \rightarrow \mathbb{E} \left[e^{it^T X} \right]$ implique que $\{X_n, n \in \mathbb{N}\}$ est bornée en probabilité. Pour tout x et tout $\delta > 0$ nous avons :

$$\mathbb{1}(|x| \geq 2/\delta) \leq 2 \left(1 - \frac{\sin(\delta x)}{\delta x} \right) = \frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \cos(tx)) dt.$$

En remplaçant x par X_n , en utilisant le théorème de Fubini :

$$\mathbb{P}(|X_n| \geq 2/\delta) \leq \frac{1}{\delta} \int_{-\delta}^{\delta} \operatorname{Re} \left(1 - \mathbb{E} \left[e^{it^T X_n} \right] \right) dt.$$

Par hypothèse, l'intégrande converge pour tout t vers $\operatorname{Re} \left(1 - \mathbb{E} \left[e^{it^T X} \right] \right)$, quand $n \rightarrow \infty$. Le théorème de convergence dominé implique donc :

$$\limsup \mathbb{P}(|X_n| \geq 2/\delta) \leq \frac{1}{\delta} \int_{-\delta}^{\delta} \operatorname{Re}(1 - \phi(t)) dt.$$

Comme la fonction ϕ est continue en zéro, pour tout $\epsilon > 0$, il existe δ tel que, pour tout $|t| \leq \delta$, $|\phi(t)| \leq \epsilon$. Par suite,

$$\limsup \mathbb{P}(|X_n| \geq 2/\delta) \leq 2\epsilon,$$

et donc la famille $\{X_n, n \in \mathbb{N}\}$ est bornée en probabilité. \square

La fonction caractéristique d'un vecteur aléatoire $X = (X_1, \dots, X_k)$ en $t \in \mathbb{R}^d$ peut être vue comme la fonction caractéristique de la v.a. $Y = t^T X$ évaluée au point 1, $\psi(t) = \mathbb{E} \left[e^{it^T X} \right] = \phi(1)$ où $\phi : u \in \mathbb{R} \rightarrow \mathbb{E} \left[e^{iuY} \right]$. Supposons que $X_n \xrightarrow{\mathcal{L}} X$. Alors, pour tout $t \in \mathbb{R}^d$ et tout $u \in \mathbb{R}$, $\mathbb{E} \left[e^{iut^T X_n} \right] \rightarrow \mathbb{E} \left[e^{iu(t^T X)} \right]$, et donc $t^T X_n \xrightarrow{\mathcal{L}} t^T X$ par application du théorème de Levy 8.3.3. Réciproquement, supposons que pour tout $t \in \mathbb{R}^d$, $t^T X_n \xrightarrow{\mathcal{L}} t^T X$. Alors $\mathbb{E} \left[e^{i(t^T X_n)} \right] \rightarrow \mathbb{E} \left[e^{i(t^T X)} \right]$ et donc $X_n \xrightarrow{\mathcal{L}} X$, encore par application du théorème de Levy. Par conséquent :

Proposition 8.3.4 (Procédé de Cramér-Wold). *Soit $(X_n, n \in \mathbb{N})$ une suite de vecteurs aléatoires à valeurs dans \mathbb{R}^d . Alors :*

$$X_n \xrightarrow{\mathcal{L}} X \quad \Leftrightarrow \quad t^T X_n \xrightarrow{\mathcal{L}} t^T X, \quad \forall t \in \mathbb{R}^d.$$

Cette approche est très utilisée pour prouver la convergence en loi de vecteurs. Elle est connue sous le nom de *procédé de Cramér-Wold*. Elle permet de réduire les problèmes de convergence de vecteurs aléatoires à des problèmes de convergence de variables aléatoires.

8.4 théorème de continuité

Une propriété fondamentale des notions de convergence que l'on a introduite est qu'elles sont conservées par une transformation continue : si la suite de v.a. $\{X_n\}$ converge vers X et que g est continue, alors $g(X_n)$ converge vers $g(X)$ et le résultat est vrai pour les trois types de convergence considérés.

Théorème 8.4.1 (Transformation continue). *Soit $g : \mathbb{R}^d \mapsto \mathbb{R}^m$ est continue en tout point d'un ensemble C tel que $\mathbb{P}(X \in C) = 1$. Alors,*

a) Si $X_n \xrightarrow{\mathcal{L}} X$, alors $g(X_n) \xrightarrow{\mathcal{L}} g(X)$,

- b) Si $X_n \xrightarrow{\mathbb{P}} X$, alors $g(X_n) \xrightarrow{\mathbb{P}} g(X)$,
 c) Si $X_n \xrightarrow{\mathbb{P}\text{-p.s.}} X$, alors $g(X_n) \xrightarrow{\mathbb{P}\text{-p.s.}} g(X)$.

Preuve. (a)) : Par définition, $\{g(X_n) \in F\} = \{X_n \in g^{-1}(F)\}$. Pour tout fermé F , on a :

$$g^{-1}(F) \subset \overline{g^{-1}(F)} \subset (g^{-1}(F) \cup C^c).$$

En effet, seule la seconde inclusion est non-triviale ; soit $x \in \overline{g^{-1}(F)}$ et montrons qu'on bien $x \in C^c$ (le complémentaire de C) ou bien $x \in g^{-1}(F)$. Il existe une suite x_m d'éléments de $g^{-1}(F)$ telle que $x_m \rightarrow x$. Si $x \in C$, $g(x_m) \rightarrow g(x)$, car g est continue au point x , et comme $g(x_m) \in F$ et F est fermé, $g(x) \in F$, ce qui implique que $x \in g^{-1}(F)$. On déduit de cette série d'inclusions et du lemme de Portmanteau, comme $X_n \xrightarrow{\mathcal{L}} X$ et $\mathbb{P}(X \in C^c) = 0$,

$$\begin{aligned} \limsup \mathbb{P}(g(X_n) \in F) &\leq \limsup \mathbb{P}(X_n \in \overline{g^{-1}(F)}) \leq \mathbb{P}(X \in \overline{g^{-1}(F)}) \\ &\leq \mathbb{P}(X \in g^{-1}(F) \cup C^c) = \mathbb{P}(X \in g^{-1}(F)) = \mathbb{P}(g(X) \in F), \end{aligned}$$

et, donc, en appliquant de nouveau le lemme de Portmanteau, $g(X_n) \xrightarrow{\mathcal{L}} g(X)$.

(b)) : Soit $\epsilon > 0$. Pour tout $\delta > 0$, soit B_δ l'ensemble des points x tels qu'il existe y tel que $\|x - y\| \leq \delta$ mais $\|g(x) - g(y)\| \geq \epsilon$. Si $X \notin B_\delta$ et $\|g(X_n) - g(X)\| \geq \epsilon$, alors $\|X_n - X\| \geq \delta$. Nous avons donc :

$$\mathbb{P}(\|g(X_n) - g(X)\| \geq \epsilon) \leq \mathbb{P}(X \in B_\delta) + \mathbb{P}(\|X_n - X\| \geq \delta).$$

Le second terme du membre de droite tend vers 0 car $X_n \xrightarrow{\mathbb{P}} X$. Nous avons $\mathbb{P}(X \in B_\delta \cap C^c) = 0$ et $\lim_{\delta \searrow 0} \mathbb{P}(X \in B_\delta \cap C) = 0$ par continuité de g .

(c)) : dans ce cas, une simple application de la définition donne le résultat. \square

8.5 Symboles o et O stochastiques

Le théorème suivant clarifie les relations entre les différentes définitions de convergence et celles entre convergence d'un vecteur et convergence de ses coordonnées.

Théorème 8.5.1. Soient X_n, X et Y_n des v.a. Nous avons :

- a) $X_n \xrightarrow{\mathbb{P}\text{-p.s.}} X$ implique $X_n \xrightarrow{\mathbb{P}} X$,
 b) $X_n \xrightarrow{\mathbb{P}} X$ implique $X_n \xrightarrow{\mathcal{L}} X$,
 c) $X_n \xrightarrow{\mathbb{P}} c$, où c est une constante, si et seulement si $X_n \xrightarrow{\mathcal{L}} c$,
 d) Si $X_n \xrightarrow{\mathcal{L}} X$ et $X_n - Y_n \xrightarrow{\mathbb{P}} 0$, alors $Y_n \xrightarrow{\mathcal{L}} X$,

- e) Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathbb{P}} c$ pour une constante c , alors, $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, c)$,
 f) Si $X_n \xrightarrow{\mathbb{P}} X$ et $Y_n \xrightarrow{\mathbb{P}} Y$ alors $(X_n, Y_n) \xrightarrow{\mathbb{P}} (X, Y)$.

Preuve. (a)) : Si $X_n \xrightarrow{\mathbb{P}-\text{p.s.}} X$, pour tout $\epsilon > 0$, $\mathbb{P}(\bigcap_n A_n) = 0$ avec $A_n = \bigcup_{k \geq n} \{\|X_k - X\| \geq \epsilon\}$. Par conséquent :

$$\lim_n \mathbb{P}(\|X_n - X\| \geq \epsilon) \leq \lim_n \mathbb{P}(A_n) = 0$$

(b)) : Soit f Lipschitzienne bornée, notons $|f|_\infty = \sup_x |f(x)|$ et $|f|_{\text{Lip}} = \sup_{x \neq y} |f(x) - f(y)| / \|x - y\|$. Pour tout $\epsilon > 0$,

$$|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leq \epsilon |f|_{\text{Lip}} + 2|f|_\infty \mathbb{P}(\|X_n - X\| \geq \epsilon).$$

Le second terme du membre de droite tend vers 0 et le premier peut être rendu arbitrairement petit. Donc, $\lim_n \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$.

(c)) : (b)) montre que si $X_n \xrightarrow{\mathbb{P}} c$, alors $X_n \xrightarrow{\mathcal{L}} c$. Pour la réciproque, soit $\epsilon > 0$ et soit $B(c, \epsilon) = \{x : \|x - c\| < \epsilon\}$ la boule ouverte de centre c et de rayon ϵ . Nous avons $\mathbb{P}(\|X_n - c\| \geq \epsilon) = \mathbb{P}(X_n \in B(c, \epsilon)^c)$. Si $X_n \xrightarrow{\mathcal{L}} c$, le lemme de Portmanteau montre que $\limsup \mathbb{P}(X_n \in B(c, \epsilon)^c) \leq \mathbb{P}(c \in B(c, \epsilon)^c) = 0$

(d)) : Soit f Lipschitzienne bornée, et $\epsilon > 0$,

$$|\mathbb{E}[f(X_n)] - \mathbb{E}[f(Y_n)]| \leq |f|_{\text{Lip}} \epsilon + |f|_\infty \mathbb{P}(\|X_n - Y_n\| \geq \epsilon).$$

Le second terme tend vers 0 et le premier peut être rendu arbitrairement petit, donc $\mathbb{E}[f(X_n)]$ et $\mathbb{E}[f(Y_n)]$ ont la même limite.

(e)) : Remarquons, en notant $d(x, y)$ la distance euclidienne, que $d((X_n, Y_n), (X_n, c)) = d(Y_n, c) \rightarrow 0$. Donc, en utilisant (d)), il suffit de prouver que $(X_n, c) \xrightarrow{\mathcal{L}} (X, c)$. Pour toute fonction continue bornée $f : (x, y) \rightarrow f(x, y)$ la fonction $f(\cdot, c) : x \rightarrow f(x, c)$ est continue et bornée et $|\mathbb{E}[f(X_n, c)] - \mathbb{E}[f(X, c)]| \rightarrow 0$, car $X_n \xrightarrow{\mathcal{L}} X$.

(f)) : conséquence directe de $d((x_1, y_1), (x_2, y_2)) \leq d(x_1, y_1) + d(x_2, y_2)$. \square

Le théorème 8.5.1 a quelques applications immédiates et utiles en pratique.

La propriété (f)) montre que la convergence en probabilité d'une suite de vecteurs aléatoires $X_n = (X_{n,1}, \dots, X_{n,k})$ est équivalente à la convergence de chacune de ses composantes individuellement. Le résultat analogue pour la convergence en distribution est faux : la convergence en distribution d'une suite de vecteurs aléatoires est une propriété plus forte que la convergence en distribution de chacune de ses composantes $X_{n,i}$.

La propriété (e)) du théorème 8.5.1 implique que si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathbb{P}} c$, alors $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, c)$. Le théorème de continuité (théorème 8.4.1) montre donc que pour toute fonction $g : (x, y) \rightarrow g(x, y)$ continue sur un ensemble $C \times \{c\}$, $\mathbb{P}(X \in C) = 1$, $g(X_n, Y_n) \xrightarrow{\mathcal{L}} g(X, c)$. Des applications particulières

de ce principe sont souvent regroupées sous la forme du lemme suivant, connue sous le nom de lemme de Slutsky.

Lemme 8.5.2 (Lemme de Slutsky). *Soit X_n, X et Y_n des v.a. réelles. Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} c$ où c est une constante, alors*

- a) $X_n + Y_n \xrightarrow{\mathcal{L}} X + c$;
- b) $Y_n X_n \xrightarrow{\mathcal{L}} cX$;
- c) $Y_n^{-1} X_n \xrightarrow{\mathcal{L}} c^{-1} X$ pour $c \neq 0$.

Le procédé de Cramér-Wold (voir section 8.3) permet d'étendre ce résultat au cas vectoriel/matriciel, pour peu que, dans (a)), c soit un vecteur de même dimension que X , et dans (b)) et (c)), $\{Y_n\}_n$ et c soient des matrices (avec c inversible pour (c)) de même dimension adaptée à celle des vecteurs $\{X_n\}$.

Il est pratique de disposer de notations simples pour exprimer qu'une suite tend vers 0 en probabilité ou est bornée en probabilité. Nous dirons qu'une suite de v.a. (scalaire ou vecteur) $X_n = o_P(1)$ si $X_n \xrightarrow{\mathbb{P}} 0$; nous noterons $X_n = O_P(1)$ si X_n est bornée en probabilité. Plus généralement, pour R_n une suite de v.a.,

$$X_n = o_P(R_n) \quad \text{signifie} \quad X_n = Y_n R_n \quad \text{avec} \quad Y_n = o_P(1), \quad (8.2)$$

$$X_n = O_P(R_n) \quad \text{signifie} \quad X_n = Y_n R_n \quad \text{avec} \quad Y_n = O_P(1), \quad (8.3)$$

$$(8.4)$$

Pour X_n et R_n des suites déterministes, les symboles o_P et O_P coïncident avec les symboles o et O de l'analyse. Les symboles o_P et O_P se manipulent d'ailleurs exactement de la même façon que les symboles o et O . Par exemple,

$$\begin{aligned} o_P(1) + o_P(1) &= o_P(1), \\ o_P(1) + O_P(1) &= O_P(1), \\ O_P(1) o_P(1) &= o_P(1), \\ (1 + o_P(1))^{-1} &= O_P(1), \\ o_P(R_n) &= R_n o_P(1), \quad O_P(R_n) = R_n O_P(1) \end{aligned}$$

La règle de calcul suivante est utile pour les développements asymptotiques.

Lemme 8.5.3. *Soit R une fonction $D \subset \mathbb{R}^k$ telle que $R(0) = 0$. Soit X_n une suite de v.a. à valeurs dans D tels que $X_n \rightarrow 0$. Alors, pour tout $p > 0$,*

- (i) *Si $R(h) = o(\|h\|^p)$ quand $h \rightarrow 0$, alors $R(X_n) = o_P(\|X_n\|^p)$,*
- (ii) *Si $R(h) = O(\|h\|^p)$ quand $h \rightarrow 0$, alors $R(X_n) = O_P(\|X_n\|^p)$.*

Preuve. Définissons $g(h) = R(h)/\|h\|^p$ pour $h \neq 0$ et $g(0) = 0$: $R(X_n) = g(X_n)\|X_n\|^p$.

(i) La fonction g est continue en 0 et le théorème de continuité montre que $g(X_n) \xrightarrow{\mathbb{P}} g(0) = 0$. La deuxième assertion se démontre de façon similaire. \square

8.6 Delta-méthode

Supposons que nous disposions d'un estimateur T_n d'un paramètre ν , mais que la quantité d'intérêt soit $g(\nu)$. Un estimateur naturel de cette quantité est $g(T_n)$. Pouvons nous déduire le comportement asymptotique de $g(T_n)$ à partir de celui de T_n ?

Nous savons déjà que si T_n converge en probabilité vers ν et que g est continue au point ν , alors $g(T_n)$ converge en probabilité vers $g(\nu)$.

Si nous savons $\sqrt{n}(T_n - \nu)$ converge en loi vers une distribution limite, pouvons nous affirmer qu'il en est de même pour $\sqrt{n}(g(T_n) - g(\nu))$? La réponse est affirmative si la fonction g est différentiable au point ν : de façon heuristique, nous avons :

$$\sqrt{n}(g(T_n) - g(\nu)) \simeq g'(\nu)\sqrt{n}(T_n - \nu),$$

et donc, si $\sqrt{n}(T_n - \nu) \rightarrow_d T$, alors $\sqrt{n}(g(T_n) - g(\nu)) \rightarrow_d [g'(\nu)]T$. En particulier, si $\sqrt{n}(T_n - \nu)$ est asymptotiquement $\mathbf{N}(0, \sigma^2)$ alors $\sqrt{n}(g(T_n) - g(\nu))$ est asymptotiquement normal $\mathbf{N}(0, [g'(\nu)]^2 \sigma^2)$.

La même question se pose lorsque $\mathbf{T}_n = (T_{n,1}, \dots, T_{n,k})$ est un vecteur aléatoire et g est une fonction de $\mathbb{R}^k \mapsto \mathbb{R}^m$. Nous connaissons la loi de $\sqrt{n}(\mathbf{T}_n - \nu)$ et nous nous intéressons à la loi de $\sqrt{n}(g(\mathbf{T}_n) - g(\nu))$. Le résultat ci-dessus s'étend directement en remplaçant la dérivée par la différentielle $\nabla_\nu g(\nu)$.

Proposition 8.6.1. *Soit $g : D_g \subset \mathbb{R}^k \mapsto \mathbb{R}^m$ une fonction définie sur un sous ensemble D_g de \mathbb{R}^k et différentiable au point ν . Soit \mathbf{T}_n des v.a. à valeurs dans D_g . Si $r_n(\mathbf{T}_n - \nu) \xrightarrow{\mathcal{L}} T$, pour une suite $r_n \nearrow \infty$, alors $r_n(g(\mathbf{T}_n) - g(\nu)) \xrightarrow{\mathcal{L}} \nabla_\nu g(\nu) \mathbf{T}$. De plus $r_n(g(\mathbf{T}_n) - g(\nu)) - \nabla_\nu g(\nu)(r_n(\mathbf{T}_n - \nu)) \xrightarrow{\mathbb{P}} 0$.*

Preuve. $r_n(\mathbf{T}_n - \nu)$ converge en distribution, la suite $r_n(\mathbf{T}_n - \nu)$ est bornée en probabilité et $\mathbf{T}_n - \nu$ tend vers 0. La différentiabilité de la fonction g implique que $g(\nu + h) = g(\nu) + \nabla_\nu g(\nu)h + R(h)$ et $R(h) = o(\|h\|)$. Le lemme 8.5.3 montre que :

$$g(T_n) - g(\nu) - \nabla_\nu g(\nu)(T_n - \nu) = R(T_n - \nu) = o_P(\|T_n - \nu\|).$$

En multipliant les deux membres de l'identité précédente par r_n , nous avons donc :

$$r_n(g(T_n) - g(\nu)) = \nabla_\nu g(\nu)(r_n(T_n - \nu)) + o_P(1).$$

Nous concluons en appliquant le lemme de Slutsky. □

8.7 Intégrabilité uniforme et Convergence des moments

Définition 8.7.1 (Uniforme Intégrabilité). Une suite de v.a. Y_n est dite *uniformément intégrable* si

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E} [\|Y_n\| \mathbb{1}(\|Y_n\| \geq M)] = 0.$$

Théorème 8.7.2. Si la suite (Y_n) est uniformément intégrable, alors

1. $\sup_n \mathbb{E} [|Y_n|] < \infty$,
2. $\mathbb{E} [\liminf Y_n] \leq \liminf \mathbb{E} [Y_n] \leq \limsup \mathbb{E} [Y_n] \leq \mathbb{E} [\limsup Y_n]$,
3. Si $Y_n \xrightarrow{\mathbb{P}\text{-p.s.}} Y_\infty$ alors, $\mathbb{E} [|Y_\infty|] < \infty$ et $\lim_{n \rightarrow \infty} \mathbb{E} [|Y_n - Y_\infty|] = 0$

La preuve est élémentaire et laissée au lecteur.

Comme le montre le théorème ci-dessous, l'uniforme intégrabilité permet de relier la convergence en loi et la convergence des moments.

Théorème 8.7.3. Soit $f : \mathbb{R}^k \mapsto \mathbb{R}$ une fonction borélienne continue en tout point de $C \in \mathcal{B}(\mathbb{R}^k)$. Supposons que $X_n \xrightarrow{\mathcal{L}} X$ et $\mathbb{P}(X \in C) = 1$. Alors, $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ si et seulement si la suite $f(X_n)$ est uniformément intégrable.

Preuve. Nous ne montrons ici que la réciproque. Posons $Y_n = f(X_n)$ et supposons que Y_n est uniformément intégrable. Nous allons montrer que $\mathbb{E}[Y_n] \rightarrow \mathbb{E}[Y]$, où $Y = f(X)$. Nous supposons sans perte de généralité que Y_n est positive (il suffit autrement de raisonner sur les parties positives et négatives séparément). Le théorème de continuité montre que $Y_n \xrightarrow{\mathcal{L}} Y$. Nous notons $a \wedge b = \inf(a, b)$. L'inégalité triangulaire donne d'une part

$$\mathbb{E}[Y \wedge M] \leq |\mathbb{E}[Y_n \wedge M] - \mathbb{E}[Y \wedge M]| + \mathbb{E}[Y_n \wedge M]$$

et d'autre part

$$\begin{aligned} |\mathbb{E}[Y_n] - \mathbb{E}[Y]| &\leq |\mathbb{E}[Y_n] - \mathbb{E}[Y_n \wedge M]| \\ &\quad + |\mathbb{E}[Y_n \wedge M] - \mathbb{E}[Y \wedge M]| + |\mathbb{E}[Y \wedge M] - \mathbb{E}[Y]|. \end{aligned}$$

Comme la fonction $y \rightarrow y \wedge M$ est continue et bornée, $|\mathbb{E}[Y_n \wedge M] - \mathbb{E}[Y \wedge M]| \rightarrow 0$ quand $n \rightarrow \infty$. Le second terme de la partie droite de la première inégalité est majorée indépendamment de M (voir ci-dessus), donc $\mathbb{E}[Y] < \infty$. Le premier terme et le troisième termes de la partie droite de la seconde inégalité peuvent être rendus arbitrairement petits en utilisant respectivement l'uniforme intégrabilité et $\mathbb{E}[Y] < \infty$, ce qui achève la démonstration de l'implication réciproque. \square

Une conséquence élémentaire de l'exemple précédent est donnée ci-après.

Proposition 8.7.4. *Si $\{X_k, k \in \mathbb{N}\}$ est une famille de vecteurs aléatoires gaussiens, alors les deux assertions suivantes sont équivalentes.*

$$(i) \lim_{k \rightarrow \infty} \mathbb{E}[X_k] = \mu \text{ et } \lim_{k \rightarrow \infty} \mathbb{V}(X_k) = \Sigma$$

$$(ii) X_k \xrightarrow{\mathcal{L}} \mathbf{N}(\mu, \Sigma).$$

Chapitre 9

Estimation de la moyenne

9.1 Consistence de la moyenne empirique

Soit $\{X_t\}$ un processus aléatoire à temps discret stationnaire au second ordre, de moyenne $\mathbb{E}[X_0] = \mu$, et de fonction d'autocovariance γ . On suppose avoir observé n échantillons consécutifs X_1, \dots, X_n du processus. L'estimateur de μ que nous considérons est la *moyenne empirique* définie par :

$$\hat{\mu}_n = \frac{1}{n} \sum_{t=1}^n X_t \quad (9.1)$$

Théorème 9.1.1. *Soit (X_t) un processus stationnaire au second ordre de moyenne μ et de fonction d'autocovariance $(\gamma(h))$. Si $\lim_{h \rightarrow \infty} \gamma(h) = 0$, alors*

$$\lim_{n \rightarrow \infty} \mathbb{E}[(\hat{\mu}_n - \mu)^2] = 0. \quad (9.2)$$

Si de plus $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, alors

$$\mathbb{E}[(\hat{\mu}_n - \mu)^2] \leq n^{-1} \left(\sum_{h=-\infty}^{\infty} |\gamma(h)| \right) \quad \text{et} \quad \lim_{n \rightarrow \infty} n \mathbb{E}[(\hat{\mu}_n - \mu)^2] = 2\pi f(0), \quad (9.3)$$

où $f(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma(\tau) e^{-i\tau\lambda}$ est la densité spectrale du processus (X_t)

Preuve. Sans perte de généralité, nous posons $\mu = 0$, ce qui implique que $\mathbb{E}[(\hat{\mu}_n - \mu)^2] = \mathbb{V}(\hat{\mu}_n)$. Nous avons

$$\begin{aligned} \mathbb{V}(\hat{\mu}_n) &= n^{-2} \sum_{t=1}^n \sum_{j=1}^n \text{Cov}(X_t, X_j) = 2n^{-2} \sum_{t=1}^n \sum_{j=1}^t \text{Cov}(X_t, X_j) - n^{-1} \gamma(0) \\ &= 2n^{-2} \sum_{t=1}^n t \text{Cov}(\hat{\mu}_t, X_t) - n^{-1} \gamma(0) \leq 2n^{-1} \sum_{t=1}^n |\text{Cov}(\hat{\mu}_t, X_t)|, \end{aligned}$$

qui converge vers 0. Lorsque $(\gamma(h))$ est absolument sommable, le théorème de la convergence dominée appliquée à

$$\mathbb{V}(\hat{\mu}_n) = \frac{1}{n^2} \sum_{s=1}^n \sum_{t=1}^n \gamma(t-s) = \frac{1}{n} \sum_{h=-n+1}^{n-1} \left(1 - \frac{|h|}{n}\right) \gamma(h) \quad (9.4)$$

montre que

$$\sum_{h=-\infty}^{\infty} \lim_{n \rightarrow \infty} \left(1 - \frac{|h|}{n}\right) \gamma(h) = \sum_{h=-\infty}^{\infty} \gamma(h) = 2\pi f(0)$$

où $f(\lambda) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-ih\lambda}$ est la densité spectrale du processus (X_t) . \square

Théorème 9.1.2. *Soit (X_t) un processus stationnaire au second ordre de moyenne μ et de fonction d'autocovariance $(\gamma(h))$. Si $\sum_{h=-\infty}^{\infty} |\gamma(h)| = 0$, alors $\hat{\mu}_n \rightarrow \mu$ -p.s.*

Preuve. Sans perte de généralité, nous supposons que $\mu = 0$. Pour tout $\epsilon > 0$, l'inégalité de Markov montre que

$$\sum_{n=1}^{\infty} \mathbb{P}(|\mu_{n^2}| \geq \epsilon) \leq \sum_h |\gamma(h)| \sum_{n=1}^{\infty} n^{-2} < \infty.$$

Le lemme de Borel-Cantelli (lemme 8.1.5) implique que $\lim_{n \rightarrow \infty} \hat{\mu}_{n^2} = 0$ p.s. . Considérons la décomposition

$$\hat{\mu}_n = \frac{m_n}{n} \hat{\mu}_{m_n} + \rho_n, \quad (9.5)$$

avec $m_n = \lfloor \sqrt{n} \rfloor^2$ et $\rho_n = n^{-1} \sum_{j=m_n+1}^n X_j$. Comme $\mathbb{E}[\rho_n^2] \leq n^{-2}(n - m_n) \sum_h |\gamma(h)|$, et

$$0 \leq n - m_n \leq n^{-2} [n - (\sqrt{n} - 1)^2] = n^{-2} (2\sqrt{n} - 1),$$

pour tout $\epsilon > 0$, nous avons

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}(|\rho_n| \geq \epsilon) &\leq \epsilon^{-2} \sum_{n=1}^{\infty} \mathbb{E}[\rho_n^2] \\ &\leq \epsilon^{-2} \sum_h |\gamma(h)| \sum_{n=1}^{\infty} n^{-2} (2\sqrt{n} - 1) < \infty, \end{aligned}$$

et le lemme de Borel-Cantelli implique que $\rho_n \rightarrow 0$, \mathbb{P} -p.s. . Comme $\mu_{m_n} \rightarrow \mu$ \mathbb{P} -p.s. et $m_n/n \rightarrow 1$, nous avons donc $\hat{\mu}_n \rightarrow \mu$, \mathbb{P} -p.s. . \square

Ces deux théorèmes montrent que la loi des grands nombres faibles et fortes, établie classiquement pour des variables aléatoires indépendantes où des incréments de martingales, est également valable pour un processus stationnaire au second ordre, du moment que la fonction d'autocovariance décroît suffisamment rapidement à l'infini.

9.2 Distribution asymptotique : processus linéaire fort

Cependant pour construire des intervalles de confiance, il est nécessaire d'obtenir un résultat plus précis portant sur la distribution limite de $\sqrt{n}(\hat{\mu}_n -$

μ). Remarquons tout d'abord que si le processus $\{Z_t, t \in \mathbb{Z}\}$ est un bruit blanc fort gaussien, la variable \bar{X}_n est elle aussi gaussienne, de moyenne nulle et de variance

$$\mathbb{V}(\bar{X}_n) = n^{-1} \sum_{|u| < n} \left(1 - \frac{|u|}{n}\right) \gamma(u),$$

et comme

$$\lim_{n \rightarrow \infty} \sum_{|u| < n} \left(1 - \frac{|u|}{n}\right) \gamma(u) = \sum_{u=-\infty}^{\infty} \gamma(u) = \sigma^2 \left(\sum_{j=-\infty}^{\infty} \psi_j \right)^2 \stackrel{\text{def}}{=} V, \quad (9.6)$$

nous avons donc $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V)$. Nous allons montrer que ce résultat reste vrai de façon générale. L'obtention de théorèmes de type limite centrale pour des suites de variables aléatoires *dépendantes* est un sujet délicat, qui a donné lieu à une vaste littérature. Il n'est bien entendu pas question ici de présenter une théorie générale et nous nous contentons donc d'énoncer un résultat valable dans le cas de processus linéaires forts, qui sont des cas particulier de processus linéaires (définition 4.1.3).

Définition 9.2.1 (Processus linéaire fort). On dira qu'un processus $(X_t)_{t \in \mathbb{Z}}$ est un processus linéaire fort de moyenne μ si on a

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

avec $\sum_k |\psi_k| < \infty$ et $Z_t \sim \text{IID}(0, \sigma^2)$.

Le fait de devoir émettre une hypothèse aussi contraignante sur la loi du processus dans un contexte où, en fait, seules les propriétés au second ordre nous intéressent est bien sûr frustrant, mais il traduit la difficulté technique d'un tel résultat. Nous allons tout d'abord établir un théorème de la limite centrale pour des processus m -dépendants, puis nous étendrons ce résultat aux processus linéaires forts.

9.2.1 Théorèmes limites pour les processus m -dépendants

Définition 9.2.2 (Processus m -dépendant). Nous dirons que le processus $\{X_k, k \in \mathbb{Z}\}$ est m -dépendant, si pour tout $n \in \mathbb{Z}$, les tribus $\mathcal{P}_n \stackrel{\text{def}}{=} \sigma(X_k, k \leq n)$ et $\mathcal{F}_{n+m+1} \stackrel{\text{def}}{=} \sigma(X_k, k \geq n+m+1)$ sont indépendantes.

Théorème 9.2.3. Soit $\{X_k, k \in \mathbb{Z}\}$ un processus stationnaire au sens strict m -dépendant. Nous supposons que $\mathbb{E}[X_0^2] < \infty$ et nous notons par $\gamma(\cdot)$ la

fonction d'autocovariance du processus. Alors,

$$\sqrt{n} (\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V),$$

où $\mu = \mathbb{E}[X_0]$ et $V = \sum_{h \in \mathbb{Z}} \gamma(h)$.

Pour établir ce théorème, nous nous appuierons sur le résultat suivant.

Théorème 9.2.4. Soient $(A_n, n \geq 0)$, $(B_m, m \geq 0)$, $(B_{m,n}, m \geq 0, n \geq 0)$ et B des vecteurs aléatoires vérifiant

- a) Pour tout m , $B_{m,n} \xrightarrow{\mathcal{L}} B_m$ quand $n \rightarrow \infty$,
- b) $B_m \xrightarrow{\mathcal{L}} B$ pour $m \rightarrow \infty$,
- c) Pour tout $\epsilon > 0$, $\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(|A_n - B_{m,n}| \geq \epsilon) = 0$.

Alors $A_n \xrightarrow{\mathcal{L}} B$ quand $n \rightarrow \infty$.

Remarque 9.2.5. Pour établir la condition (c)) il est souvent pratique d'utiliser l'inégalité de Markov.

Preuve. En appliquant le théorème de Levy (voir théorème 8.3.3), il suffit de montrer que, pour tout λ , $\phi_{A_n}(\lambda) \rightarrow \phi_B(\lambda)$, où $\phi_Z(\lambda)$ est la fonction caractéristique du vecteur aléatoire Z . L'inégalité triangulaire montre que

$$|\phi_{A_n}(\lambda) - \phi_B(\lambda)| \leq |\phi_{A_n}(\lambda) - \phi_{B_{m,n}}(\lambda)| + |\phi_{B_{m,n}}(\lambda) - \phi_{B_m}(\lambda)| + |\phi_{B_m}(\lambda) - \phi_B(\lambda)|.$$

Considérons tout d'abord le premier terme.

$$\begin{aligned} |\phi_{A_n}(\lambda) - \phi_{B_{m,n}}(\lambda)| &= \left| \mathbb{E} \left[e^{i\lambda^t A_n} \right] - \mathbb{E} \left[e^{i\lambda^t B_{m,n}} \right] \right| \\ &\leq \mathbb{E} \left[\left| e^{i\lambda^t A_n} \left(1 - e^{i(\lambda^t B_{m,n} - \lambda^t A_n)} \right) \right| \right] \\ &\leq \mathbb{E} \left[\left| 1 - e^{i(\lambda^t B_{m,n} - \lambda^t A_n)} \right| \right]. \end{aligned}$$

Pour $\delta > 0$, posons $\mathbf{A}_{m,n}(\delta) \stackrel{\text{def}}{=} \{|\lambda^t B_{m,n} - \lambda^t A_n| \geq \delta\}$. Nous avons donc :

$$\begin{aligned} |\phi_{A_n}(\lambda) - \phi_{B_{m,n}}(\lambda)| &\leq \mathbb{E} \left[\left| 1 - e^{i(\lambda^t B_{m,n} - \lambda^t A_n)} \right| \mathbb{1}_{\mathbf{A}_{m,n}(\delta)} \right] \\ &\quad + \mathbb{E} \left[\left| 1 - e^{i(\lambda^t B_{m,n} - \lambda^t A_n)} \right| \mathbb{1}_{\mathbf{A}_{m,n}^c(\delta)} \right]. \end{aligned}$$

Pour λ et $\epsilon > 0$, nous choisissons $\delta(\epsilon)$ tel que $\left| 1 - e^{i(\lambda^t B_{m,n} - \lambda^t A_n)} \right| \mathbb{1}_{\mathbf{A}_{m,n}^c(\delta(\epsilon))} < \epsilon$. Comme $\left| 1 - e^{i(\lambda^t B_{m,n} - \lambda^t A_n)} \right| \leq 2$, l'inégalité précédente implique

$$|\phi_{A_n}(\lambda) - \phi_{B_{m,n}}(\lambda)| \leq 2\mathbb{P}[\mathbf{A}_{m,n}(\delta(\epsilon))] + \epsilon.$$

Comme $\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}[\mathbf{A}_{m,n}(\delta(\epsilon))] = 0$, nous pouvons choisir un entier $m(\epsilon)$ tel que, pour tout $m \geq m(\epsilon)$ il existe un entier $n(m, \epsilon)$ tel que

$$\mathbb{P}[A_{m,n}(\delta(\epsilon))] \leq \epsilon, \quad \text{pour tout } n \geq n(m, \epsilon).$$

La condition (b)) montre que l'on peut choisir $m \geq m(\epsilon)$ assez grand pour que $|\phi_{B_m}(\lambda) - \phi_B(\lambda)| \geq \epsilon$. La condition (a)) montre que l'on peut choisir $n \geq n(m, \epsilon)$ suffisamment grand pour que $|\phi_{B_{m,n}}(\lambda) - \phi_{B_m}(\lambda)| < \epsilon$, ce qui conclut la preuve du théorème. \square

Preuve (Preuve du théorème 9.2.3). Sans perte de généralité, nous supposons que $\mu = 0$. Nous allons utiliser le théorème d'approximation 9.2.4 en construisant un tableau de variables aléatoires $\{B_{m,n}, (m, n) \in \mathbb{N} \times \mathbb{N}\}$ approchant $A_n \stackrel{\text{def}}{=} n^{1/2} \bar{X}_n$. Pour $m \geq 2M$, considérons

$$B_{m,n} = n^{-1/2} \sum_{k=1}^{r_n} Z_{m,k}$$

où $r_n = \lfloor n/m \rfloor$ et $Z_{m,k} = (X_{(k-1)m+1} + \dots + X_{km-M})$. Cette approximation contient une partie des termes de la somme \bar{X}_n , mais les variables aléatoires $Z_{m,1}, \dots, Z_{m,r_n}$ sont indépendantes. Comme le processus est stationnaire au sens strict, les vecteurs aléatoires $\{Z_{m,1}, \dots, Z_{m,r}\}$ sont indépendants et identiquement distribués de moyenne nulle et de variance

$$S_{m-M} \stackrel{\text{def}}{=} \sum_{|h| \leq M} (m - M - |h|) \gamma(h), \quad m \geq 2M. \quad (9.7)$$

Nous allons vérifier que cette approximation satisfait les conditions du théorème 9.2.4.

- a) En appliquant le théorème de Limite Centrale aux vecteurs aléatoires $\{Z_{m,k}, k \geq 0\}$, nous obtenons

$$B_{m,n} = n^{-1/2} \sum_{k=1}^{r_n} Z_{m,k} = (n/r_n)^{-1/2} r_n^{-1/2} \sum_{k=1}^{r_n} Z_{m,k},$$

et en utilisant $\lim_{n \rightarrow \infty} (n/r_n)^{-1/2} = m^{1/2}$, nous obtenons $B_{m,n} \xrightarrow{\mathcal{L}} B_m$ lorsque $n \rightarrow \infty$, où B_m est une variable aléatoire Gaussienne de moyenne nulle et de variance S_{m-M}/m .

- b) Comme $\lim_{m \rightarrow \infty} S_{m-M}/m = V_M$, nous avons $B_m \xrightarrow{\mathcal{L}} B$, où B est une variable aléatoire gaussienne de moyenne nulle et de variance V_M . En effet, la fonction caractéristique est donnée par $\mathbb{E}[e^{i\lambda B_m}] = e^{-\lambda^2 S_{m-M}/(2m)}$ et, donc, pour tout $\lambda \in \mathbb{R}$, $\lim_{m \rightarrow \infty} \mathbb{E}[e^{i\lambda B_m}] = e^{-\lambda^2 V_M/2}$. Notons que $\lambda \mapsto e^{-\lambda^2 V_M/2}$ est la fonction caractéristique d'une variable aléatoire gaussienne centrée de variance V_M . le résultat est donc une conséquence de la caractérisation du théorème Levy (voir Théorème 8.3.3).
- c) Pour vérifier la dernière condition, considérons la différence

$$n^{1/2}\bar{X}_n - B_{m,n} = n^{-1/2}(W_{m,1} + \cdots + W_{m,r_n}) ,$$

où $W_{m,k} = X_{km-M+1} + \cdots + X_{km}$ pour $1 \leq k \leq r_n - 1$ et $W_{m,r_n} = X_{r_nm-M+1} + \cdots + X_n$. Les variables aléatoires $W_{m,k}$, $1 \leq k \leq r_n$ sont indépendantes. La variance des $r_n - 1$ premières variables est égale à S_M où S_M est défini par (9.7). La variance de W_{m,r_n} est donnée par

$$\begin{aligned} \mathbb{V}(W_{m,r_n}) &= \sum_{|u| \leq m-M} (n - \lfloor n/m \rfloor m + M - |u|) \gamma(u) \\ &\leq \sum_{|u| \leq m-M} (m + M - |u|) \gamma(u) . \end{aligned}$$

En utilisant

$$\mathbb{V}(n^{1/2}\bar{X}_n - B_{m,n}) = n^{-1}[(r_n - 1)S_M + \mathbb{V}(W_{m,r_n})] ,$$

nous avons

$$\limsup_{n \rightarrow \infty} \mathbb{V}(n^{1/2}\bar{X}_n - B_{m,n}) = m^{-1}S_M ,$$

et donc

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{V}(n^{-1}\bar{X}_n - B_{m,n}) = 0 .$$

□

9.2.2 Théorème limite pour des processus linéaires forts

Théorème 9.2.6. Soit $\{X_t\}$ un processus linéaire fort de moyenne μ . On pose $\hat{\mu}_n = n^{-1} \sum_{t=1}^n X_t$. Alors :

$$\sqrt{n}(\hat{\mu}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 2\pi f(0)) \quad (9.8)$$

où $f(0)$ est la densité spectrale de $(X_t)_{t \in \mathbb{Z}}$ prise à la fréquence nulle.

Preuve. Sans perte de généralité, nous supposons que $\mu = 0$. On écrit donc

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

avec $(Z_t)_{t \in \mathbb{Z}} \sim \text{IID}(0, \sigma^2)$. On a alors

$$f(0) = \sigma^2 |\hat{\psi}(0)|^2 / (2\pi) \quad \text{avec} \quad \hat{\psi}(\lambda) = \sum_{j=-\infty}^{\infty} \psi_j e^{ij\lambda} .$$

Pour établir ce résultat, nous allons utiliser encore le théorème d'approximation 9.2.4. L'idée est d'approcher le processus $\{X_t, t \in \mathbb{Z}\}$, par une suite de processus $(2m+1)$ -dépendant, $X_t^m = \sum_{i=-m}^m \psi_i Z_{t-i}$ et nous construisons les variables :

$$Y_{m,n} = n^{-1/2} \sum_{t=1}^n X_t^m.$$

1. En appliquant le théorème 9.2.3, nous avons, pour tout m , $Y_{m,n} \xrightarrow{\mathcal{L}} Y_m$ quand $n \rightarrow \infty$, où Y_m est une variable aléatoire gaussienne de moyenne nulle et de variance V_m , avec

$$V_m = \sum_{h=-2m}^{2m} \gamma^m(h) = \sigma^2 \left(\sum_{j=-m}^m \psi_j \right)^2. \quad (9.9)$$

2. Comme $V_m \rightarrow V$ quand $m \rightarrow \infty$, nous avons, $Y_m \xrightarrow{\mathcal{L}} Y$, où Y est une variable aléatoire gaussienne de moyenne nulle et de variance $V = \sigma^2 \left(\sum_{j=-\infty}^{\infty} \psi_j \right)^2$.
3. Finalement, en utilisant (9.3),

$$\mathbb{V} \left(n^{1/2} \bar{X}_n - Y_{m,n} \right) = \mathbb{V} \left(n^{-1/2} \sum_{t=1}^n \sum_{|j|>m} \psi_j Z_{t-j} \right) \leq \sigma^2 \left(\sum_{|j|>m} |\psi_j| \right)^2$$

qui converge vers 0 quand $m \rightarrow \infty$.

□

Exemple 9.2.7 (Moyenne empirique pour un processus AR(1) (fort)). Soit X_t un processus autorégressif d'ordre 1 fort, de moyenne μ , solution stationnaire au second ordre défini par l'équation de récurrence

$$X_t - \mu = \phi(X_{t-1} - \mu) + Z_t$$

où $\{Z_t\} \sim \text{IID}(0, \sigma^2)$ et $|\phi| < 1$. Nous rappelons que la fonction d'autocovariance d'un processus AR(1) pour $|\phi| < 1$ est donnée par

$$\gamma_X(k) = \frac{\sigma^2}{(1 - \phi^2)} \phi^{|k|}$$

et que la densité spectrale de ce processus a pour expression

$$f(\lambda) = \frac{\sigma^2}{2\pi |1 - \phi e^{-i\lambda}|^2}$$

Dans ce cas, la variance limite qui intervient dans l'équation (9.8), est donnée par $2\pi f(0) = \sigma^2/(1 - \phi)^2$. Cette valeur est à comparer avec la vari-

ance de X_t donnée par $\gamma(0) = \sigma^2/(1 - \phi^2)$. On constate que le rapport $2\pi f(0)/\gamma(0) = (1 + \phi)/(1 - \phi)$ tend vers 0 lorsque $\phi \rightarrow -1$ et vers $+\infty$ lorsque $\phi \rightarrow 1$. Ce qui implique par exemple lorsque l'on considère l'intervalle de confiance asymptotique de niveau 95% pour la moyenne μ donné par $[\hat{\mu}_n - 1.96\sigma n^{-1/2}/(1 - \phi), \hat{\mu}_n + 1.96\sigma n^{-1/2}/(1 - \phi)]$ que la longueur de l'intervalle de confiance est minimale lorsque $\phi = -1$ (corrélacion négative). et maximale lorsque $\phi \rightarrow 1$ (corrélacion positive).

9.3 Extension à des processus linéaires plus généraux

Il est possible d'affaiblir l'hypothèse pour couvrir le cas de processus linéaires qui ne sont pas linéaires forts. Nous établissons tout d'abord un théorème central limite pour un filtre à réponse impulsionnelle finie.

Proposition 9.3.1. *Soit (X_t) un processus tel que*

$$X_t = \mu + \sum_{i=-m}^m \psi_i Z_{t-i}, \quad t \in \mathbb{Z},$$

où (Z_t) est un $\text{BB}(0, \sigma^2)$. Supposons que $n^{-1/2} \sum_{t=1}^n Z_t$ converge en loi vers une loi normale de moyenne nulle et de variance σ^2 . Alors,

$$\sqrt{n}(\hat{\mu}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 2\pi f(0)) \quad (9.10)$$

où $f(\lambda)$ est la densité spectrale de (X_t) donnée par $f(\lambda) = \sigma^2 \left| \sum_{j=-m}^m \psi_j e^{-ij\lambda} \right|^2$.

Preuve. Nous posons $\hat{\mu}_n^X = n^{-1} \sum_{t=1}^n X_t$ et $\hat{\mu}_n^Z = n^{-1} \sum_{t=1}^n Z_t$. Nous avons

$$\hat{\mu}_n^X = \mu + \sum_{j=-m}^m \psi_j \left(n^{-1} \sum_{t=1}^n Z_{t-j} \right).$$

Pour $j \geq 0$, nous pouvons écrire

$$\begin{aligned} n^{-1} \sum_{t=1}^n Z_{t-j} &= n^{-1} \sum_{t=1-j}^{n-j} Z_t \\ &= \hat{\mu}_n^Z + n^{-1} \sum_{t=1-j}^0 Z_t - n^{-1} \sum_{t=n-j+1}^n Z_t \stackrel{\text{def}}{=} \hat{\mu}_n^Z + R_{n,j}. \end{aligned}$$

De même, pour $j < 0$, nous avons

$$n^{-1} \sum_{t=1}^n Z_{t-j} = \hat{\mu}_n^Z - n^{-1} \sum_{t=1}^{-j} Z_t + n^{-1} \sum_{t=n+1}^{n-j} Z_t \stackrel{\text{def}}{=} \hat{\mu}_n^Z + R_{n,j}.$$

En utilisant les notations introduites ci-dessus, nous avons donc

$$\sqrt{n} (\hat{\mu}_n^X - \mu) = \left(\sum_{j=-m}^m \psi_j \right) \sqrt{n} \hat{\mu}_n^Z + n^{-1/2} \sum_{j=-m}^m \psi_j R_{n,j}.$$

Un calcul élémentaire montre que

$$\begin{aligned} \mathbb{E} \left[\left(\sum_{j=0}^m \psi_j \sum_{t=1-m}^0 Z_t \mathbb{1}_{\{t \geq 1-j\}} \right)^2 \right] &= n^{-1} \sigma^2 \sum_{t=1-m}^0 \left(\sum_{j=1-t}^m \psi_j \right)^2 \\ &\leq \sigma^2 n^{-1} m \left(\sum_{j=0}^m |\psi_j| \right)^2, \end{aligned}$$

ce qui implique que $\sum_{j=0}^m \psi_j \sum_{t=1-m}^0 Z_t \mathbb{1}_{\{t \geq 1-j\}} = o_P(1)$. Nous montrons de même $\sum_{j=-m}^m \psi_j R_{n,j} = o_P(1)$. \square

Nous pouvons étendre maintenant le résultat au cas d'un filtre à réponse impulsionnelle infinie.

Théorème 9.3.2. *Soit (X_t) un processus linéaire admettant la représentation*

$$X_t = \mu + \sum_{i=-\infty}^{\infty} \psi_i Z_{t-i}, \quad t \in \mathbb{Z},$$

où $\sum_{i=-\infty}^{\infty} |\psi_i| < \infty$ et (Z_t) est un BB(0, σ^2). Supposons que $n^{-1/2} \sum_{t=1}^n Z_t$ converge en loi vers une loi normale de moyenne nulle et de variance σ^2 . Alors la conclusion du théorème 9.2.6 reste vraie.

Preuve. La preuve est similaire à celle du Théorème 9.2.6. Sans perte de généralité, nous posons $\mu = 0$. Nous approchons le processus (X_t) par une suite de processus $X_t^m = \sum_{j=-m}^m \psi_j Z_{t-j}$. La proposition 9.3.1 montre que $\sqrt{n} (\hat{\mu}_n^{X^m} - \mu)$ converge vers une variable aléatoire gaussienne centrée de variance V_m , avec $V_m = \sigma^2 \left(\sum_{j=-m}^m \psi_j \right)^2$. On conclut comme dans Théorème 9.2.6. \square

Problèmes

9.3.3. On considère le problème d'estimer la moyenne dans des observations dépendantes, Y_1, \dots, Y_n

$$Y_t = \mu + X_t,$$

où (X_t) est un processus stationnaire au second ordre, de moyenne nulle $\mathbb{E}[X_t] = 0$ et de covariance $\gamma(\tau) = \text{Cov}(X_\tau, X_0)$. Nous supposons que $\lim_{\tau \rightarrow \infty} \gamma(\tau) = 0$. On note considère des estimateurs de la forme $\hat{\mu}(\mathbf{b}_n) = \mathbf{b}_n^T \mathbf{Y}_n$, où $\mathbf{Y}_n = (Y_1, \dots, Y_n)^T$, où $\mathbf{b}_n^T \mathbf{1}_n = 1$, où $\mathbf{1}_n = (1, \dots, 1)^T$. On note que $\mathbb{E}[\hat{\mu}(\mathbf{b}_n)] = \mu$ pour tout $\mu \in \mathbb{R}$, et donc que des tels estimateurs sont sans biais.

1. Déterminer la variance de la moyenne empirique $\bar{Y}_n = n^{-1} \sum_{y=1}^n Y_t$.
2. Calculer la variance $\mathbb{V}(\hat{\mu}(\mathbf{b}_n))$ en fonction de \mathbf{b}_n et de la matrice Γ_n de covariance du vecteur $(X_1, \dots, X_n)^T$.
3. Justifier (sans démontrer) pourquoi la matrice Γ_n est inversible.
4. Montrer que l'estimateur $\hat{\mu}(\mathbf{b}_n^*)$ minimisant la variance $\mathbf{b}_n \mapsto \mathbb{V}(\hat{\mu}(\mathbf{b}_n))$ sous la contrainte $\mathbf{b}_n^T \mathbf{1}_n = 1$ est donné par : $\hat{\mu}(\mathbf{b}_n^*) = (\mathbf{1}_n^T \Gamma_n^{-1} \mathbf{1}_n)^{-1} \mathbf{1}_n^T \Gamma_n^{-1} \mathbf{Y}_n$.
5. Montrer que sa variance est donnée par $\mathbb{V}(\hat{\mu}(\mathbf{b}_n^*)) = (\mathbf{1}_n^T \Gamma_n^{-1} \mathbf{1}_n)^{-1}$.

On suppose maintenant dans la suite que (X_t) est un processus AR(p) vérifiant

$$X_t = \sum_{i=1}^p \phi_i X_{t-i}, \phi(z) = 1 - \sum_{i=1}^p \phi_i z^i. \quad (9.11)$$

6. Montrer que

$$\mathbb{V}(\hat{\mu}(\mathbf{b}_n^*)) = \left[\sum_{i=0}^{p-1} \sigma_i^{-2} \left(1 - \sum_{j=1}^i \phi_{i,j} \right)^2 + (n-p) \sigma^{-2} \left(1 - \sum_{j=1}^p \phi_j \right)^2 \right]^{-1}.$$

7. En déduire une expression de

$$\lim_{n \rightarrow \infty} n \mathbb{V}(\hat{\mu}(\mathbf{b}_n^*))$$

en fonction de la densité spectrale du processus X .

8. Comparer $\lim_{n \rightarrow \infty} n \mathbb{V}(\bar{Y}_n)$ et $\lim_{n \rightarrow \infty} n \mathbb{V}(\hat{\mu}(\mathbf{b}_n^*))$. Qu'en conclure ?

9.3.4. Soit (Z_t) une suite de variables m -dépendantes telles que $\mathbb{E}[Z_t] = 0$, $\mathbb{E}[Z_t^2] = \sigma^2 < \beta < \infty$ et $\mathbb{E}[|Z_t|^{2+2\delta}] \leq \beta^{2+2\delta}$ pour $\delta > 0$. Posons

$$V_t = \mathbb{E}[Z_{t+m}^2] + 2 \sum_{j=1}^m \mathbb{E}[Z_{t+m-j} Z_{t+m}].$$

On suppose que la limite $\lim_{p \rightarrow \infty} p^{-1} \sum_{j=1}^p V_{t+j}$ existe et est égale à V . Montrer que $n^{-1/2} \sum_{t=1}^n Z_t$ converge vers une variable gaussienne centrée de variance V .

9.3.5. Soit (X_t) un processus linéaire fort $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, où (Z_t) est un IID($0, \sigma^2$). Soit (c_t) une suite déterministe satisfaisant

1. $\sum_{t=1}^{\infty} c_t^2 = \infty$,
2. $\sup_{t \geq 0} |c_t| < \infty$,
3. $\lim_{n \rightarrow \infty} \left(\sum_{t=1}^n c_t^2 \right)^{-1} \sum_{t=1}^{n-h} c_t c_{t+h} = g(h)$, pour $h \in \mathbb{Z}$.

Supposons que $V = \sum_{h=-\infty}^{\infty} g(h) \gamma(h) \neq 0$, où $\gamma(h) = \text{Cov}(X_h, X_0)$.

Pour $m \in \mathbb{N}$, on pose $X_t^m = \sum_{j=0}^m \psi_j Z_{t-j}$ et, pour $t \in \{1, \dots, n\}$,

$$b_{n,m,t} = \left(\sum_{t=1}^n c_t^2 \right)^{-1/2} \sum_{j=0}^m \psi_j c_{t+j}.$$

On pose $B_{n,m} = \sum_{t=1}^n b_{n,m,t} Z_t$.

1. Montrer que $\left(\sum_{t=1}^n c_t^2 \right)^{-1/2} \sum_{t=1}^n c_t X_t = B_{n,m} + R_{n,m}$, où pour tout $\epsilon > 0$,

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(|R_{n,m}| > \epsilon) = 0.$$

2. Montrer que $\lim_{n \rightarrow \infty} \sum_{t=1}^n b_{n,m,t}^2 = V_m$, où $V_m \stackrel{\text{def}}{=} \sum_{h=-m}^m g(h) \gamma_m(h)$ avec $\gamma_m(h) = \text{Cov}(X_0^m, X_h^m)$.

3. Montrer que $B_{n,m} \xrightarrow{\mathcal{L}} \mathcal{N}(0, V_m)$.

4. En déduire que

$$\left(\sum_{t=1}^n c_t^2 \right)^{-1/2} \sum_{t=1}^n c_t X_t \xrightarrow{\mathcal{L}} \mathcal{N}(0, V).$$

9.3.6. Soient $\lambda \in]0, \pi[$ et (X_t) un processus linéaire fort $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, où (Z_t) est un IID($0, \sigma^2$).

1. Montrer que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \cos^2(t\lambda) = 1/2,$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \sin^2(t\lambda) = 1/2,$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \cos(t\lambda) \sin(t\lambda) = 0.$$

Soient $d_{n,X}^R(\lambda)$ et $d_{n,X}^I(\lambda)$ les parties réelles et imaginaires de la transformée de Fourier à temps discret de $\{X_1, \dots, X_n\}$ à la fréquence λ ,

$$d_{n,X}^R(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n X_t \cos(t\lambda) \quad \text{et} \quad d_{n,X}^I(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n X_t \sin(t\lambda) .$$

2. Montrer (en utilisant les résultats de l'exercice 9.3.5) que le vecteur $[d_{n,X}^R(\lambda), d_{n,X}^I(\lambda)]$ converge en loi vers une variable gaussienne dont on déterminera la matrice de covariance V .
3. En déduire la distribution asymptotique que périodogramme

$$I_{n,X}(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^n \right|^2$$

9.3.7. On cherche ici à déterminer la moyenne d'un processus en présence de données manquantes. Soient $(C_t)_{t \in \mathbb{Z}}$ une suite de variables de Bernoulli indépendantes de paramètres p et (X_t) un processus linéaire fort $X_t = \mu + \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, où (Z_t) est un IID($0, \sigma^2$). On suppose que les processus (Z_t) et (X_t) sont indépendants et on considère le processus $Y_t = C_t X_t$.

1. Montrer que le processus $(Y_t)_{t \in \mathbb{Z}}$ est stationnaire au sens strict. Calculer sa moyenne et sa fonction d'autocovariance. du processus $(Y_t)_{t \in \mathbb{Z}}$.
2. Montrer que l'estimateur

$$\hat{\mu}_n = \frac{\sum_{t=1}^n C_t X_t}{\sum_{t=1}^n C_t} ,$$

est une suite d'estimateurs consistants de la moyenne μ_n .

3. On suppose que (X_t) est un processus AR(1), $X_t = \phi X_{t-1} + Z_t$ où $(Z_t)_{t \in \mathbb{Z}}$ est un IID($0, \sigma^2$). Comparer en fonction de la proportion des données manquantes les distributions de l'estimateur de moyenne empirique $n^{-1} \sum_{t=1}^n X_t$ et de $(\sum_{t=1}^n c_t)^{-1} \sum_{t=1}^n c_t X_t$

Chapitre 10

Estimation de la covariance

10.1 Estimation des coefficients d'autocovariance et d'autocorrélation

Considérons à nouveau un processus $\{X_t\}$ stationnaire au second ordre, de moyenne μ et de fonction d'autocovariance $\gamma(h)$ supposée de module sommable. Pour estimer la suite $\gamma(h)$, nous considérons les estimateurs, dits de covariances empiriques, définis par :

$$\hat{\gamma}_n(h) = \begin{cases} n^{-1} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \hat{\mu}_n)(X_t - \hat{\mu}_n) & \text{si } |h| \leq n-1 \\ 0 & \text{sinon} \end{cases} \quad (10.1)$$

où $\hat{\mu}_n = n^{-1} \sum_{t=1}^n X_t$. Remarquons que le nombre d'observations, dont nous disposons, étant précisément égal à n , il n'existe pas de paires d'observations séparées de plus de $n-1$ intervalles de temps et donc l'expression (10.1) ne permet pas d'estimer les valeurs de $\gamma(h)$ pour $|h| \geq n$. De plus, lorsque $|h|$ est proche de n , il est clair que l'estimateur (10.1) de la covariance n'est pas fiable, dans la mesure où on ne dispose que de peu de paires d'observations $(X_t, X_{t+|h|})$, ce qui implique que l'effet de moyennage statistique ne peut pas jouer. La partie la plus utile de la fonction d'autocovariance empirique est celle qui correspond aux valeurs du décalage h significativement plus faibles que le nombre d'observations n .

Certains auteurs remplacent n^{-1} par $(n-|h|)^{-1}$ dans l'expression (10.1). L'inconvénient de cette normalisation (qui peut sembler plus naturelle car on divise la somme par le nombre de termes intervenant dans celle-ci) est que la suite d'estimateurs des coefficients d'autocovariance ainsi définie n'est pas de type positif. Ajouté au fait que ces deux estimateurs sont asymptotiquement équivalents, l'utilisation de la normalisation $(n-|h|)^{-1}$ ne présente que peu d'intérêt pour les séries temporelles et n'est que très rarement utilisée en pratique.

Une propriété importante de cet estimateur est que la suite $\hat{\gamma}_n(h)$ est de type positif. En effet, si on définit le *périodogramme* par¹

$$I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^n (X_t - \hat{\mu}_n) e^{-it\lambda} \right|^2 \quad (10.2)$$

Par construction, $\lambda \mapsto I_n(\lambda)$ est une fonction positive pour $\lambda \in [-\pi, \pi]$. Par ailleurs,

$$\int_{-\pi}^{\pi} e^{i\lambda h} I_n(\lambda) d\lambda = \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n (X_t - \hat{\mu}_n)(X_s - \hat{\mu}_n) \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\lambda(h-t+s)} d\lambda = \hat{\gamma}_n(h)$$

1. Le périodogramme joue un rôle fondamental pour l'estimation de la densité spectrale étudiée dans le chapitre ??.

Par conséquent, d'après le théorème d'Herglotz 3.3.1, la suite $\hat{\gamma}_n(h)$ est de type positif.

Proposition 10.1.1. *Si $\hat{\gamma}_n(0) > 0$ alors, pour tout $p \leq n$, la matrice $\hat{\Gamma}_{n,p}$ définie par*

$$\hat{\Gamma}_n = \begin{bmatrix} \hat{\gamma}_n(0) & \hat{\gamma}_n(1) & \cdots & \hat{\gamma}_n(p-1) \\ \hat{\gamma}_n(1) & \hat{\gamma}_n(0) & \cdots & \hat{\gamma}_n(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}_n(p-1) & \hat{\gamma}_n(p-2) & \cdots & \hat{\gamma}_n(0) \end{bmatrix} \quad (10.3)$$

est de rang plein et est donc inversible.

Preuve. La suite $\hat{\gamma}_n(h)$ est de type positif, $\hat{\gamma}_n(0) > 0$ et $\hat{\gamma}_n(h)$ tend vers 0 quand n tend vers l'infini. On en déduit, d'après la propriété 3.3.9, que, pour tout p , la matrice est inversible. \square

L'estimateur d'autocovariance empirique est asymptotiquement sans biais est consistance, si nous supposons que le processus $\{X_t\}$ est linéaire au sens fort.

Théorème 10.1.2. *Soit $\{X_t, t \in \mathbb{Z}\}$ un processus linéaire fort défini par $X_t = \mu + \sum_{s=-\infty}^{\infty} \psi_s Z_{t-s}$ avec $\sum_s |\psi_s| < \infty$. On suppose que $Z_t \sim \text{IID}(0, \sigma^2)$ vérifie $\mathbb{E}[Z_t^4] = \eta\sigma^4$. Alors, pour tout $(p, q) \in \mathbb{N} \times \mathbb{N}$,*

$$\mathbb{E}[\hat{\gamma}_n(p)] = \gamma(p) + O(n^{-1}), \quad (10.4)$$

$$\lim_{n \rightarrow \infty} n \text{Cov}(\hat{\gamma}_n(p), \hat{\gamma}_n(q)) = V_{p,q}, \quad (10.5)$$

où

$$V_{p,q} \stackrel{\text{def}}{=} (\eta - 3)\gamma(p)\gamma(q) + \sum_{-\infty}^{\infty} [\gamma(u)\gamma(u-p+q) + \gamma(u+q)\gamma(u-p)]. \quad (10.6)$$

Les coefficients d'autocovariance empiriques interviennent quasiment dans tous les problèmes d'inférence statistique portant sur les processus stationnaires. A l'instar de la moyenne empirique, il est donc indispensable de disposer de résultats concernant leur distribution. Cependant, même pour les modèles de processus les plus simples, il est en général impossible de déterminer la distribution exacte de la suite des coefficients d'autocovariance empiriques $\hat{\gamma}_n(0), \dots, \hat{\gamma}_n(K)$ pour un nombre d'échantillons n donné. Nous ne considérons ici que des résultats asymptotiques concernant la distribution limite jointe des coefficients d'autocovariance empirique $[\hat{\gamma}_n(0), \dots, \hat{\gamma}_n(K)]$, pour un nombre de retard K fixé et lorsque le nombre d'échantillons n tend vers l'infini.

Théorème 10.1.3. *Soit $\{X_t, t \in \mathbb{Z}\}$ un processus linéaire défini par $X_t = \mu + \sum_{s=-\infty}^{\infty} \psi_s Z_{t-s}$ avec $\sum_s |\psi_s| < \infty$. On suppose que $Z_t \sim \text{IID}(0, \sigma^2)$ vérifie $\mathbb{E}[Z_t^4] = \eta\sigma^4$. Pour $K \geq 1$, notons $\hat{\gamma}_n \stackrel{\text{def}}{=} [\hat{\gamma}_n(1), \dots, \hat{\gamma}_n(K)]^T$, $\gamma \stackrel{\text{def}}{=}$*

$[\gamma(1), \dots, \gamma(K)]^T$ et V la matrice de dimension $K \times K$ dont l'élément $V_{p,q}$ est donné par (10.6). Alors,

$$\sqrt{n}(\hat{\gamma}_n - \gamma) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V).$$

La preuve de ce résultat est donnée dans le paragraphe 10.2.

Il est aussi intéressant, et souvent plus pratique, de considérer la distribution limite des coefficients d'autocorrélations $[\hat{\rho}_n(0), \dots, \hat{\rho}_n(K)]$. On rappelle que les coefficients d'autocorrélation sont définis par

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

et qu'ils vérifient $|\rho(h)| \leq \rho(0) = 1$. On définit les coefficients d'autocorrélation empiriques par

$$\hat{\rho}_n(h) = \frac{\hat{\gamma}_n(h)}{\hat{\gamma}_n(0)} \quad (10.7)$$

où $\hat{\gamma}_n(h)$ est donné par (10.1).

Théorème 10.1.4. Soit $\{X_t\}$ un processus linéaire défini par $X_t - \mu = \sum_{s=-\infty}^{\infty} \psi_s Z_{t-s}$ avec $\sum_s |\psi_s| < \infty$. On suppose que $Z_t \sim \text{IID}(0, \sigma^2)$ vérifie $\mathbb{E}[Z_t^4] < \infty$. Pour $K \geq 1$, on note $\hat{\rho}_n = (\hat{\rho}_n(1), \dots, \hat{\rho}_n(K))^T$, $\rho = (\rho(1), \dots, \rho(K))^T$ et $W = [w_{p,q}, p, q = 1, \dots, K]$ la matrice de dimension $K \times K$ définie, pour $1 \leq p, q \leq K$, par l'élément :

$$w_{p,q} \stackrel{\text{def}}{=} \sum_{u=-\infty}^{\infty} \{\rho(u+p) + \rho(u-p) - 2\rho(u)\rho(p)\} \\ \times \{\rho(u+q) + \rho(u-q) - 2\rho(u)\rho(q)\}. \quad (10.8)$$

Alors :

$$\sqrt{n}(\hat{\rho}_n - \rho) \rightarrow_d \mathcal{N}(0, W) \quad (10.9)$$

Il est remarquable de noter que la distribution des coefficients d'autocorrélation ne dépend pas des moments du processus Z_t (on a uniquement supposé que $Z_t \sim \text{IID}(0, \sigma^2)$ avec un moment du 4ème ordre fini). Comme dans le cas du théorème 9.2.6, on constate qu'il est nécessaire d'admettre des hypothèses relativement fortes pour garantir ce résultat. La preuve est donnée dans le paragraphe 10.2.

Exemple 10.1.5 (Bruit blanc fort). Soit $\{X_t\} \sim \text{IID}(0, \sigma^2)$. Dans ce cas $\rho(h) = 0$ pour tout $h \neq 0$ et la matrice de covariance asymptotique W est égale à la matrice identité. L'expression (10.9) montre que, lorsque la taille de l'échantillon n tend vers l'infini, le vecteur des coefficients d'autocorrélation empiriques multiplié par la racine carrée du nombre d'échantillons

$\sqrt{n}[\hat{\rho}_n(1), \dots, \hat{\rho}_n(K)]$ converge vers un vecteur gaussien, centré et de covariance identité. On en déduit que, si $\{X_t\}$ est un bruit blanc fort, pour tout $h \neq 0$:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[-1.96n^{-1/2} \leq \hat{\rho}_n(h) \leq 1.96n^{-1/2} \right] = 0.95 \quad (10.10)$$

Ce résultat peut être utilisé pour définir des tests asymptotiques de l'hypothèse nulle H_0 : " $\{X_t\}$ est un bruit blanc fort". Considérons en effet la procédure de test consistant à accepter l'hypothèse nulle si $\hat{\rho}_n(h)$ appartient à l'intervalle $[-1.96n^{-1/2}, 1.96n^{-1/2}]$ et à la rejeter sinon est un test de l'hypothèse nulle dont l'erreur de première espèce tend, lorsque $n \rightarrow \infty$, vers 0,05. Nous avons représenté figure 10.1 les 60 premiers coefficients d'auto-

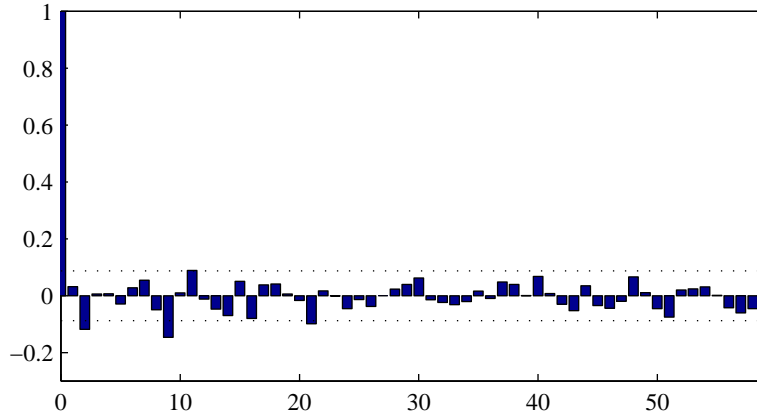


Fig. 10.1 Fonction d'autocorrélation empirique pour un échantillon de bruit blanc fort, gaussien, centré, de variance $\sigma^2 = 1$. Le nombre d'échantillons est égal à $n = 500$. Les droites en pointillé représentent les extrémités des intervalles $[-1.96n^{-1/2}, 1.96n^{-1/2}]$.

corrélations empiriques d'un échantillon de taille $n = 500$, d'un bruit blanc fort, gaussien, centré, de variance $\sigma^2 = 1$. En utilisant la formule (10.10), nous avons représenté l'intervalle asymptotique $[-1.96n^{-1/2}, 1.96n^{-1/2}]$ autour de la vraie valeur $\rho(h) = 0$. Pour les valeurs des retards h pour lesquelles l'autocorrélation empirique appartient à l'intervalle $[-1.96n^{-1/2}, 1.96n^{-1/2}]$ le test d'hypothèse décrit ci-dessus est accepté.

Ce type de visualisation où l'on représente les coefficients d'autocorrélation empiriques ainsi que les extrémités des intervalles $[-1.96n^{-1/2}, 1.96n^{-1/2}]$ pour les estimateurs correspondants dans le cas du bruit blanc (fort) est classique dans le domaine des séries temporelles où il est désigné sous le nom de corrélogramme. Il permet de détecter visuellement les retards pour lesquels

les coefficients de corrélation sont ou ne sont pas compatibles avec l'hypothèse de bruit blanc fort (comme dans le cas de la figure 10.2 par exemple).

Il faut toutefois faire attention quand on considère simultanément plusieurs valeurs de retards, car si le risque de première espèce de chaque test individuel est asymptotique égal à 0,05, il est difficile d'évaluer le risque d'un test consistant à prendre en considération simultanément un ensemble de valeurs de retards, car nous ne prenons pas en compte la dépendance entre les différents tests.

Nous pouvons toutefois déduire du théorème précédent un test de l'hypothèse nulle que le processus soit un bruit blanc en prenant en compte simultanément plusieurs valeurs de retard. En effet, le théorème 10.1.4 montre que, sous l'hypothèse que $\{X_t, t \in \mathbb{Z}\}$ est un bruit blanc fort, la statistique $T_n = \sum_{l=1}^K \hat{\rho}_n(l)^2$ est distribuée suivant une loi du χ^2 centré à K degrés de liberté. Le test consistant à accepter l'hypothèse nulle si la valeur prise par la statistique T_n est inférieure au quantile à 95% de la loi du χ^2 centré à K degrés de liberté, a une erreur de première espèce asymptotique égale à 0,05.

Exemple 10.1.6 (Processus MA(1)). On considère le processus MA(1) défini par $X_t = Z_t + \theta_1 Z_{t-1}$ où Z_t est un bruit blanc fort, centré, de variance σ^2 . Ici, la suite des coefficients d'autocorrélation est donnée par :

$$\rho(h) = \begin{cases} 1 & \text{pour } h = 0 \\ \frac{\theta_1}{1 + \theta_1^2} & \text{pour } |h| = 1 \\ 0 & \text{pour } |h| \geq 2 \end{cases}$$

On en déduit, d'après (10.8), que les éléments diagonaux de la matrice de covariance de la distribution limite des coefficients d'autocovariance empiriques ont pour expression :

$$W_{h,h} = \begin{cases} 1 - 3\rho^2(1) + 4\rho^4(1) & \text{pour } |h| = 1 \\ 1 + 2\rho(1)^2 & \text{pour } |h| \geq 2 \end{cases}$$

Par conséquent les intervalles de confiance de taux de couverture 95% pour les coefficients d'autocorrélation empiriques sont donnés, pour $h = 1$, par :

$$\hat{\rho}_n(1) \in \left[\rho(1) - 1.96W_{1,1}^{1/2} n^{-1/2} \quad \rho(1) + 1.96W_{1,1}^{1/2} n^{-1/2} \right]$$

et, pour $h \geq 2$, par :

$$\hat{\rho}_n(h) \in \left[-1.96W_{2,2}^{1/2} n^{-1/2} \quad + 1.96W_{2,2}^{1/2} n^{-1/2} \right]$$

Notons ici que ces régions dépendent, par l'intermédiaire de $\rho(1)$, de la quantité a priori inconnue θ_1 . Nous avons représenté figure 10.2 les 60 premiers co-

efficients d'autocorrélation empiriques d'un échantillon de longueur $n = 500$ d'un processus MA(1) défini par $\theta_1 = -0.8$ et $\sigma = 1$. Les traits en pointillé représentent les bornes asymptotiques autour des vraies valeurs au niveau 95%.

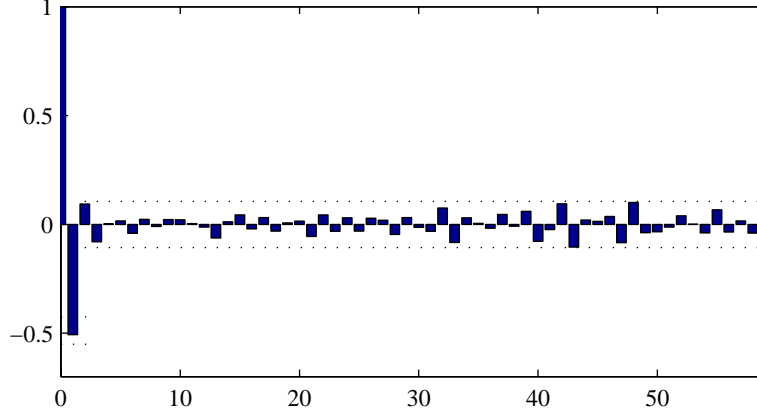


Fig. 10.2 Fonction d'autocorrélation empirique d'un échantillon de longueur $n = 500$ d'un processus MA(1) pour $\theta_1 = -0.8$ et donc $\rho(1) = -0.4878$.

Exemple 10.1.7 (Processus autorégressif fort d'ordre 1). On considère le processus aléatoire X_t défini par :

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

où $\{Z_t\} \sim \text{IID}(0, \sigma^2)$ et où $|\phi| < 1$. La fonction d'autocorrélation d'un tel processus est donnée par $\rho(h) = \phi^{|h|}$ et les éléments diagonaux de la matrice de covariance W sont donnés par

$$\begin{aligned} W_{h,h} &= \sum_{m=1}^h \phi^{2h} (\phi^{-m} - \phi^m)^2 + \sum_{m=h+1}^{\infty} \phi^{2m} (\phi^{-i} - \phi^i)^2 \\ &= (1 - \phi^{2h})(1 + \phi^2)(1 - \phi^2)^{-1} - 2h\phi^{2h} \end{aligned}$$

Considérons la séquence, de longueur $n = 1800$, des battements cardiaques représentés figure 1.1 (chapitre 1). La figure 3.2 qui représente les couples (X_t, X_{t-1}) suggère fortement la présence d'une relation linéaire entre les variables X_t et X_{t-1} et invite donc à tester la validité d'un modèle autorégressif d'ordre 1. Pour estimer le paramètre ϕ du modèle autorégressif, une méthode naturelle, compte tenu de l'allure de la fonction d'autocorrélation de l'AR(1),

consiste à utiliser comme estimateur $\hat{\phi}_n = \hat{\rho}_n(1)$ qui donne $\hat{\phi}_n = 0.966$. Pour tester la validité du modèle, deux solutions s'offrent à nous : (i) tester que les résidus de prédiction donnés par $\hat{Z}_t = X_t - \hat{\mu}_n - \hat{\phi}_n(X_{t-1} - \hat{\mu}_n)$ sont compatibles avec un modèle de bruit blanc, (ii) vérifier directement que les coefficients d'autocorrélation empiriques sont compatibles avec ceux d'un modèle AR(1). Les résidus de prédiction sont reportés figure 10.3 et la fonction d'autocorrélation de ces résidus figure 10.4, où nous avons aussi indiqué les bornes de la zone crédible à 95% pour le bruit blanc avec un nombre d'observations $n = 1800$. Les corrélations empiriques, en particulier pour $h = 2$, sont significativement à l'extérieur des intervalles de confiance du bruit blanc, ce qui conduit à rejeter le modèle de bruit blanc pour les résidus et donc le modèle autorégressif d'ordre 1 pour les observations. Les résultats de l'analyse de la suite des coefficients d'autocorrélation empiriques du processus et des zones crédibles à 95% sous l'hypothèse d'un modèle AR(1) avec $\phi = 0.966$ sont reportés figure 10.5. On observe que les premières valeurs des coefficients de corrélation sont nettement à l'extérieur de cette zone, ce qui contribue ici encore à rejeter le modèle AR(1).

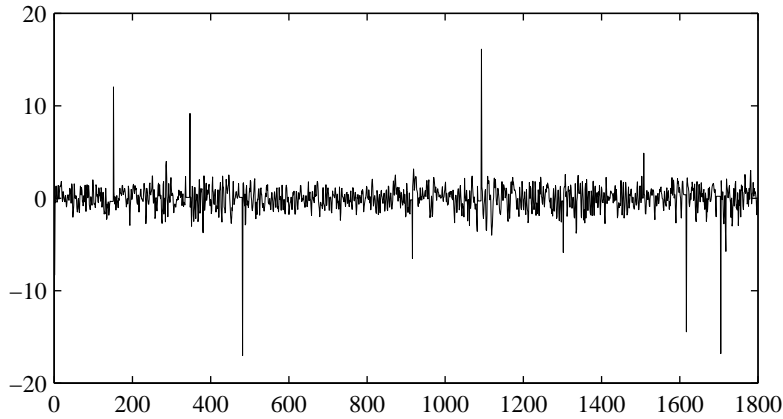


Fig. 10.3 *Série des battements cardiaques : Résidu de prédiction $\hat{Z}_t = (X_t - \hat{\mu}_n) - \hat{\phi}_n(X_{t-1} - \hat{\mu}_n)$.*

10.2 Preuves des Théorèmes 10.1.4 et 10.1.3

Nous allons maintenant étudier les distributions asymptotique de la fonction d'autocovariance et d'autocorrélation. Dans toute cette partie, nous sup-

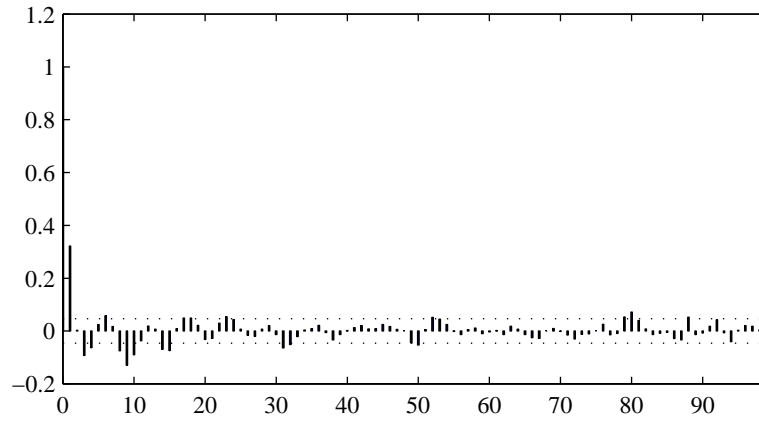


Fig. 10.4 Série des battements cardiaques : coefficients d'autocorrélation empiriques des résidus de prédiction $\tilde{Z}_t = (X_t - \hat{\mu}_n) - \hat{\phi}_n(X_{t-1} - \hat{\mu}_n)$ et zones crédibles à 95% pour le bruit blanc ($n = 1800$).

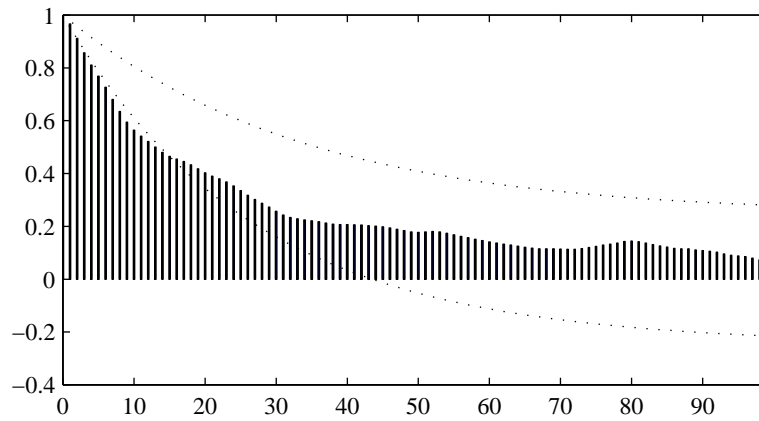


Fig. 10.5 Série des battements cardiaques : coefficients d'autocorrélation empiriques de la série et bornes des zones crédibles à 95% pour un modèle $AR(1)$ de paramètre $\phi = 0.966$.

posons que $\{X_t, t \in \mathbb{Z}\}$ est un processus linéaire défini par

$$X_t = \mu + \sum_{s=-\infty}^{\infty} \psi_s Z_{t-s}, \quad \sum_s |\psi_s| < \infty, \quad (10.11)$$

où

$$Z_t \sim \text{IID}(0, \sigma^2) \text{ vérifie } \mathbb{E}[Z_t^4] = \eta\sigma^4. \quad (10.12)$$

Sans perte de généralité, nous posons $\mu = 0$. Nous notons $\{\gamma(h), h \in \mathbb{Z}\}$ la fonction d'autocovariance de ce processus,

$$\gamma(h) = \sigma^2 \sum_{i=-\infty}^{\infty} \psi_i \psi_{i+h}, \quad h \in \mathbb{Z}. \quad (10.13)$$

Considérons, pour $h \geq 0$,

$$\tilde{\gamma}_n(h) = n^{-1} \sum_{t=1}^n (X_{t+h} - \mu)(X_t - \mu). \quad (10.14)$$

Cette quantité est plus facile à étudier que l'autocovariance empirique $\hat{\gamma}_n(h)$. De plus, $\tilde{\gamma}_n(h)$ est une *bonne* approximation de l'autocovariance empirique comme le démontre la Proposition 10.2.2. Nous aurons besoin dans la suite du résultat suivant :

Lemme 10.2.1. *Supposons que $\{X_t\}$ satisfait (10.11) et (10.12). Alors, pour tout $(k, l, p, q) \in \mathbb{Z}^4$*

$$\begin{aligned} \mathbb{E}[X_k X_\ell X_p X_q] &= (\eta - 3)\sigma^4 \sum_{i=-\infty}^{\infty} \psi_{k+i} \psi_{\ell+i} \psi_{p+i} \psi_{q+i} \\ &+ \sigma^4 \{ \gamma(k - \ell)\gamma(p - q) + \gamma(k - p)\gamma(\ell - q) + \gamma(k - q)\gamma(\ell - p) \}. \end{aligned} \quad (10.15)$$

De plus, il existe une constante C telle que, pour tout $m \in \mathbb{N}$,

$$\mathbb{E} \left[\left(\sum_{t=1}^m X_t \right)^4 \right] \leq C m^2. \quad (10.16)$$

Preuve. Ce calcul est élémentaire mais un peu compliqué ; nous n'en donnons que les éléments essentiels en laissant les détails au lecteur. Nous évaluons ensuite $\mathbb{E}[Z_i Z_j Z_k Z_\ell]$. Un calcul élémentaire montre que

$$\mathbb{E}[Z_i Z_j Z_k Z_\ell] = \begin{cases} \eta\sigma^4 & \text{si } i = j = k = \ell \\ \sigma^4 & \text{si } i = j \neq k = \ell \text{ ou } i = k \neq j = \ell \text{ ou } i = \ell \neq j = k \\ 0 & \text{sinon} \end{cases} \quad (10.17)$$

Le résultat (10.15) en découle.

$$\mathbb{E} \left[\left(\sum_{t=1}^m X_t \right)^4 \right] = \sum_{k,\ell,p,q=1}^m \mathbb{E} [X_k X_\ell X_p X_q] = A + B$$

avec

$$A = (\eta - 3)\sigma^4 \sum_{k,\ell,p,q=1}^m \sum_{i=-\infty}^{\infty} \psi_{k+i} \psi_{\ell+i} \psi_{p+i} \psi_{q+i}$$

$$B = \sigma^4 \sum_{k,\ell,p,q=1}^m \{ \gamma(k-\ell)\gamma(p-q) + \gamma(k-p)\gamma(\ell-q) + \gamma(k-q)\gamma(\ell-p) \}.$$

Notons tout d'abord que

$$\begin{aligned} \sum_{k,\ell,p,q=1}^m \left| \sum_{i=-\infty}^{\infty} \psi_{k+i} \psi_{\ell+i} \psi_{p+i} \psi_{q+i} \right| &\leq \sum_{k=1}^n \sum_{\ell,p,q=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} |\psi_{k+i} \psi_{\ell+i} \psi_{p+i} \psi_{q+i}| \\ &\leq \sum_{k=1}^m \sum_{i=-\infty}^{\infty} |\psi_{k+i}| \left(\sum_{j=-\infty}^{\infty} |\psi_j| \right)^3, \end{aligned}$$

ce qui montre que

$$A \leq m \left(\sum_{j=-\infty}^{\infty} |\psi_j| \right)^4.$$

Remarquons d'autre part que

$$\begin{aligned} \sum_{k,\ell,p,q=1}^m |\gamma(k-\ell)\gamma(p-q) + \gamma(k-p)\gamma(\ell-q) + \gamma(k-q)\gamma(\ell-p)| \\ \leq 3 \left(\sum_{h=-m+1}^{m-1} (m-|h|)|\gamma(h)| \right)^2. \end{aligned}$$

Par conséquent

$$B \leq 3m^2 \left(\sum_{h=-\infty}^{\infty} |\gamma(h)| \right)^2.$$

□

Proposition 10.2.2. *Supposons que $\{X_t\}$ satisfait (10.11) et (10.12). Pour tout $h \geq 0$,*

$$\mathbb{E} \left[(\hat{\gamma}_n(h) - \tilde{\gamma}_n(h))^2 \right] = O(n^{-2}). \quad (10.18)$$

De plus, pour tout $h \in \mathbb{Z}$

$$\mathbb{E} [\hat{\gamma}_n(h)] = \mathbb{E} [\tilde{\gamma}_n(h)] + O(n^{-1}), \quad (10.19)$$

et pour tout $(p, q) \in \mathbb{Z} \times \mathbb{Z}$,

$$\text{Cov}(\hat{\gamma}_n(p), \hat{\gamma}_n(q)) = \text{Cov}(\tilde{\gamma}_n(p), \tilde{\gamma}_n(q)) + O(n^{-1}). \quad (10.20)$$

Preuve. Sans perte de généralité, nous posons $\mu = 0$. Un calcul élémentaire montre que

$$\hat{\gamma}_n(h) = \frac{1}{n} \sum_{t=1}^{n-h} X_{t+h} X_t + \frac{n-h}{n} \bar{X}_n^2 - \frac{1}{n} \bar{X}_n \sum_{t=1}^{n-h} \{X_t + X_{t+h}\}.$$

Nous avons d'autre part

$$n^{-1} \sum_{t=1}^{n-h} \{X_t + X_{t+h}\} = 2\bar{X}_n - n^{-1} \sum_{t=1}^h X_t - n^{-1} \sum_{t=n-h+1}^n X_t.$$

Par conséquent

$$\hat{\gamma}_n(h) = \tilde{\gamma}_n(h) - \left(1 + \frac{h}{n}\right) \bar{X}_n^2 - \bar{X}_n \left[n^{-1} \sum_{t=1}^h X_t + n^{-1} \sum_{t=n-h+1}^n X_t \right].$$

En utilisant le Lemme 10.2.1, nous avons $\mathbb{E} [\bar{X}_n^4] \leq Cn^{-2}$ et

$$\mathbb{E} \left[\left(\bar{X}_n n^{-1} \sum_{t=1}^h X_t \right)^2 \right] \leq n^{-2} (\mathbb{E} [\bar{X}_n^4])^{1/2} \left(\mathbb{E} \left[\left(\sum_{t=1}^h X_t \right)^4 \right] \right)^{1/2} \leq Cn^{-3}h.$$

De même,

$$\mathbb{E} \left[\left(\bar{X}_n n^{-1} \sum_{t=n-h+1}^n X_t \right)^2 \right] \leq Cn^{-3}h,$$

ce qui conclut la preuve de (10.18). L'équation (10.19) découle de

$$|\mathbb{E} [\hat{\gamma}_n(h)] - \mathbb{E} [\tilde{\gamma}_n(h)]| \leq \mathbb{E} [|\hat{\gamma}_n(h) - \tilde{\gamma}_n(h)|] \leq (\mathbb{E} [|\hat{\gamma}_n(h) - \tilde{\gamma}_n(h)|^2])^{1/2}.$$

L'équation (10.20) s'obtient de façon similaire. \square

Nous allons tout d'abord calculer la variance et la variance asymptotique de $\tilde{\gamma}_n(h)$ lorsque $\{X_t, t \in \mathbb{Z}\}$ est un processus linéaire fort vérifiant (10.11) et (10.12).

Proposition 10.2.3. *Soit $\{X_t, t \in \mathbb{Z}\}$ un processus linéaire vérifiant (10.11) et (10.12). Alors :*

$$\begin{aligned} \lim_{n \rightarrow \infty} n \text{Cov}(\tilde{\gamma}_n(p), \tilde{\gamma}_n(q)) &= (\eta - 3)\gamma(p)\gamma(q) \\ &+ \sum_{u=-\infty}^{\infty} [\gamma(u)\gamma(u+p-q) + \gamma(u+p)\gamma(u-q)] . \end{aligned} \quad (10.21)$$

Preuve. Sans perte de généralité, nous posons ici $\mu = 0$. Notons que $\mathbb{E}[\tilde{\gamma}_n(h)] = \gamma(h)$. Nous allons vérifier tout d'abord que pour tout $p, q \geq 0$,

$$\text{Cov}(\tilde{\gamma}_n(p), \tilde{\gamma}_n(q)) = n^{-1} \sum_{u=-(n-1)}^{(n-1)} \left(1 - \frac{|u|}{n}\right) V_u , \quad (10.22)$$

où

$$\begin{aligned} V_u &\stackrel{\text{def}}{=} \gamma(u)\gamma(u+p-q) + \gamma(u+p)\gamma(u-q) \\ &+ (\eta - 3)\sigma^4 \sum_i \psi_{i+u+q}\psi_{i+u}\psi_{i+p}\psi_i . \end{aligned} \quad (10.23)$$

L'absolue sommabilité de $\{\psi_t, t \in \mathbb{Z}\}$ implique que $\sum_{u \in \mathbb{Z}} |V_u| < \infty$ et le théorème de convergence dominé implique (10.21). Notons tout d'abord que :

$$\mathbb{E}[\tilde{\gamma}_n(p)\tilde{\gamma}_n(q)] = n^{-2} \sum_{s,t} \sum_{i,j,k,\ell} \psi_{s+p-i}\psi_{s-j}\psi_{t+q-k}\psi_{t-\ell} \mathbb{E}[Z_i Z_j Z_k Z_\ell] .$$

La formule s'obtient ensuite directement en utilisant (10.17). \square

Nous disposons maintenant de l'ensemble des résultats permettant d'établir le Théorème 10.1.2

Preuve (Preuve du Théorème 10.1.2). Le théorème découle de (10.19), (10.20) et de la Proposition 10.2.3. \square

La proposition 10.2.2 montre que

$$n^{1/2}(\hat{\gamma}_n(h) - \tilde{\gamma}_n(h)) = o_P(1) .$$

Par conséquent, les distributions limites de $\tilde{\gamma}_n(h)$ et de $\hat{\gamma}_n(h)$ coïncident. En utilisant la proposition précédente, nous pouvons obtenir un premier résultat sur la distribution asymptotique des coefficients d'autocovariance empirique :

Théorème 10.2.4. *Soit $\{X_t, t \in \mathbb{Z}\}$ un processus linéaire vérifiant (10.11) et (10.12). Pour $K \geq 1$, notons $\hat{\gamma}_n \stackrel{\text{def}}{=} [\hat{\gamma}_n(1), \dots, \hat{\gamma}_n(K)]^T$, $\gamma \stackrel{\text{def}}{=} [\gamma(1), \dots, \gamma(K)]^T$ et V la matrice de dimension $K \times K$ dont l'élément $V_{p,q}$ est donné par*

$$V_{p,q} \stackrel{\text{def}}{=} (\eta - 3)\gamma(p)\gamma(q) + \sum_{u=-\infty}^{\infty} [\gamma(u)\gamma(u-p+q) + \gamma(u+q)\gamma(u-p)] . \quad (10.24)$$

Alors,

$$\sqrt{n}(\hat{\gamma}_n - \gamma) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V) .$$

Preuve. Nous allons établir ce résultat en remplaçant $\hat{\gamma}_n$ par $\tilde{\gamma}_n = [\tilde{\gamma}_n(1), \dots, \tilde{\gamma}_n(K)]^T$, car nous savons que ces deux quantités sont asymptotiquement équivalentes. Nous considérons tout d'abord le processus stationnaire au sens strict $(2m + K)$ -dépendant défini par

$$Y_t^m = \begin{bmatrix} (X_t^m - \mu)^2 \\ (X_{t+1}^m - \mu)(X_t^m - \mu) \\ \vdots \\ (X_{t+K}^m - \mu)(X_t^m - \mu) \end{bmatrix} ,$$

où $X_t^m \stackrel{\text{def}}{=} \mu + \sum_{|j| \leq m} \psi_j Z_{t-j}$. La moyenne empirique de ce processus est donnée par

$$\bar{Y}_n^m = n^{-1} \sum_{t=1}^n Y_t^m = \begin{bmatrix} \tilde{\gamma}_n^m(0) \\ \tilde{\gamma}_n^m(1) \\ \vdots \\ \tilde{\gamma}_n^m(K) \end{bmatrix} ,$$

où, pour $h \geq 0$,

$$\tilde{\gamma}_n^m(h) = n^{-1} \sum_{t=1}^n (X_{t+h}^m - \mu)(X_t^m - \mu) .$$

Nous remarquons que

$$\mathbb{E} [\bar{Y}_{m,n}] = \begin{bmatrix} \gamma_n^m(0) \\ \gamma_n^m(1) \\ \vdots \\ \gamma_n^m(K) \end{bmatrix}$$

où $\gamma^m(h) = \text{Cov}(X_0^m, X_h^m)$. Nous allons appliquer le théorème d'approximation 9.2.4. Considérons le vecteur

$$B_{m,n} = n^{1/2} (\bar{Y}_{m,n} - \mathbb{E} [\bar{Y}_{m,n}]) ,$$

qui approche le vecteur

$$A_n = n^{1/2} (\hat{\gamma}_n - \gamma) .$$

1. Soit \mathbf{c} un $(K + 1) \times 1$ vecteur déterministe, et appliquons le théorème Central Limite au processus $(2m + K)$ -dépendant $\mathbf{c}^T Y_t$. Nous obtenons

$$B_{m,n} = n^{1/2} (\bar{Y}_t^m - \mathbb{E} [\bar{Y}_0^m]) \xrightarrow{\mathcal{L}} B_m$$

où B_m est une variable aléatoire gaussienne de moyenne nulle et de variance $\mathbf{c}^T V^m \mathbf{c}$, les éléments $V_{p,q}^m$ de la matrice V^m étant donnés par,

$$V_{p,q}^m \stackrel{\text{def}}{=} (\eta - 3)\gamma^m(p)\gamma^m(q) + \sum_{u=-\infty}^{\infty} [\gamma^m(u)\gamma^m(u-p+q) + \gamma^m(u+q)\gamma^m(u-p)] .$$

2. En remarquant que comme $\lim_{m \rightarrow \infty} V_m = V$, nous avons $B_m \xrightarrow{\mathcal{L}} B$ quand $m \rightarrow \infty$, où B est une variable aléatoire gaussienne de moyenne nulle et de variance $\mathbf{c}^T V \mathbf{c}$.
3. Il reste à établir l'uniformité de la convergence : pour tout $\epsilon > 0$,

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(|B_{m,n} - A_n| \geq \epsilon) = 0 .$$

Comme

$$\mathbb{P}(|B_{m,n} - A_n| \geq \epsilon) \leq \sum_{h=0}^K \mathbb{P}(n^{1/2}|\tilde{\gamma}_n^m(h) - \hat{\gamma}_n(h)| \geq \epsilon_K)$$

il suffit d'établir (en utilisant l'inégalité de Bienaymé-Tchebyshev) que

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} n\mathbb{V}(\tilde{\gamma}_n^m(h) - \hat{\gamma}_n(h)) = 0 . \quad (10.25)$$

En utilisant l'identité

$$n\mathbb{V}(\tilde{\gamma}_n^m(h) - \hat{\gamma}_n(h)) = n[\mathbb{V}(\tilde{\gamma}_n^m(h)) + n\mathbb{V}(\hat{\gamma}_n(h)) + 2\text{Cov}(\tilde{\gamma}_n^m(h), \hat{\gamma}_n(h))] ,$$

et des calculs similaires à ceux que nous avons effectué pour établir (10.22) and (10.21), nous obtenons (10.25).

□

Nous allons étendre le résultat précédent aux coefficients d'autocorrélation $\hat{\rho}_n(h) = \hat{\gamma}_n(h)/\hat{\gamma}_n(0)$.

Preuve (Preuve du Théorème 10.1.4). La preuve est une application directe du théorème 10.1.3 par application de la delta-méthode (voir paragraphe 8.6). On considère la fonction $g : \mathbb{R}^{K+1} \rightarrow \mathbb{R}^K$ définie pour $x_0 \neq 0$ par

$$g(x_0, x_1, \dots, x_K) = [x_1/x_0, \dots, x_K/x_0]^T .$$

Nous avons,

$$g(\hat{\gamma}_n(0), \dots, \hat{\gamma}_n(K)) = (\hat{\rho}_n(1), \dots, \hat{\rho}_n(K))^T .$$

et une application directe de la Proposition 8.6.1 montre que

$$\sqrt{n}(g(\hat{\gamma}_n(0), \dots, \hat{\gamma}_n(K)) - g(\gamma(0), \dots, \gamma(K))) \xrightarrow{\mathcal{L}} \mathcal{N}(0, DV D^T)$$

où V est la matrice de covariance des coefficients d'autocovariance donnée par (10.24) et D est la matrice jacobienne de la fonction g au point $[\gamma(0), \dots, \gamma(K)]$,

$$D = \frac{1}{\gamma(0)} \begin{bmatrix} -\gamma(1) & \gamma(0) & 0 & \dots & 0 \\ -\gamma(2) & 0 & \gamma(0) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\gamma(K) & 0 & 0 & \dots & \gamma(0) \end{bmatrix}.$$

Cette matrice peut se réécrire de façon plus compacte

$$D = \frac{1}{\gamma(0)} [-\boldsymbol{\rho} I_K]$$

où I_K est la matrice identité $K \times K$. En écrivant la matrice V sous la forme

$$V = \begin{bmatrix} v_{0,0} & \mathbf{v}_1^T \\ \mathbf{v}_1 & V_{2,2} \end{bmatrix}$$

la matrice W peut s'écrire

$$W = \gamma^{-2}(0) [v_0 \boldsymbol{\rho} \boldsymbol{\rho}^T - \boldsymbol{\rho} \mathbf{v}_1^T - \mathbf{v}_1 \boldsymbol{\rho}^T + V_{2,2}],$$

où $\mathbf{v}_1 = [v_{1,0}, v_{2,0}, \dots, v_{K,0}]^T$ et $V_{2,2} = [v_{p,q}, p, q = 1, \dots, K]$. Un calcul élémentaire montre que

$$\begin{aligned} w_{p,q} &= \gamma^{-2}(0) [v_{p,q} - \rho(p)v_{0,q} - \rho(q)v_{p,0} + \rho(p)\rho(q)v_{0,0}] \\ &= \sum_{u=-\infty}^{\infty} [\rho(u)\rho(u-p+q) + \rho(u-p)\rho(u+q) + 2\rho(p)\rho(q)\rho^2(u) - 2\rho(p)\rho(u)\rho(u+q) - 2\rho(q)\rho(u)\rho(u-p)] \end{aligned}$$

□

Index

- δ -méthode, 152
- π -système, 146
- équation de Riccati, 120

- Bruit blanc
 - faible, 43
 - fort, 43

- Cauchy-Schwarz (Inégalité de), 21
- Champ aléatoire
 - densité, 82
- Champ aléatoire, 53
- Champ aléatoire
 - Intégrale stochastique, 54
- Champ spectral, 56
- Convergence
 - étroite, 142, 143
 - en loi, 142

- Densité spectrale, 48

- Enveloppe Linéaire d'un processus, 56

- Famille des répartitions finies, 8
- Filtrage
 - Champ spectral, 83
 - Définition, 83
 - Inverse, 83
 - Moyenne ajustée infinie, $MA(\infty)$, 83
- Filtrage de Kalman
 - prédiction, 118
- Filtre de Kalman, 115
- Fonction d'autocovariance
 - positivité, 42
- Fonction caractéristique, 147
- Fonction d'autocovariance
 - positivité, 40

- Fonction d'autocovariance
 - symétrie hermitienne, 42
- Fonction de répartition, 146
- Fonction de transfert, 83
 - inverse, 85
- Forward-backward
 - in state-space model, 138

- Gaussian linear model, 134

- Information parameterization, 133
- Innovation, 117
 - partielle, 91

- Kalman
 - filter, 122
 - gain, 122
 - filtering with non-zero means, 123
 - predictor, 120
 - gain, 119

- Lemme de Portmanteau, 145
- Lemme de Slutsky, 151
- Likelihood
 - in state-space model, 121
- Linear prediction, 118
- Lissage, 131
 - Perturbation, 124
- Lissage arrière
 - mise à jour, 133
 - moyenne, 133
 - noyau, 133
- Loi, 9
- Loi marginale, 14

- Matrice
 - compagnon, 113

- Matrix inversion lemma, 134
- Mesure image, *voir* Loi
- Modèle d'état
 - linéaire gaussien, 129
- Modèle d'états
 - linéaire gaussien, 127
- Modèle linéaire gaussien, 129
- Non-deterministic process, 117
- Prediction lineaire
 - Prédicteur linéaire optimal, 115
- Prédiction linéaire, 115
- Procédé de Cramér–Wold, 148
- Processus
 - m -dépendant, 15, 158
 - Aléatoire, 7
 - canonique, 9
 - Champ spectral, 56
 - Enveloppe Linéaire, 56
 - Filtré, 83
 - gaussien, 12
- Image, 83
- linéaire, 67
- fort, 158
- Série chronologique, *voir* Série temporelle
- Série temporelle, 4
- Stationnarité
 - stricte, 14
- Suite bornée en probabilité, 143
- Suite uniformément tendue, *voir* Suite
 - bornée en probabilité
- Tension, 143
- théorème de Prohorov, 144
- théorème de continuité, 148
- théorème de Lévy, 147
- Trajectoire, 7
- Uniforme Intégrabilité, 153
- Variables aléatoires
 - gaussiennes, 10

Littérature

- [1] J. Jacod and P. Protter. *Probability essentials*. Universitext. Springer-Verlag, Berlin, second edition, 2003.
- [2] R. E. Kalman and R. Bucy. New results in linear filtering and prediction theory. *J. Basic Eng., Trans. ASME, Series D*, 83(3) :95–108, 1961.
- [3] H. Rauch, F. Tung, and C. Striebel. Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, 3(8) :1445–1450, 1965.
- [4] H. L. Royden. *Real analysis*. Macmillan Publishing Company, New York, third edition, 1988.