

Introduction aux séries temporelles

Justin Salez

18 décembre 2020

Table des matières

1	Cadre de travail	5
1.1	Modélisation des séries temporelles	5
1.2	Processus aléatoires	7
1.3	Stationnarité	8
2	Filtrage linéaire	11
2.1	Existence et propriétés	11
2.2	Composition de filtres	14
2.3	Représentation en série de puissances	16
2.4	Défiltrage	17
3	Processus ARMA	21
3.1	Définition et cas particuliers	21
3.2	Étude du cas général	23
3.3	Décorrélacion rapide	25
4	Analyse spectrale	29
4.1	Rappels sur les coefficients de Fourier	29
4.2	Le théorème de représentation spectrale d'Herglotz	32
4.3	Mesure spectrale d'un processus stationnaire	33
4.4	Retour sur les processus ARMA	35
5	Prédiction linéaire	37
5.1	Cadre théorique	37
5.2	Équation de Yule-Walker	39
5.3	Sur l'unicité dans l'équation de Yule-Walker	40
5.4	Algorithme de Levinson-Durbin	42
6	Estimation statistique	45
6.1	Rappels sur les estimateurs	45
6.2	Estimation de la moyenne	46
6.3	Estimation de l'auto-covariance	47

Chapitre 1

Cadre de travail

1.1 Modélisation des séries temporelles

Définition. Une **série temporelle** est une suite cohérente de valeurs numériques représentant l'évolution d'une grandeur physique à des intervalles de temps réguliers, par exemple :

1. Population mondiale
2. Prix du baril de pétrole
3. Consommation électrique
4. Trafic internet
5. Hauteurs des crues du Nil
6. Température des océans
7. Concentration atmosphérique en CO₂
8. Taux de glucose dans le sang
9. Côte de popularité du président
10. Nombre de cas de covid-19

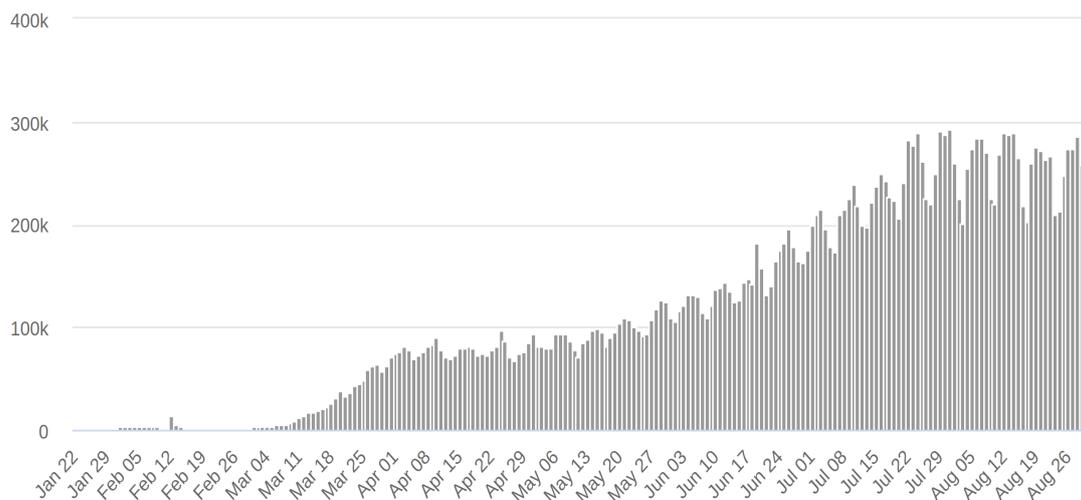


FIGURE 1.1 – Nombre de décès quotidiens mondiaux dûs au covid-19

Modèle. Dans ce cours, nous allons apprendre à modéliser de telles séries, à les analyser, et à les prédire. Pour cela, nous supposons que la série observée est la réalisation (partielle) d'un processus aléatoire $X = (X_t)_{t \in \mathbb{Z}}$ de la forme

$$X_t = m_t + s_t + Z_t, \quad (1.1)$$

où les composantes $(m_t)_{t \in \mathbb{Z}}$, $(s_t)_{t \in \mathbb{Z}}$ et $(Z_t)_{t \in \mathbb{Z}}$, bien visibles sur la Figure 1.1, sont comme suit :

1. La **tendance** $(m_t)_{t \in \mathbb{Z}}$ est une suite déterministe polynomiale, c'est-à-dire de la forme

$$m_t = a_0 + a_1 t + a_2 t^2 + \dots + a_n t^n. \quad (1.2)$$

Elle représente l'évolution moyenne lente de la grandeur considérée. Notons qu'imposer une forme polynomiale n'est pas restrictif, puisque toute suite finie peut-être réalisée par un polynôme bien choisi (interpolation). Le degré n doit simplement être pris aussi grand que nécessaire pour rendre compte de la complexité du profil observé.

2. La **saisonnalité** $(s_t)_{t \in \mathbb{Z}}$ est une suite déterministe périodique de la forme

$$s_t = \sum_{k=0}^n b_k \cos\left(\frac{2\pi kt}{d}\right) + \sum_{k=0}^n c_k \sin\left(\frac{2\pi kt}{d}\right). \quad (1.3)$$

Elle représente la composante "cyclique" du phénomène considéré. La période d sera dictée par le contexte (saisonnalité journalière, hebdomadaire, mensuelle, annuelle, etc.). Notons qu'ici encore, la forme choisie n'est pas restrictive, puisque toute suite périodique peut s'écrire comme un polynôme trigonométrique (quitte à choisir le degré n suffisamment grand pour épouser la complexité du profil observé).

3. Le **bruit** $(Z_t)_{t \in \mathbb{Z}}$ est un processus aléatoire stationnaire, c'est-à-dire dont les caractéristiques statistiques (espérance, covariance) ne varient pas dans le temps. Nous y reviendrons.

Élimination de la tendance et de la saisonnalité. L'analyse des processus de la forme (1.1) peut en réalité se réduire à celle des processus aléatoires stationnaires, grâce à l'application d'opérateurs de différentiation. En effet, pour prédire l'observation future X_t à partir des observations passées X_{t-1}, X_{t-2}, \dots , il est clair qu'il suffit de prédire la valeur de l'incrément

$$(\Delta X)_t := X_t - X_{t-1}. \quad (1.4)$$

On peut donc remplacer l'étude de X par celle de ΔX . Or la transformation linéaire $X \mapsto \Delta X$ a pour effet de diminuer le degré de la tendance de 1, tout en préservant le caractère d -périodique de la saisonnalité et le caractère stationnaire du bruit. En itérant cette transformation $n + 1$ fois, on obtient un signal dont la tendance est nulle. De la même manière, on peut ensuite éliminer la saisonnalité à l'aide de l'opérateur

$$(\Delta_d X)_t := X_t - X_{t-d}. \quad (1.5)$$

Cette transformation a pour effet d'éliminer la composante d -périodique, tout en préservant le caractère stationnaire du bruit. Après application de ces deux transformations, on se trouve ramené au problème de prédire un signal ne contenant que du bruit, c'est-à-dire un processus stationnaire. C'est donc à l'étude des processus aléatoires stationnaires que le cours sera consacré.

1.2 Processus aléatoires

Dans tout le cours, nous supposons donné un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$.

Définition 1.1 (Processus). Nous appellerons *processus aléatoire* (ou simplement *processus*) une suite doublement infinie $X = (X_t)_{t \in \mathbb{Z}}$ de variables aléatoires réelles définies sur $(\Omega, \mathcal{F}, \mathbb{P})$.

Attention : pour décrire la loi de X , il ne suffit pas de donner la loi de X_t pour chaque $t \in \mathbb{Z}$! Il faut spécifier toutes les **marginales fini-dimensionnelles**, c'est-à-dire la loi jointe des n variables $(X_{t_1}, \dots, X_{t_n})$, pour tout $n \geq 1$ et tout $(t_1, \dots, t_n) \in \mathbb{Z}^n$. En pratique, nous renoncerons à faire cela, et nous nous intéresserons seulement à deux statistiques simples mais fondamentales : les moyennes et covariances. Leur existence nécessite une hypothèse supplémentaire, que nous ferons désormais systématiquement.

Définition 1.2 (Second-ordre, moyenne et covariance). Nous dirons que le processus X est **du second ordre** si ses coordonnées $(X_t)_{t \in \mathbb{Z}}$ sont toutes dans l'espace $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Autrement dit,

$$\forall t \in \mathbb{Z}, \quad \mathbb{E}[X_t^2] < \infty. \quad (1.6)$$

On peut alors définir la **fonction moyenne** $\mu_X: \mathbb{Z} \rightarrow \mathbb{R}$ par la formule

$$\mu_X(t) := \mathbb{E}[X_t], \quad (1.7)$$

et la **fonction d'auto-covariance** $\gamma_X: \mathbb{Z}^2 \rightarrow \mathbb{R}$ par la formule

$$\gamma_X(s, t) := \text{Cov}(X_s, X_t) = \mathbb{E}[X_s X_t] - \mathbb{E}[X_s] \mathbb{E}[X_t]. \quad (1.8)$$

Les fonctions μ_X et γ_X ne caractérisent pas les marginales fini-dimensionnelles de X , mais elles en capturent les caractéristiques essentielles. Ainsi, pour tout $n \geq 1$ et tout $(t_1, \dots, t_n) \in \mathbb{Z}^n$,

1. La moyenne $m = (m_i)_{1 \leq i \leq n}$ du vecteur $(X_{t_1}, \dots, X_{t_n})$ est donnée par $m_i = \mu_X(t_i)$.
2. La matrice de covariance $\Gamma = (\Gamma_{ij})_{1 \leq i, j \leq n}$ de $(X_{t_1}, \dots, X_{t_n})$ est donnée par $\Gamma_{ij} = \gamma_X(t_i, t_j)$.

Nous verrons que ces deux statistiques suffisent amplement en pratique à comprendre, analyser et prédire le processus X . En particulier, la fonction d'auto-covariance va jouer un rôle fondamental, et il est naturel de commencer par en rappeler les propriétés caractéristiques.

Théorème 1.1 (Propriétés de l'auto-covariance). La fonction d'auto-covariance γ_X est toujours

1. **symétrique** : pour tous $s, t \in \mathbb{Z}$, on a $\gamma_X(s, t) = \gamma_X(t, s)$;
2. **de type positif** : pour tout $n \geq 1$, tout $(t_1, \dots, t_n) \in \mathbb{Z}^n$, et tout $(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$,

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma_X(t_i, t_j) \geq 0. \quad (1.9)$$

Réciproquement, nous admettrons que toute fonction $\gamma: \mathbb{Z}^2 \rightarrow \mathbb{R}$ vérifiant ces deux propriétés est la fonction d'auto-covariance d'un processus du second ordre.

Démonstration. La première propriété découle du fait que la covariance est symétrique :

$$\gamma_X(s, t) = \text{Cov}(X_s, X_t) = \text{Cov}(X_t, X_s) = \gamma_X(t, s).$$

1.3. Stationnarité

Pour la seconde propriété, on utilise la bilinéarité de la covariance :

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma_X(t_i, t_j) &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \text{Cov}(X_{t_i}, X_{t_j}) \\ &= \text{Cov} \left(\sum_{i=1}^n \lambda_i X_{t_i}, \sum_{j=1}^n \lambda_j X_{t_j} \right) \\ &= \text{Var} \left(\sum_{i=1}^n \lambda_i X_{t_i} \right) \geq 0, \end{aligned}$$

puisque une variance est toujours positive ($\text{Var}(U) = \mathbb{E}[(U - \mathbb{E}[U])^2]$). \square

Remarque 1.1 (Matrice de covariance). *D'un point-de-vue algébrique, les propriétés 1 et 2 ci-dessus affirment simplement que pour tout $n \geq 1$ et $(t_1, \dots, t_n) \in \mathbb{Z}^n$, la matrice $\Gamma = (\Gamma_{ij})_{1 \leq i, j \leq n}$ définie par*

$$\Gamma_{ij} := \gamma_X(t_i, t_j) \quad (1.10)$$

est une matrice symétrique positive. Or Γ n'est rien d'autre que la matrice de covariance du vecteur aléatoire $(X_{t_1}, \dots, X_{t_n})$. Nous venons donc simplement de redémontrer une propriété bien connue : les matrices de covariance sont toujours des matrices symétriques positives.

1.3 Stationnarité

Pour les raisons évoquées plus haut, nous sommes amenés à considérer des processus dont les caractéristiques statistiques n'évoluent pas au cours du temps. La façon la plus naturelle de formaliser cela serait de supposer que la loi du processus décalé $(X_{t+s})_{t \geq 0}$ ne dépend pas de $s \in \mathbb{Z}$ (**stationnarité forte**). Mais puisque seules les moyennes et covariances nous intéressent, nous pouvons en fait affaiblir considérablement cette hypothèse en remplaçant *loi* par *moyennes et covariances*. Cela conduit à la définition suivante, que nous adopterons dans ce cours.

Définition 1.3 (Stationnarité). *Nous dirons que le processus $X = (X_t)_{t \in \mathbb{Z}}$ est **stationnaire** si les deux conditions suivantes sont remplies :*

1. *la moyenne $\mu_X(t)$ ne dépend pas de t ;*
2. *la covariance $\gamma_X(s, t)$ ne dépend que de la différence $t - s$ (i.e. $\gamma_X(t, t + h)$ ne dépend pas de t).*

Lorsque c'est le cas, nous noterons μ_X au lieu de $\mu_X(t)$ et $\gamma_X(h)$ au lieu de $\gamma_X(t, t + h)$. Avec ces notations, le Théorème 1.1 prend la forme simplifiée suivante dans le cas stationnaire.

Corollaire 1.1 (Auto-covariance des processus stationnaires). *Lorsque X est stationnaire, sa fonction d'auto-covariance $\gamma_X: \mathbb{Z} \rightarrow \mathbb{R}$ est toujours*

1. **paire** : $\gamma_X(h) = \gamma_X(-h)$ pour tout $h \in \mathbb{Z}$;
2. **de type positif** : pour tout $n \geq 1$, tout $(t_1, \dots, t_n) \in \mathbb{Z}^n$, et tout $(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$,

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma_X(t_i - t_j) \geq 0. \quad (1.11)$$

Réciproquement, toute fonction $\gamma: \mathbb{Z} \rightarrow \mathbb{R}$ vérifiant ces deux propriétés est la fonction d'auto-covariance d'un processus stationnaire.

Nous terminons ce chapitre en donnant trois exemples élémentaires mais fondamentaux de processus stationnaires. Le premier, en particulier, nous servira de brique de base pour en construire de nombreux autres au chapitre suivant.

Exemple 1.1 (Bruit blanc). Soient $(X_t)_{t \in \mathbb{Z}}$ des variables aléatoires dans L^2 qui sont

1. De même moyenne $\mu \in \mathbb{R}$.
2. De même variance $\sigma^2 \in \mathbb{R}_+$.
3. Décorrélées : $\text{Cov}(X_t, X_s) = 0$ pour tout $t \neq s$.

Alors le processus $X = (X_t)_{t \in \mathbb{Z}}$ est évidemment stationnaire, et ses paramètres sont

$$\mu_X = \mu, \quad \text{et} \quad \gamma_X(h) = \begin{cases} \sigma^2 & \text{si } h = 0 \\ 0 & \text{sinon.} \end{cases} \quad (1.12)$$

On dit que X est un **bruit blanc** et on note $X \sim \text{BB}(\mu, \sigma^2)$. Notons qu'il s'agit d'une version très affaiblie de la notion de suite i.i.d. (parfois appelée **bruit blanc fort**) : les hypothèses 1 et 2 remplacent le caractère identiquement distribué, tandis que l'hypothèse 3 remplace le caractère indépendant. Intuitivement, le bruit blanc est un signal sans structure. Dans le chapitre suivant, il constituera la brique élémentaire à partir de laquelle nous construirons tous les processus dits linéaires, qui seront au cœur de notre étude.

Exemple 1.2 (Signal constant). Soit U une variable aléatoire dans L^2 de moyenne $\mu \in \mathbb{R}$ et de variance $\sigma^2 \in \mathbb{R}_+$. On définit un processus $X = (X_t)_{t \in \mathbb{Z}}$ en posant

$$X_t := U, \quad (1.13)$$

pour tout $t \in \mathbb{Z}$. Alors $X = (X_t)_{t \in \mathbb{Z}}$ est bien-sûr stationnaire, avec

$$\mu_X = \mu, \quad \text{et} \quad \gamma_X(h) = \sigma^2. \quad (1.14)$$

Notons que les coordonnées ont exactement la même moyenne et la même variance que dans l'exemple précédent, mais que les corrélations sont radicalement différentes !

Exemple 1.3 (Processus harmonique). Soient $\theta \in [-\pi, \pi]$ un paramètre et $U, V \in L^2$ deux variables aléatoires décorrélées ($\text{Cov}(U, V) = 0$), de moyenne 0 et variance 1. On définit un processus $X = (X_t)_{t \in \mathbb{Z}}$ (appelé **processus harmonique** de fréquence θ) en posant pour tout $t \in \mathbb{Z}$,

$$X_t := U \cos(\theta t) + V \sin(\theta t). \quad (1.15)$$

D'abord, le processus $X = (X_t)_{t \in \mathbb{Z}}$ est bien du second ordre, puisque ses coordonnées sont combinaisons linéaires d'éléments de l'espace vectoriel L^2 . De plus, par linéarité de l'espérance, on a

$$\mu_X(t) = \cos(\theta t) \mathbb{E}[U] + \sin(\theta t) \mathbb{E}[V] = 0,$$

qui ne dépend pas de t . Enfin, en utilisant la bilinéarité de la covariance et les hypothèses $\text{Var}(U) = \text{Var}(V) = 1$ et $\text{Cov}(U, V) = 0$, on trouve aisément

$$\begin{aligned} \gamma_X(s, t) &= \cos(\theta t) \cos(\theta s) + \sin(\theta t) \sin(\theta s) \\ &= \cos(\theta(t - s)), \end{aligned}$$

qui ne dépend que de la différence $t - s$. Ainsi, X est stationnaire, de paramètres

$$\mu_X = 0, \quad \text{et} \quad \gamma_X(h) = \cos(\theta h). \quad (1.16)$$

Ce processus, exotique à bien des égards, nous servira souvent de contre-exemple : il possède des propriétés bien différentes des processus linéaires que nous considérerons par la suite.

Chapitre 2

Filtrage linéaire

Étant donné un processus $X = (X_t)_{t \in \mathbb{Z}}$ et une suite réelle $\alpha = (\alpha_t)_{t \in \mathbb{Z}}$, on aimerait construire un nouveau processus $Y = (Y_t)_{t \in \mathbb{Z}}$ en posant pour tout $t \in \mathbb{Z}$,

$$Y_t := \sum_{k \in \mathbb{Z}} \alpha_k X_{t-k}. \quad (2.1)$$

Bien-sûr, quelques hypothèses sont nécessaires pour garantir que cette somme infinie aléatoire ait bien un sens. Lorsque c'est le cas, on appelle Y le **filtrage** de X par le **filtre** α , et l'on note $Y = F_\alpha(X)$. On parlera en outre de filtre **causal** lorsque $\alpha_k = 0$ pour tout $k < 0$. Le filtrage est une opération fondamentale, à laquelle le présent chapitre est entièrement consacré.

2.1 Existence et propriétés

Commençons par quelques petits rappels.

Rappel 2.1. Pour $p \geq 1$, $\ell^p = \ell^p(\mathbb{Z})$ désigne l'espace vectoriel des suites $\alpha = (\alpha_k)_{k \in \mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}}$ avec

$$\|\alpha\|_p := \|\alpha\|_{\ell^p} = \left(\sum_{k \in \mathbb{Z}} |\alpha_k|^p \right)^{1/p} < \infty. \quad (2.2)$$

Il ne faut pas le confondre avec l'espace vectoriel $L^p = L^p(\Omega, \mathcal{F}, \mathbb{P})$ des (classes d'équivalences de) variables aléatoires réelles X sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ vérifiant

$$\|X\|_p := \|X\|_{L^p} := (\mathbb{E}[|X|^p])^{1/p} < \infty. \quad (2.3)$$

Les deux sont des espaces de Banach pour la norme $\|\cdot\|_p$ associée, mais attention aux inclusions :

$$q > p \implies \begin{cases} \ell^q \supsetneq \ell^p \\ L^q \subsetneq L^p. \end{cases} \quad (2.4)$$

Rappelons également qu'une suite $\alpha = (\alpha_k)_{k \in \mathbb{Z}}$ qui est dans $\ell^1(\mathbb{Z})$ est dite **absolument sommable**. Dans ce cas, sa somme $\sum_{k \in \mathbb{Z}} \alpha_k$ existe et c'est la limite des sommes partielles :

$$\sum_{k \in [-n, n]} \alpha_k \xrightarrow{n \rightarrow \infty} \sum_{k \in \mathbb{Z}} \alpha_k. \quad (2.5)$$

De plus, on peut remplacer $[-n, n]$ par n'importe quelle partie finie $K_n \subseteq \mathbb{Z}$ telle que $K_n \nearrow \mathbb{Z}$.

2.1. Existence et propriétés

Théorème 2.1 (Filtrage des processus bornés). *On se donne une suite $\alpha = (\alpha_k)_{k \in \mathbb{Z}} \in \ell^1(\mathbb{Z})$, un nombre $p \in [1, \infty[$, et un processus stochastique $X = (X_t)_{t \in \mathbb{Z}}$ borné dans L^p , i.e.*

$$\sup_{t \in \mathbb{Z}} \|X_t\|_{L^p} < \infty. \quad (2.6)$$

Alors, on a les résultats suivants.

1. Pour tout $t \in \mathbb{Z}$, la somme (2.1) est presque-sûrement absolument convergente.
2. Le processus aléatoire $Y = (Y_t)_{t \in \mathbb{Z}}$ défini par (2.1) est borné dans L^p .
3. Pour tout $t \in \mathbb{Z}$, on a presque-sûrement et dans L^p ,

$$\sum_{k \in [-n, n]} \alpha_k X_{t-k} \xrightarrow[n \rightarrow \infty]{} Y_t. \quad (2.7)$$

De plus, on peut remplacer $[-n, n]$ par n'importe quelle partie finie $K_n \subseteq \mathbb{Z}$ telle que $K_n \nearrow \mathbb{Z}$.

Démonstration. Pour tout $n \in \mathbb{N}$, l'inégalité triangulaire pour la norme $\|\cdot\|_{L^p}$ donne

$$\begin{aligned} \mathbb{E} \left[\left(\sum_{k=-n}^n |\alpha_k X_{t-k}| \right)^p \right]^{\frac{1}{p}} &= \left\| \sum_{k=-n}^n |\alpha_k X_{t-k}| \right\|_{L^p} \\ &\leq \sum_{k=-n}^n |\alpha_k| \cdot \|X_{t-k}\|_{L^p} \\ &\leq \|\alpha\|_1 \left(\sup_{t \in \mathbb{Z}} \|X_t\|_{L^p} \right). \end{aligned}$$

On peut alors faire tendre $n \rightarrow \infty$ et invoquer la convergence monotone pour obtenir :

$$\mathbb{E} \left[\left(\sum_{k \in \mathbb{Z}} |\alpha_k X_{t-k}| \right)^p \right]^{\frac{1}{p}} \leq \|\alpha\|_1 \left(\sup_{t \in \mathbb{Z}} \|X_t\|_{L^p} \right). \quad (2.8)$$

Le membre droit est fini par hypothèse. En particulier, la variable sous l'espérance à gauche est finie p.-s., ce qui est le point 1. Ainsi Y_t est bien définie et l'on a par inégalité triangulaire

$$|Y_t| \leq \sum_{k \in \mathbb{Z}} |\alpha_k X_{t-k}|. \quad (2.9)$$

Combinée avec la précédente, cette inégalité entraîne

$$\|Y_t\|_{L^p} \leq \|\alpha\|_1 \left(\sup_{t \in \mathbb{Z}} \|X_t\|_{L^p} \right), \quad (2.10)$$

ce qui établit le point 2. En remplaçant α_k par 0 pour $k \in K_n$, cette inégalité devient

$$\left\| Y_t - \sum_{k \in K_n} \alpha_k X_{t-k} \right\|_{L^p} \leq \left(\sum_{k \in \mathbb{Z} \setminus K_n} |\alpha_k| \right) \left(\sup_{t \in \mathbb{Z}} \|X_t\|_{L^p} \right), \quad (2.11)$$

et la sommabilité de α assure que le membre droit tend vers 0 lorsque $n \rightarrow \infty$. Cela établit la convergence L^p promise par le point 3. La convergence p.-s. est évidente d'après le point 1. \square

Nous allons appliquer ce théorème au cas particulier des processus stationnaires. Avant cela, nous avons besoin d'un petit rappel sur la continuité dans L^2 des espérances et covariances.

Rappel 2.2 (Cauchy-Schwartz). *Sur un espace de Hilbert \mathcal{H} , l'inégalité de Cauchy-Schwartz affirme :*

$$\forall U, V \in \mathcal{H}, \quad |\langle U, V \rangle| \leq \|U\| \|V\|. \quad (2.12)$$

Une conséquence importante de cette inégalité est que le produit scalaire $\langle \cdot, \cdot \rangle$ est continu sur \mathcal{H} :

$$\begin{cases} U_n \xrightarrow[n \rightarrow \infty]{\mathcal{H}} U \\ V_n \xrightarrow[n \rightarrow \infty]{\mathcal{H}} V \end{cases} \implies \langle U_n, V_n \rangle \xrightarrow[n \rightarrow \infty]{} \langle U, V \rangle. \quad (2.13)$$

En effet, la bilinéarité du produit scalaire permet d'écrire

$$\langle U_n, V_n \rangle - \langle U, V \rangle = \langle U_n - U, V \rangle + \langle U_n - U, V_n - V \rangle + \langle U, V_n - V \rangle, \quad (2.14)$$

et les trois termes de droite tendent vers zéro lorsque $n \rightarrow \infty$ par (2.12). En particulier, cela est vrai dans l'espace de Hilbert $L^2(\Omega, \mathcal{F}, \mathbb{P})$ muni du produit scalaire $\langle U, V \rangle := \mathbb{E}[UV]$. Or dans cet espace, les espérances et covariances peuvent s'écrire à l'aide du produit scalaire comme suit :

$$\begin{aligned} \mathbb{E}[U] &= \langle U, 1 \rangle \\ \text{Cov}(U, V) &= \langle U, V \rangle - \langle U, 1 \rangle \langle V, 1 \rangle. \end{aligned}$$

On en déduit aussitôt le résultat important suivant : l'espérance $U \mapsto \mathbb{E}[U]$ et la covariance $(U, V) \mapsto \text{Cov}(U, V)$ sont continues sur $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Autrement dit,

$$\begin{cases} U_n \xrightarrow{L^2} U \\ V_n \xrightarrow{L^2} V. \end{cases} \implies \begin{cases} \mathbb{E}[U_n] \rightarrow \mathbb{E}[U] \\ \mathbb{E}[V_n] \rightarrow \mathbb{E}[V] \\ \text{Cov}(U_n, V_n) \rightarrow \text{Cov}(U, V) \end{cases} \quad (2.15)$$

Notons enfin une autre conséquence intéressante de l'inégalité de Cauchy-Schwartz : si X est un processus du second ordre, alors en prenant $U = X_s - \mathbb{E}[X_s]$ et $V = X_t - \mathbb{E}[X_t]$ dans (2.12), il vient

$$\forall s, t \in \mathbb{Z}, \quad |\gamma_X(s, t)| \leq \sqrt{\gamma_X(s, s) \gamma_X(t, t)}. \quad (2.16)$$

En particulier, si X est stationnaire, cette inégalité prend la forme plus simple suivante :

$$\forall h \in \mathbb{Z}, \quad |\gamma_X(h)| \leq \gamma_X(0). \quad (2.17)$$

Ainsi, la fonction d'auto-covariance d'un processus stationnaire est bornée par sa valeur en zéro.

Corollaire 2.1 (Cas stationnaire). *Soit $\alpha \in \ell^1(\mathbb{Z})$ et soit $X = (X_t)_{t \in \mathbb{Z}}$ un processus du second ordre stationnaire. Alors le processus $Y = F_\alpha(X)$ est bien défini, du second ordre, stationnaire, avec*

$$\mu_Y = \mu_X \sum_{k \in \mathbb{Z}} \alpha_k, \quad \gamma_Y(h) = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \alpha_j \alpha_k \gamma_X(h + j - k).$$

Démonstration. Le processus X est borné dans L^2 , puisque l'on a pour tout $t \in \mathbb{Z}$,

$$\|X_t\|_{L^2} = \sqrt{\mathbb{E}[X_t^2]} = \sqrt{\mathbb{E}[X_t]^2 + \text{Var}(X_t)} = \sqrt{\mu_X + \gamma_X(0)}. \quad (2.18)$$

Ainsi, le théorème précédent avec $p = 2$ affirme l'existence du processus $Y = F_\alpha(X)$, garantit qu'il est du second ordre, et assure que pour tout $t \in \mathbb{Z}$, on a p.-s. et dans L^2 ,

$$Y_t = \lim_{n \rightarrow \infty} \sum_{k=-n}^n \alpha_k X_{t-k} \quad (2.19)$$

2.2. Composition de filtres

Par continuité dans L^2 de l'espérance et de la covariance (remarque ci-dessus), on en déduit que

$$\begin{aligned}
 \mathbb{E}[Y_t] &= \lim_{n \rightarrow \infty} \mathbb{E} \left[\sum_{k=-n}^n \alpha_k X_{t-k} \right] \\
 &= \lim_{n \rightarrow \infty} \sum_{k=-n}^n \alpha_k \mathbb{E}[X_{t-k}] \\
 &= \mu_X \sum_{k \in \mathbb{Z}} \alpha_k. \\
 \text{Cov}(Y_t, Y_{t+h}) &= \lim_{n \rightarrow \infty} \text{Cov} \left(\sum_{j=-n}^n \alpha_j X_{t-j}, \sum_{k=-n}^n \alpha_k X_{t+h-k} \right) \\
 &= \lim_{n \rightarrow \infty} \sum_{j=-n}^n \sum_{k=-n}^n \alpha_j \alpha_k \gamma_X(h+j-k) \\
 &= \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \alpha_j \alpha_k \gamma_X(h+j-k).
 \end{aligned}$$

L'absolue convergence de la double somme découle de la majoration $|\gamma_X(h+j-k)| \leq \gamma_X(0)$ établie en (2.17). Les quantités ci-dessus ne dépendent pas de t , donc Y est bien stationnaire. \square

Exemple 2.1 (Processus linéaire). Soit Z un bruit blanc $BB(0,1)$, et soit $\alpha \in \ell^1(\mathbb{Z})$. Alors le processus $X = F_\alpha(Z)$ est un processus stationnaire d'espérance nulle et d'auto-covariance

$$\gamma_X(h) = \sum_{k \in \mathbb{Z}} \alpha_k \alpha_{k+h}. \quad (2.20)$$

Un tel processus (obtenu par filtrage d'un bruit blanc) s'appelle **processus linéaire**.

Exercice 2.1 (Un exemple de processus non-linéaire). Montrer que si X est un processus linéaire, alors $\gamma_X \in \ell^1(\mathbb{Z})$. En déduire que le processus harmonique n'est pas un processus linéaire.

Exercice 2.2 (Identifiabilité). Montrer que si Z est un bruit blanc standard et $\alpha, \beta \in \ell^1(\mathbb{Z})$, alors

$$F_\alpha(Z) = F_\beta(Z) \iff \alpha = \beta. \quad (2.21)$$

Indication : on pourra calculer $\text{Cov}(Z_0, (F_\alpha(Z))_t)$.

2.2 Composition de filtres

Lorsque l'on applique un filtre $\beta \in \ell^1(\mathbb{Z})$ à un processus stationnaire X , on obtient un nouveau processus stationnaire $F_\beta(X)$, par le Théorème 2.1. On peut donc appliquer à nouveau un filtre $\alpha \in \ell^1(\mathbb{Z})$, et obtenir ainsi un troisième processus $F_\alpha(F_\beta(X))$. Le Théorème 2.2 ci-dessus affirme qu'une telle composition de filtres revient en réalité à n'appliquer qu'un seul filtre, à savoir la **convolution** $\alpha \star \beta$, définie comme suit : pour tout $k \in \mathbb{Z}$,

$$(\alpha \star \beta)_k := \sum_{j \in \mathbb{Z}} \alpha_j \beta_{k-j}. \quad (2.22)$$

Notons qu'en utilisant Fubini-Tonelli puis un changement d'indice, on a

$$\sum_{k \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} |\alpha_j \beta_{k-j}| = \sum_{j \in \mathbb{Z}} |\alpha_j| \sum_{k \in \mathbb{Z}} |\beta_{k-j}| = \|\alpha\|_1 \|\beta\|_1,$$

ce qui montre que la suite $\alpha \star \beta$ est bien définie, et qu'elle est dans $\ell^1(\mathbb{Z})$.

Théorème 2.2 (Composition). Soit X un processus stationnaire et $\alpha, \beta \in \ell^1(\mathbb{Z})$ deux filtres. Alors,

$$F_\alpha(F_\beta(X)) = F_{\alpha \star \beta}(X). \quad (2.23)$$

Démonstration. Pour tout $t \in \mathbb{Z}$, on a

$$\begin{aligned} (F_\alpha(F_\beta(X)))_t &= \sum_{k \in \mathbb{Z}} \alpha_k (F_\beta(X))_{t-k} \\ &= \sum_{k \in \mathbb{Z}} \alpha_k \left(\sum_{j \in \mathbb{Z}} \beta_j X_{t-k-j} \right) \\ &= \sum_{k \in \mathbb{Z}} \alpha_k \left(\sum_{j \in \mathbb{Z}} \beta_{j-k} X_{t-j} \right) \\ &= \sum_{j \in \mathbb{Z}} \left(\sum_{k \in \mathbb{Z}} \alpha_k \beta_{j-k} \right) X_{t-j} \\ &= \sum_{j \in \mathbb{Z}} (\alpha \star \beta)_j X_{t-j} \\ &= (F_{\alpha \star \beta}(X))_t. \end{aligned}$$

L'interversion des deux sommes à la quatrième ligne est justifiée, puisque l'on a

$$\sum_{k \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} |\alpha_k \beta_{j-k} X_{t-j}| < \infty \quad (2.24)$$

presque-sûrement, comme on le voit en prenant l'espérance. \square

L'opération de convolution \star possède les propriétés remarquables suivantes, qui font de $(\ell^1(\mathbb{Z}), +, \star)$ une *algèbre de Banach* (pour la culture), et dont nous laissons la preuve en exercice.

Proposition 2.1 (Propriétés de la convolution). L'opération de convolution \star est

1. *commutative* : pour $\alpha, \beta \in \ell^1(\mathbb{Z})$, on a

$$\alpha \star \beta = \beta \star \alpha;$$

2. *associative* : pour $\alpha, \beta, \gamma \in \ell^1(\mathbb{Z})$, on a

$$(\alpha \star \beta) \star \gamma = \alpha \star (\beta \star \gamma);$$

3. *distributive* : pour $\alpha, \beta, \gamma \in \ell^1(\mathbb{Z})$ et $\lambda \in \mathbb{R}$, on a

$$(\alpha + \lambda\beta) \star \gamma = \alpha \star \gamma + \lambda(\beta \star \gamma);$$

4. avec *élément neutre* : la suite $\mathbf{e} = (\mathbf{e}_k)_{k \in \mathbb{Z}}$ définie par $\mathbf{e}_k = \mathbf{1}_{k=0}$ vérifie, pour $\alpha \in \ell^1(\mathbb{Z})$,

$$\alpha \star \mathbf{e} = \alpha;$$

5. *sous-multiplicative* : pour $\alpha, \beta \in \ell^1(\mathbb{Z})$, on a

$$\|\alpha \star \beta\|_1 \leq \|\alpha\|_1 \cdot \|\beta\|_1.$$

2.3 Représentation en série de puissances

Par la suite, il nous sera utile de représenter un filtre $\alpha \in \ell^1(\mathbb{Z})$ par sa **série de puissances**

$$P_\alpha(z) := \sum_{k \in \mathbb{Z}} \alpha_k z^k, \quad (2.25)$$

qui est absolument convergente sur le cercle unité

$$\mathbb{U} := \{z \in \mathbb{C} : |z| = 1\}. \quad (2.26)$$

Le résultat suivant montre que l'application $\alpha \mapsto P_\alpha$ est un homomorphisme injectif de l'algèbre $(\ell^1(\mathbb{Z}), +, \star)$ vers l'algèbre $(\mathcal{C}_\mathbb{C}^0(\mathbb{U}), +, \times)$ des fonctions complexes continues sur le cercle unité. Cette phrase compliquée signifie simplement que la représentation en séries de puissances est une "bonne" représentation : elle ne perd pas d'information sur les filtres, mais elle réduit leur manipulation à celle d'objets beaucoup plus classiques (les fonctions continues).

Proposition 2.2 (Propriétés de la représentation en série de puissances).

1. La série de puissance P_α d'un filtre $\alpha \in \ell^1(\mathbb{Z})$ est une fonction continue de \mathbb{U} dans \mathbb{C} .
2. La représentation $\alpha \mapsto P_\alpha$ est un **homomorphisme d'algèbres** : pour $\alpha, \beta \in \ell^1(\mathbb{Z}), \lambda \in \mathbb{C}, z \in \mathbb{U}$,

$$P_{\lambda\alpha}(z) = \lambda P_\alpha(z); \quad (2.27)$$

$$P_{\alpha+\beta}(z) = P_\alpha(z) + P_\beta(z); \quad (2.28)$$

$$P_{\alpha \star \beta}(z) = P_\alpha(z) \times P_\beta(z) \quad (2.29)$$

$$P_{\mathbf{e}}(z) = 1. \quad (2.30)$$

3. On peut retrouver le filtre α à partir de sa série de puissances P_α grâce à la **formule d'inversion** :

$$\forall k \in \mathbb{Z}, \quad \alpha_k = \frac{1}{2\pi} \int_0^{2\pi} P_\alpha(e^{i\theta}) e^{-ik\theta} d\theta. \quad (2.31)$$

En particulier, la représentation en série de puissances est **injective** : pour $\alpha, \beta \in \ell^1(\mathbb{Z})$, on a

$$P_\alpha = P_\beta \implies \alpha = \beta. \quad (2.32)$$

Démonstration. La fonction P_α est bien continue sur le cercle unité, par une application immédiate du théorème de convergence dominée. Les propriétés (2.28) et (2.27) découlent de la linéarité de Σ pour les séries absolument convergentes. Pour (2.29), on écrit

$$\begin{aligned} P_{\alpha \star \beta}(z) &= \sum_{n \in \mathbb{Z}} (\alpha \star \beta)_n z^n \\ &= \sum_{n \in \mathbb{Z}} \left(\sum_{k \in \mathbb{Z}} \alpha_k \beta_{n-k} \right) z^n \\ &= \sum_{k \in \mathbb{Z}} \alpha_k z^k \left(\sum_{n \in \mathbb{Z}} \beta_{n-k} z^{n-k} \right) \\ &= \sum_{k \in \mathbb{Z}} \alpha_k z^k \left(\sum_{n \in \mathbb{Z}} \beta_n z^n \right) \\ &= P_\alpha(z) P_\beta(z), \end{aligned}$$

où l'on a utilisé le théorème de Fubini puis le changement d'indice bijectif $n \mapsto n + k$. Enfin, pour la formule d'inversion, on utilise à nouveau Fubini pour écrire

$$\begin{aligned} \int_0^{2\pi} P_\alpha(e^{i\theta})e^{-ik\theta} d\theta &= \int_0^{2\pi} \left(\sum_{n \in \mathbb{Z}} \alpha_n e^{i\theta n} \right) e^{-ik\theta} d\theta \\ &= \sum_{n \in \mathbb{Z}} \alpha_n \left(\int_0^{2\pi} e^{i\theta(n-k)} d\theta \right). \end{aligned}$$

Mais la dernière intégrale vaut 0 si $n \neq k$ et 2π si $n = k$, et l'on obtient exactement (2.31). \square

2.4 Défiltrage

Dans ce paragraphe, nous nous intéressons à la question du **défiltrage** : peut-on annuler l'effet d'un filtre donné $\alpha \in \ell^1(\mathbb{Z})$ par l'application d'un autre filtre $\beta \in \ell^1(\mathbb{Z})$ bien choisi? D'après le Théorème 2.2, il suffit de choisir $\beta \in \ell^1(\mathbb{Z})$ tel que :

$$\alpha \star \beta = \epsilon. \quad (2.33)$$

Autrement dit, il s'agit de trouver une suite $\beta \in \ell^1(\mathbb{Z})$ qui résout le système infini d'équations

$$\forall k \in \mathbb{Z}, \quad \sum_{j \in \mathbb{Z}} \alpha_j \beta_{k-j} = \begin{cases} 1 & \text{si } k = 0 \\ 0 & \text{sinon.} \end{cases} \quad (2.34)$$

Attention : cela n'est pas toujours possible, comme le montre l'exemple $\alpha = 0$ (ce n'est pas le seul). Nous dirons que le filtre $\alpha \in \ell^1(\mathbb{Z})$ est **inversible** s'il existe un filtre $\beta \in \ell^1(\mathbb{Z})$ tel que l'égalité (2.33) est vérifiée. Notons qu'un tel β est alors unique. En effet, si $\gamma \in \ell^1(\mathbb{Z})$ vérifie aussi

$$\alpha \star \gamma = \epsilon, \quad (2.35)$$

alors en convolant par β des deux côtés, on obtient $\gamma = \beta$ par commutativité et associativité de \star , d'où l'unicité. L'unique $\beta \in \ell^1(\mathbb{Z})$ vérifiant (2.33) sera appelé **inverse** de α et noté α^{-1} . Notons que α^{-1} est alors trivialement inversible, d'inverse $(\alpha^{-1})^{-1} = \alpha$. On vérifie tout aussi facilement que si α et β sont inversibles, alors il en est de même de $\alpha \star \beta$, et

$$(\alpha \star \beta)^{-1} = \beta^{-1} \star \alpha^{-1}. \quad (2.36)$$

En pratique, il n'est pas forcément évident de résoudre le système (2.34) pour inverser un filtre. Le théorème suivant fournit une solution simple dans le cas où le filtre α est **polynomial**.

Définition 2.1 (Filtre polynomial). *Un filtre $\alpha \in \ell^1(\mathbb{Z})$ est dit **polynômial** si sa série de puissances P_α est un polynôme, c'est-à-dire si son support $\text{supp}(\alpha) := \{k \in \mathbb{Z} : \alpha_k \neq 0\}$ est une partie finie de \mathbb{N} .*

Théorème 2.3 (Inversion des filtres polynomiaux). *Soit $\alpha \in \ell^1(\mathbb{Z})$ un filtre polynomial. Alors les conditions suivantes sont équivalentes :*

- (i) Le filtre α est inversible.
- (ii) Le polynôme $P_\alpha(z) = \sum_{k \in \mathbb{Z}} \alpha_k z^k$ n'a aucune racine de module 1.
- (iii) La fraction rationnelle $\frac{1}{P_\alpha}$ est développable en série de puissances sur le cercle unité : il existe $\beta \in \ell^1(\mathbb{Z})$ tel que pour tout $z \in \mathbb{U}$,

$$\frac{1}{P_\alpha(z)} = \sum_{k \in \mathbb{Z}} \beta_k z^k. \quad (2.37)$$

De plus, lorsque c'est le cas, on a nécessairement $\alpha^{-1} = \beta$.

Démonstration de (iii) \Rightarrow (i). On utilise la Proposition 2.2 pour écrire, pour $z \in \mathbb{U}$,

$$P_{\alpha \star \beta}(z) = P_\alpha(z)P_\beta(z) = \frac{P_\alpha(z)}{P_\alpha(z)} = 1. \quad (2.38)$$

Par injectivité de $\alpha \mapsto P_\alpha$, on en déduit que $\alpha \star \beta = \epsilon$, donc que α est inversible d'inverse β . \square

Démonstration de (i) \Rightarrow (ii). Si α admet un inverse β , la Proposition 2.2 donne pour tout $z \in \mathbb{U}$,

$$P_\alpha(z)P_\beta(z) = P_{\alpha \star \beta}(z) = P_\epsilon(z) = 1, \quad (2.39)$$

ce qui implique bien en particulier que $P_\alpha(z) \neq 0$. \square

Démonstration de (ii) \Rightarrow (iii). En notant a_1, \dots, a_d les racines distinctes de P_α , m_1, \dots, m_d leur multiplicités respectives, et $c \in \mathbb{R}$ le coefficient dominant, on a

$$P_\alpha(z) = c(z - a_1)^{m_1} \cdots (z - a_d)^{m_d}. \quad (2.40)$$

Par hypothèse, nous savons que $c \neq 0$ et $|a_1|, \dots, |a_d| \neq 1$. Par décomposition en éléments simples, la fraction rationnelle $1/P_\alpha$ peut toujours se mettre sous la forme

$$\frac{1}{P_\alpha(z)} = \sum_{j=1}^d \sum_{k=1}^{m_d} \frac{c_{j,k}}{(z - a_j)^k}, \quad (2.41)$$

avec $c_{j,k} \in \mathbb{C}$. Comme une combinaison linéaire de séries de puissances est encore une série de puissances, il nous suffit maintenant de montrer que toute fraction simple de la forme

$$z \mapsto \frac{1}{(z - a)^k}, \quad (2.42)$$

avec $|a| \neq 1$ et $k \in \mathbb{N}$ est décomposable en séries de puissances sur le cercle unité. Pour cela, rappelons que pour $|x| < 1$, on a

$$\frac{1}{1 - x} = \sum_{n=0}^{\infty} x^n.$$

En dérivant $k - 1$ fois des deux côtés, on obtient pour $|x| < 1$

$$\frac{(k - 1)!}{(1 - x)^k} = \sum_{n=k-1}^{\infty} n(n - 1) \cdots (n - k + 2)x^{n-k+1},$$

c'est-à-dire, après changement d'indice et division par $(k - 1)!$,

$$\frac{1}{(1 - x)^k} = \sum_{n=0}^{\infty} \binom{n + k - 1}{k - 1} x^n. \quad (2.43)$$

En particulier, pour $z \in \mathbb{U}$ et $|a| > 1$, on peut prendre $x = z/a$ pour obtenir

$$\frac{1}{(z - a)^k} = \frac{1}{(-a)^k (1 - \frac{z}{a})^k} = \frac{1}{(-a)^k} \sum_{n=0}^{\infty} \binom{n + k - 1}{k - 1} z^n. \quad (2.44)$$

À l'inverse, pour $z \in \mathbb{U}$ et $|a| < 1$, on peut prendre $x = a/z$ pour obtenir

$$\frac{1}{(z - a)^k} = \frac{1}{z^k (1 - \frac{a}{z})^k} = \sum_{n=0}^{\infty} \binom{n + k - 1}{k - 1} \frac{a^n}{z^{n+k}}. \quad (2.45)$$

Dans les deux cas, on obtient bien un développement en série de puissances de $z \mapsto \frac{1}{(z-a)^k}$ sur le cercle unité. En appliquant cela à chacun des termes de la décomposition (2.41), on obtient finalement bien un développement en séries de puissances de la forme (2.37). Notons que, bien que les racines de P_α soient a priori complexes, la suite $\beta = (\beta_n)_{n \in \mathbb{Z}}$ finalement obtenue est en fait nécessairement réelle, car l'égalité $\overline{P_\alpha(\bar{z})} = P_\alpha(z)$ impose, pour $z \in \mathbb{U}$,

$$\sum_{n \in \mathbb{Z}} \overline{\beta_n} z^n = \sum_{n \in \mathbb{Z}} \beta_n z^n. \quad (2.46)$$

□

Remarque 2.1 (Causalité). *Remarquons que les puissances apparaissant dans la série (2.44) (cas $|a| < 1$) sont négatives, tandis que celles apparaissant dans la série (2.45) sont positives (cas $|a| > 1$). On en déduit que α^{-1} est causal si et seulement si P_α n'a pas de racine dans le disque unité.*

Exemple 2.2. *Inverser le filtrage $X_t = Z_t + \frac{3}{2}Z_{t-1} - Z_{t-2}$.*

Chapitre 3

Processus ARMA

3.1 Définition et cas particuliers

Définition 3.1 (ARMA). Soient $p, q \in \mathbb{N}$, $\varphi = (\varphi_1, \dots, \varphi_p) \in \mathbb{R}^p$, et $\theta = (\theta_1, \dots, \theta_q) \in \mathbb{R}^q$. Soit également $Z = (Z_t)_{t \in \mathbb{Z}}$ un bruit blanc standard. Un processus stationnaire $X = (X_t)_{t \in \mathbb{Z}}$ est appelé :

1. **Processus AR(p)** (Auto Regressive) s'il vérifie une équation de la forme

$$X_t = \sum_{k=1}^p \varphi_k X_{t-k} + Z_t. \quad (3.1)$$

2. **Processus MA(q)** (Moving Average) s'il vérifie une équation de la forme

$$X_t = \sum_{k=1}^q \theta_k Z_{t-k} + Z_t \quad (3.2)$$

3. **Processus ARMA(p, q)** (Auto Regressive Moving Average) s'il vérifie une équation de la forme

$$X_t = \sum_{k=1}^p \varphi_k X_{t-k} + Z_t + \sum_{k=1}^q \theta_k Z_{t-k}. \quad (3.3)$$

Les processus MA(q), $q < \infty$ sont très simples : ce sont les processus linéaires causaux à support fini. Effet, l'équation (3.2) s'écrit $X = F_{\alpha_\theta}(Z)$, où le filtre α_θ est défini par

$$(\alpha_\theta)_k := \begin{cases} 1 & \text{si } k = 0; \\ \theta_k & \text{si } k \in \{1, 2, \dots, q\}; \\ 0 & \text{sinon.} \end{cases} \quad (3.4)$$

On a trivialement $\theta \in \ell^1(\mathbb{Z})$, donc l'équation ci-dessus définit toujours un unique processus X , lequel est stationnaire, causal, de moyenne nulle et d'auto-covariance

$$\gamma_X(h) = \begin{cases} \sum_{k=0}^{q-|h|} \theta_k \theta_{k+|h|} & \text{si } |h| \leq q; \\ 0 & \text{sinon.} \end{cases} \quad (3.5)$$

En revanche, ni l'existence ni l'unicité ne sont garanties en général pour l'équation (3.1) et plus généralement (3.3), et c'est à la résolution de ces équations que le chapitre est consacré. En guise d'échauffement, commençons par le cas particulier simple des processus AR(1).

Théorème 3.1 (Cas des processus AR(1)). Dans le cas où $p = 1$, l'équation (3.1) prend la forme

$$X_t = \varphi X_{t-1} + Z_t. \quad (3.6)$$

3.1. Définition et cas particuliers

1. Si $|\varphi| = 1$, l'équation (3.6) n'admet pas de solution stationnaire.
2. Si $|\varphi| < 1$, alors l'équation (3.6) admet une unique solution stationnaire, donnée par

$$X_t = \sum_{k=0}^{\infty} \varphi^k Z_{t-k}. \quad (3.7)$$

X est linéaire, causal, de moyenne $\mu_X = 0$ et d'auto-covariance $\gamma_X(h) = \varphi^{|h|} / (1 - \varphi^2)$.

3. Si $|\varphi| > 1$, alors l'équation (3.6) admet une unique solution stationnaire, donnée par

$$X_t = - \sum_{k=1}^{\infty} \frac{Z_{t+k}}{\varphi^k}. \quad (3.8)$$

X est linéaire, non-causal, de moyenne $\mu_X = 0$ et d'auto-covariance $\gamma_X(h) = \varphi^{-|h|} / (\varphi^2 - 1)$.

Démonstration. Commençons par noter que si X est solution de (3.6), alors en itérant $n \geq 1$ fois,

$$X_t = \varphi^n X_{t-n} + \sum_{k=0}^{n-1} \varphi^k Z_{t-k}. \quad (3.9)$$

1. Cas $|\varphi| = 1$. D'après l'équation ci-dessus et l'orthogonalité des $(Z_t)_{t \in \mathbb{Z}}$, on a

$$\|X_t - \varphi^n X_{t-n}\|_{L^2} = \sqrt{n}. \quad (3.10)$$

D'autre part, si X est stationnaire, on peut écrire

$$\|X_t - \varphi^n X_{t-n}\|_{L^2} \leq \|X_t\|_{L^2} + \|X_{t-n}\|_{L^2} = 2\sqrt{\mu_X^2 + \gamma_X(0)} \quad (3.11)$$

En faisant tendre $n \rightarrow \infty$ dans l'équation précédente, on obtient une contradiction.

2. Cas $|\varphi| < 1$. Le filtre $\alpha := (\varphi^k \mathbf{1}_{k \geq 0})_{k \in \mathbb{Z}}$ étant dans $\ell^1(\mathbb{Z})$, le théorème de filtrage entraîne

$$\sum_{k=0}^{n-1} \varphi^k Z_{t-k} \xrightarrow[n \rightarrow \infty]{p.-s.} (F_\alpha(Z))_t = \sum_{k=0}^{\infty} \varphi^k Z_{t-k}. \quad (3.12)$$

Par ailleurs, la stationnarité de X entraîne

$$\|\varphi^n X_{t-n}\|_{L^2} = |\varphi|^n \sqrt{\mu_X^2 + \gamma_X(0)} \xrightarrow[n \rightarrow \infty]{} 0. \quad (3.13)$$

Ainsi, lorsque $n \rightarrow \infty$, le membre droit de (3.9) converge dans L^2 vers $(F_\alpha(Z))_t$. Comme le membre gauche ne dépend pas de n , on en déduit que nécessairement,

$$X_t = (F_\alpha(Z))_t. \quad (3.14)$$

Réciproquement, le processus $X := F_\alpha(Z)$ est bien défini, stationnaire et l'on a

$$X_t = \sum_{k=0}^{\infty} \varphi^k Z_{t-k} \quad (3.15)$$

$$= Z_t + \varphi \sum_{k=1}^{\infty} \varphi^{k-1} Z_{t-k} \quad (3.16)$$

$$= Z_t + \varphi \sum_{k=0}^{\infty} \varphi^k Z_{t-1-k} \quad (3.17)$$

$$= Z_t + \varphi X_{t-1}, \quad (3.18)$$

ce qui montre que X est bien solution de (3.6). Enfin, le théorème de filtrage assure que $\mu_X = (\sum_k \alpha_k) \mu_Z = 0$ et que $\gamma_X(h) = \sum_k \alpha_k \alpha_{k+|h|} = \sum_{k \geq 0} \varphi^{2k+|h|} = \frac{\varphi^{|h|}}{1-\varphi^2}$.

3. Cas $|\varphi| > 1$. On remarque que X est solution stationnaire de (3.6) si et seulement si le processus $\tilde{X} = (\tilde{X}_t)_{t \in \mathbb{Z}}$ défini $\tilde{X}_t := \varphi X_{-t}$ est solution stationnaire de l'équation

$$\tilde{X}_t = \tilde{Z}_t + \tilde{\varphi} \tilde{X}_{t-1}, \quad (3.19)$$

où l'on a posé $\tilde{Z}_t = -Z_{1-t}$ et $\tilde{\varphi} = 1/\varphi$. Notons que \tilde{Z} est encore un bruit blanc, et que l'on a $|\tilde{\varphi}| < 1$. Ainsi, le cas précédent assure que cette équation admet pour unique solution

$$\tilde{X}_t = \sum_{k=0}^{\infty} \tilde{\varphi}^k \tilde{Z}_{t-k}, \quad (3.20)$$

c'est-à-dire

$$X_t = -\sum_{k=1}^{\infty} \varphi^{-k} Z_{t+k}. \quad (3.21)$$

Notons enfin que $\mu_X = \mu_{\tilde{X}}/\varphi = 0$ et que $\gamma_X(h) = \varphi^{-2} \gamma_{\tilde{X}}(h) = \frac{\varphi^{-|h|}}{\varphi^2 - 1}$.

□

Remarque 3.1 (Solutions non-stationnaires). *Le théorème ci-dessus ne parle que des solutions stationnaires, qui sont les seules qui nous intéressent ici. En fait, à part dans le cas trivial où $\varphi = 0$, il existe une infinité de solutions non-stationnaires puisque l'on peut choisir X_0 de manière arbitraire et poser ensuite :*

$$X_t := \begin{cases} \sum_{k=0}^{t-1} \varphi^k Z_{t-k} + \varphi^t X_0 & \text{pour } t > 0; \\ -\sum_{k=1}^{-t} \varphi^{-k} Z_{t+k} + \varphi^t X_0 & \text{pour } t < 0. \end{cases} \quad (3.22)$$

Nous restreindrons toujours notre étude aux solutions stationnaires.

3.2 Étude du cas général

Nous allons à présent nous attaquer à l'équation ARMA(p,q) générale :

$$X_t = \sum_{k=1}^p \varphi_k X_{t-k} + Z_t + \sum_{k=1}^q \theta_k Z_{t-k}. \quad (3.23)$$

Pour cela, on définit deux filtres $\beta, \gamma \in \ell^1(\mathbb{Z})$ de la façon suivante :

$$\beta_k := \begin{cases} 1 & \text{si } k = 0; \\ -\varphi_k & \text{si } k \in \{1, 2, \dots, p\}; \\ 0 & \text{sinon.} \end{cases} \quad \gamma_k := \begin{cases} 1 & \text{si } k = 0; \\ \theta_k & \text{si } k \in \{1, 2, \dots, q\}; \\ 0 & \text{sinon.} \end{cases}$$

On notera Φ et Θ les représentations en séries de puissance associées (qui sont des polynômes) :

$$\Phi(z) = 1 - (\varphi_1 z + \dots + \varphi_p z^p), \quad \Theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q. \quad (3.24)$$

Ce sont les racines de module 1 de ces deux polynômes qui vont déterminer l'existence et l'unicité des solutions à l'équation (3.23), que nous appellerons désormais équation ARMA(Φ, Θ). Nous dirons qu'une racine z de Φ est **compensée** par Θ si z est aussi racine de Θ , avec une multiplicité au moins aussi grande. Par exemple, dans $\Phi(z) = (z+1)^2(z-1)$, la racine $z = 1$ est

3.2. Étude du cas général

compensée par $\Theta(z) = (z+1)(z-1)$, mais la racine $z = -1$ ne l'est pas. Lorsque Φ et Θ n'ont pas de racines en commun, nous dirons que l'équation ARMA(Φ, Θ) est **irréductible**. Dans le cas général, il sera utile de se débarrasser des racines communes en écrivant

$$\Phi = D\tilde{\Phi}, \quad \Theta = D\tilde{\Theta}, \quad (3.25)$$

où D désigne le PGCD de Φ, Θ . L'équation ARMA($\tilde{\Phi}, \tilde{\Theta}$) obtenue par cette procédure est appelée équation **réduite**. Elle est irréductible par construction. Avec cette terminologie, voici le théorème de résolution générale des équations ARMA.

Théorème 3.2 (Résolution générale de l'équation ARMA(Φ, Θ)). *Il y a trois cas.*

Cas 1. *Si Φ n'a pas de racines de module 1, alors l'équation ARMA(Φ, Θ) admet une unique solution stationnaire, et celle-ci est linéaire : c'est $X = F_\alpha(Z)$, où $\alpha \in \ell^1(\mathbb{Z})$ est caractérisé par*

$$\forall z \in \mathbb{U}, \quad P_\alpha(z) = \frac{\Theta(z)}{\Phi(z)}. \quad (3.26)$$

Cas 2. *Si Φ a des racines de module 1, et que l'une de ces racines de module 1 n'est pas compensée par Θ , alors l'équation ARMA(Φ, Θ) n'admet aucune solution stationnaire.*

Cas 3. *Si Φ a des racines de module 1, mais que celles-ci sont toutes compensées par Θ , alors l'équation ARMA(Φ, Θ) admet une infinité de solutions stationnaires, mais seule une d'entre elles est linéaire : celle qui vérifie l'équation réduite ARMA($\tilde{\Phi}, \tilde{\Theta}$).*

Démonstration. Le point crucial consiste à remarquer que l'équation ARMA(Φ, Θ) s'écrit

$$F_\beta(X) = F_\gamma(Z), \quad (3.27)$$

où les filtres β, γ sont polynômiaux et ont pour séries de puissances Φ et Θ , respectivement.

Cas 1. Puisque Φ n'a pas de racine de module 1, le théorème 2.3 assure que β admet un inverse $\beta^{-1} \in \ell^1(\mathbb{Z})$. En appliquant $F_{\beta^{-1}}$ aux deux membres de l'équation ci-dessus, il vient

$$X = F_{\beta^{-1}\star\gamma}(Z). \quad (3.28)$$

Réciproquement, ce processus stationnaire est bien solution de (3.27), comme on le voit en appliquant F_β des deux côtés. Enfin, l'égalité $\beta \star \alpha = \gamma$ s'écrit en séries de puissances :

$$\forall z \in \mathbb{U}, \quad \Phi(z)P_\alpha(z) = \Theta(z). \quad (3.29)$$

Comme Φ n'a pas de racine dans \mathbb{U} , on peut tranquillement diviser par $\Phi(z)$ des deux côtés pour obtenir la formule cherchée.

Cas 2. Nous verrons cela au chapitre suivant, quand nous aurons parlé de densité spectrale.

Cas 3. Pour un processus de la forme $X = F_\alpha(Z)$ avec $\alpha \in \ell^1(\mathbb{Z})$, l'équation (3.27) se réécrit :

$$\begin{aligned} F_\alpha(X) = F_\gamma(Z) &\iff F_\beta(F_\alpha(Z)) = F_\gamma(Z) \\ &\iff F_{\beta\star\alpha}(Z) = F_\gamma(Z) \\ &\iff \beta \star \alpha = \gamma \quad (\text{voir l'exercice 2.2}) \\ &\iff \forall z \in \mathbb{U}, \Phi(z)P_\alpha(z) = \Theta(z) \\ &\iff \forall z \in \mathbb{U}, D(z)\tilde{\Phi}(z)P_\alpha(z) = D(z)\tilde{\Theta}(z) \\ &\iff \forall z \in \mathbb{U}, \tilde{\Phi}(z)P_\alpha(z) = \tilde{\Theta}(z). \end{aligned}$$

La simplification par $D(z)$ à la dernière ligne n'est a priori valable que pour les points $z \in \mathbb{U}$ tels que $D(z) \neq 0$, mais la conclusion s'étend en fait bien à \mathbb{U} tout entier par continuité. Maintenant, l'hypothèse de compensation des racines garantit que $\tilde{\Phi}(z)$ ne s'annule pas sur \mathbb{U} , de sorte que l'on s'est ramené au cas 1. L'existence et l'unicité de la solution linéaire $X = F_\alpha(Z)$ en découle aussitôt. Enfin, nous vérifierons en exercice que si H est un processus harmonique décorréllé de X et dont la fréquence θ est choisie telle que $\Phi(e^{i\theta}) = 0$, alors le processus stationnaire $X + \lambda H$ est solution pour tout $\lambda \in \mathbb{R}$. □

Remarque 3.2 (Cas irréductible). *Lorsque l'équation ARMA est irréductible, le cas 3 ne se produit pas. C'est vrai par exemple pour les équations MA(q) ($\Phi(z) = 1$) et AR(p) ($\Theta(z) = 1$). En particulier, le lecteur est invité à retrouver les résultats obtenus précédemment dans le cas particulier AR(1).*

Remarque 3.3 (Causalité). *Dans les cas 1 et 3, la solution linéaire sera causale si et seulement si $\tilde{\Phi}$ n'a pas de racine dans le disque unité $\{z \in \mathbb{C} : |z| \leq 1\}$. En effet, un processus linéaire $X = F_\alpha(Z)$ est causal si et seulement si la série de puissances $P_\alpha(z)$ est absolument convergente dans le disque unité. Ici, P_α est la fraction rationnelle irréductible $\tilde{\Phi}/\tilde{\Theta}$, qui est développable en série entière partout où le dénominateur ne s'annule pas.*

Remarque 3.4 (Fraction rationnelle associée à l'équation ARMA). *Deux équations ARMA qui ont la même fraction rationnelle réduite $\frac{\tilde{\Theta}}{\tilde{\Phi}}$ admettent les mêmes solutions linéaires.*

Exemple 3.1. Résoudre l'équation $X_t = Z_t + 2Z_{t-1} + \frac{X_{t-1}}{2}$, et dire tout ce qu'on peut sur la solution.

Exemple 3.2. De même pour l'équation $X_t = Z_t + 2Z_{t-1} - \frac{5}{3}X_{t-1} + \frac{2}{3}X_{t-2}$.

3.3 Décorrélation rapide

Nous allons voir que les processus ARMA ont la *mémoire courte* : leur passé lointain a très peu d'influence sur leur valeur présente. La définition suivante formalise cette notion vague.

Définition 3.2 (Décorrélation rapide). *Une suite réelle $\alpha = (\alpha_k)_{k \in \mathbb{Z}}$ est dite à **décroissance rapide** s'il existe des constantes $C < \infty$ et $\rho < 1$ telles que*

$$\forall k \in \mathbb{Z}, \quad |\alpha_k| \leq C\rho^{|k|}. \tag{3.30}$$

*Un processus stationnaire X est dit à **décroissance rapide** si sa fonction d'auto-covariance $(\gamma_X(h))_{h \in \mathbb{Z}}$ forme une suite à décroissance rapide.*

Un bruit blanc est bien-sûr un exemple extrême de processus à décorrélation rapide. Le but de ce chapitre est de démontrer le résultat beaucoup plus général suivant.

Théorème 3.3 (Décorrélation rapide des processus ARMA). *La solution stationnaire d'une équation ARMA bien posée est toujours un processus à décorrélation rapide.*

Pour démontrer ce résultat, commençons par rappeler que la solution X d'une équation ARMA bien posée est de la forme $X = F_\alpha(Z)$ avec Z un bruit blanc et $\alpha = \alpha_\varphi^{-1} \star \alpha_\theta$, où $\alpha_\theta, \alpha_\varphi$ sont les filtres associés aux polynômes Φ et Θ . Voici une condition suffisante simple pour que le résultat d'un filtrage soit à décorrélation rapide.

Lemme 3.1 (Décorrélation rapide et filtrage). *Soient Z un processus à décorrélation rapide (par exemple un bruit blanc) et α un filtre à décroissance rapide. Alors $X := F_\alpha(Z)$ est à décorrélation rapide.*

3.3. Décorrélation rapide

Démonstration. En appliquant l'hypothèse, on a pour tout $h \in \mathbb{Z}$,

$$\begin{aligned} |\gamma_X(h)| &= \left| \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \alpha_j \alpha_k \gamma_Z(h+j-k) \right| \\ &\leq C^3 \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \rho^{|j|+|k|+|h+j-k|} \\ &\leq C^3 \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \rho^{\frac{|j|+|k|+|h|}{2}} \\ &\leq \frac{4C^3}{(1-\sqrt{\rho})^2} (\sqrt{\rho})^{|h|}. \end{aligned}$$

où l'on a utilisé l'inégalité $|j| + |k| + |h+j-k| \geq \max(|j| + |k|, |h|) \geq \frac{|j|+|k|+|h|}{2}$. \square

Puisque le bruit blanc Z est à décroissance rapide, il nous suffit pour démontrer le théorème 3.3, d'établir que le filtre $\alpha = \alpha_\varphi^{-1} \star \alpha_\theta$ est à décroissance rapide. Comme α_φ et α_θ sont des filtres polynomiaux, cela découlera simplement du résultat général suivant.

Proposition 3.1 (Stabilité des filtres à décroissance rapide).

- (i) la convolution de deux filtres à décroissance rapide est un filtre à décroissance rapide ;
- (ii) un filtre polynomial est toujours à décroissance rapide ;
- (iii) l'inverse d'un filtre polynomial, lorsqu'il existe, est lui aussi à décroissance rapide.

Pour démontrer cette proposition, nous utiliserons le résultat suivant, qui caractérise les filtres à décroissance rapide en termes de série de puissances.

Lemme 3.2 (Caractérisation de la décroissance rapide en termes de série de puissances). *Un filtre α est à décroissance rapide si et seulement si sa série de puissance est absolument convergente dans une couronne ouverte contenant \mathbb{U} : il existe $r < 1 < R$ tels que*

$$r < |z| < R \implies \sum_{k \in \mathbb{Z}} |\alpha_k z^k| < \infty. \quad (3.31)$$

Démonstration. Si α est à décroissance rapide, alors il existe $C < \infty$ et $\rho < 1$ tels que

$$\forall k \in \mathbb{Z}, \quad |\alpha_k| \leq C\rho^{|k|}. \quad (3.32)$$

Alors, pour tout $z \in \mathbb{C}$, on a

$$\begin{aligned} |z| < \frac{1}{\rho} &\implies \sum_{k=0}^{\infty} |\alpha_k z^k| \leq C \sum_{k=0}^{\infty} |z|^k \rho^k = \frac{C}{1-|z|\rho} < \infty \\ |z| > \rho &\implies \sum_{k=0}^{\infty} |\alpha_{-k} z^{-k}| \leq C \sum_{k=0}^{\infty} \left(\frac{\rho}{|z|}\right)^k = \frac{C}{1-\frac{\rho}{|z|}} < \infty. \end{aligned}$$

En combinant ces deux observations, on conclut que la série $\sum_{k \in \mathbb{Z}} \alpha_k z^k$ est absolument convergente dans la couronne $\rho < |z| < \frac{1}{\rho}$, donc (ii) est vérifié avec $r = \rho$ et $R = 1/\rho$. Réciproquement, si la série $\sum_{k \in \mathbb{Z}} \alpha_k z^k$ est absolument convergente dans une couronne $\{z \in \mathbb{C} : r < |z| < R\}$ avec $r < 1 < R$, alors en choisissant $\rho < 1$ suffisamment proche de 1 pour que $r < \rho, \rho^{-1} < R$, on a

$$\sup_{k \geq 0} |\alpha_k| \rho^{-k} \leq \sum_{k \in \mathbb{Z}} |\alpha_k| \rho^{-k} < \infty \quad (3.33)$$

$$\sup_{k \geq 0} |\alpha_{-k}| \rho^k \leq \sum_{k \in \mathbb{Z}} |\alpha_k| \rho^k < \infty, \quad (3.34)$$

ce qui montre que $|\alpha_k| \leq C\rho^{|k|}$ pour tout $k \in \mathbb{Z}$ avec $C = \max(\sum_{k \in \mathbb{Z}} |\alpha_k| \rho^{-k}, \sum_{k \in \mathbb{Z}} |\alpha_k| \rho^k)$. \square

Démonstration de la proposition 3.1. On utilise la caractérisation ci-dessus :

(i) Soient α, β des filtres à décroissance rapide. Par inégalité triangulaire, on a

$$\sum_{k \in \mathbb{Z}} |(\alpha \star \beta)_k z^k| \leq \sum_{k \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} |\alpha_n z^n \beta_{k-n} z^{k-n}| = \left(\sum_{n \in \mathbb{Z}} |\alpha_n z^n| \right) \times \left(\sum_{k \in \mathbb{Z}} |\beta_k z^k| \right).$$

Chacune des séries de droite est convergente dans une couronne ouverte contenant \mathbb{U} , et l'intersection de ces couronnes est encore une couronne ouverte contenant \mathbb{U} .

- (ii) Si α est un filtre polynomial, alors sa série de puissances est en fait un polynôme, donc elle est trivialement absolument convergente dans \mathbb{C} tout entier.
- (iii) Si α est un filtre polynomial inversible, alors P_α n'a pas de racines de module 1. On peut donc trouver $r < 1 < R$ tels que P_α n'a pas de racines dans la couronne $\mathcal{C} := \{z \in \mathbb{C} : r < |z| < R\}$. La fraction rationnelle $P_{\alpha^{-1}}(z) = 1/P_\alpha(z)$ est alors développable en série de puissances dans \mathcal{C} , car c'est une combinaison linéaire de fractions de la forme $\frac{1}{(z-a)^k}$ avec k entier et a racine de P_α , et ces fractions sont développables en série de puissances dans \mathcal{C} . En effet, pour $|a| > R > |z|$, on a $|z/a| < 1$, de sorte que l'on peut écrire

$$\frac{1}{(z-a)^k} = \frac{1}{(-a)^k \left(1 - \frac{z}{a}\right)^k} = \frac{1}{(-a)^k} \sum_{n=0}^{\infty} \binom{n+k-1}{k-1} \left(\frac{z}{a}\right)^n,$$

tandis que pour $|a| < r < |z|$, on a $|a/z| < 1$, de sorte que l'on peut écrire

$$\frac{1}{(z-a)^k} = \frac{1}{z^k \left(1 - \frac{a}{z}\right)^k} = \sum_{n=0}^{\infty} \binom{n+k-1}{k-1} \frac{a^n}{z^{n+k}}.$$

□

Chapitre 4

Analyse spectrale

4.1 Rappels sur les coefficients de Fourier

Dans tout ce paragraphe, ν désigne une **mesure positive finie** sur $[-\pi, \pi]$ (en d'autres termes, ν est une brave mesure de probabilité sur $[-\pi, \pi]$, que l'on a ensuite simplement multipliée par une constante positive arbitraire). Nous ferons toujours, sans le re-préciser, l'hypothèse que les conditions au bord sont périodiques :

$$\nu(\{-\pi\}) = \nu(\{\pi\}). \quad (4.1)$$

Il s'agit néanmoins d'un point mineur, que le lecteur pourra ignorer en première lecture.

Définition 4.1 (Coefficient de Fourier). *Le **coefficient de Fourier** de fréquence $h \in \mathbb{Z}$ de ν est*

$$\widehat{\nu}(h) := \int_{-\pi}^{\pi} e^{ihu} \nu(du) \in \mathbb{C}.$$

La **transformée de Fourier** de ν est simplement la suite des coefficients de Fourier $\widehat{\nu} = (\widehat{\nu}(h))_{h \in \mathbb{Z}}$.

Remarque 4.1 (Fréquence nulle). *On a en particulier $\widehat{\nu}(0) = \nu([-\pi, \pi])$.*

Exercice 4.1 (Quelques exemples). *Calculer la transformée de Fourier des mesures suivantes :*

1. $\nu = \delta_0$.
2. $\nu = \frac{\delta_{\theta} + \delta_{-\theta}}{2}$, $\theta \in [-\pi, \pi]$.
3. ν est la mesure uniforme sur $[-\pi, \pi]$.

Comme nous allons le voir, la suite des coefficients de Fourier $\widehat{\nu}$ fournit une excellente représentation de la mesure ν . Tout d'abord, cette représentation est fidèle, au sens où elle n'induit pas de perte d'information : la suite $\widehat{\nu} := (\widehat{\nu}(h))_{h \in \mathbb{Z}}$ caractérise la mesure ν .

Théorème 4.1 (Injectivité de la transformée de Fourier). *L'application $\nu \mapsto \widehat{\nu}$ est injective.*

Démonstration. Soient ν_1, ν_2 mesures positives finies sur $[-\pi, \pi]$ telles que $\widehat{\nu}_1 = \widehat{\nu}_2$. Alors l'égalité

$$\int_{-\pi}^{\pi} g(u) \nu_1(du) = \int_{-\pi}^{\pi} g(u) \nu_2(du), \quad (4.2)$$

est vérifiée, successivement, pour

- Toute fonction trigonométrique élémentaire $g(u) = e^{ihu}$ ($h \in \mathbb{Z}$), par hypothèse.
- Tout polynôme trigonométrique $g(u) = \sum_{h=-n}^n c_h e^{ihu}$, par linéarité de (4.2).

4.1. Rappels sur les coefficients de Fourier

- Toute fonction $g: [-\pi, \pi] \rightarrow \mathbb{R}$ continue et telle que $g(\pi) = g(-\pi)$, parce qu'une telle fonction est la limite uniforme d'une suite polynômes trigonométriques (théorème de Stone-Weierstrass).
- Toute indicatrice d'un intervalle $]a, b[\subseteq [-\pi, \pi]$, parce qu'une telle fonction est la limite d'une suite croissante de fonctions continues positives telle que $g(\pi) = g(-\pi) = 0$.

Le dernier point nous dit que $\nu_1(]a, b]) = \nu_2(]a, b])$ pour tout intervalle $]a, b[\subseteq [-\pi, \pi]$, et le lemme des classes monotones nous permet de déduire que ν_1 et ν_2 coïncident sur tout borélien de $] - \pi, \pi[$. Pour conclure, il ne nous reste plus qu'à montrer que ν_1 et ν_2 coïncident au bord, c'est-à-dire que $\nu_1(\{x\}) = \nu_2(\{x\})$ pour $x = \pm\pi$. C'est ici (et seulement ici) qu'on utilise l'hypothèse (4.1), qui permet d'écrire pour $i \in \{1, 2\}$,

$$\nu_i(\pi) = \nu_i(-\pi) = \frac{1}{2} (\widehat{\nu}_i(0) - \nu_i(] - \pi, \pi])).$$

Or nous avons montré que $\nu_1(] - \pi, \pi]) = \nu_2(] - \pi, \pi])$, et nous savons par ailleurs que $\widehat{\nu}_1(0) = \widehat{\nu}_2(0)$ par hypothèse, ce qui termine la preuve. \square

Remarque 4.2 (Conditions périodique au bord). *L'hypothèse (4.1) est réellement nécessaire ici. En effet, les mesures $\nu_+ = \delta_{+\pi}$ et $\nu_- = \delta_{-\pi}$ sont distinctes mais ont les mêmes transformée de Fourier.*

En conséquence, on doit pouvoir reconstruire entièrement une mesure ν à partir de ses coefficients de Fourier. Le célèbre théorème suivant donne une formule explicite simple pour cela dans le cas particulier où les coefficients de Fourier sont absolument sommables.

Théorème 4.2 (Formule d'inversion de Lévy). *Si $\widehat{\nu} \in \ell^1(\mathbb{Z})$, alors la mesure ν admet une densité continue explicite, donnée par la formule*

$$\forall u \in [-\pi, \pi], \quad f(u) := \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \widehat{\nu}(h) e^{-iuh}.$$

(En particulier, f est à valeurs réelles positives, ce qui n'est pas évident a priori.)

Démonstration du Théorème 4.2. D'abord, l'hypothèse $\widehat{\nu} \in \ell^1(\mathbb{Z})$ assure que la fonction f est bien définie et continue sur \mathbb{R} . De plus, pour tout $h_0 \in \mathbb{Z}$, on a par Fubini

$$\int_{-\pi}^{\pi} e^{ih_0 u} f(u) \, du = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \widehat{\nu}(h) \int_{-\pi}^{\pi} e^{i(h_0 - h)u} \, du = \widehat{\nu}(h_0).$$

En d'autre termes, on a vérifié l'égalité

$$\int_{-\pi}^{\pi} g(u) f(u) \, du = \int_{-\pi}^{\pi} g(u) \nu(du)$$

lorsque $g: [-\pi, \pi] \rightarrow \mathbb{C}$ est une fonction trigonométrique élémentaire. Comme dans la preuve précédente, l'égalité s'étend systématiquement aux fonctions indicatrices d'intervalles. Ainsi, pour tout $a, b \in [-\pi, \pi]$, on peut prendre $g = \mathbf{1}_{[a,b]}$ pour obtenir

$$\nu([a, b]) = \int_a^b f(u) \, du.$$

Cela entraîne aisément que f est réelle positive en tout point. Comme les mesures $\nu(du)$ et $f(u)du$ coïncident sur les intervalles, elles doivent être égales, par le lemme des classes monotones. \square

Exemple 4.1 (Cas simples). Reprenons nos trois exemples précédents : dans le cas où ν est la loi uniforme sur $]-\pi, \pi[$, on a $\widehat{\nu}(h) = \mathbf{1}_{h=0}$ et la formule ci-dessus est clairement vérifiée. En revanche, lorsque $\nu = \delta_0$ ou $\nu = \frac{\delta_\theta + \delta_{-\theta}}{2}$, la mesure ν n'a pas de densité. Ce n'est pas une contradiction, puisque l'hypothèse $\widehat{\nu} \in \ell^1(\mathbb{Z})$ n'est pas vérifiée.

La seconde propriété fondamentale des coefficients de Fourier est qu'ils capturent la convergence faible d'une manière simple. Rappelons que la **convergence faible** $\nu_n \rightarrow \nu$ signifie ici

$$\int_{-\pi}^{\pi} g(u) \nu_n(du) \xrightarrow{n \rightarrow \infty} \int_{-\pi}^{\pi} g(u) \nu(du), \quad (4.3)$$

pour toute fonction continue $g: [-\pi, \pi] \rightarrow \mathbb{C}$ (notons que g est nécessairement bornée).

Théorème 4.3 (Théorème de continuité de Lévy). Soit $(\nu_n)_{n \geq 1}$ une suite de mesures sur $[-\pi, \pi]$. Alors les deux conditions suivantes sont équivalentes :

- (i) La suite $(\nu_n)_{n \geq 1}$ converge faiblement vers une mesure finie ν .
- (ii) La suite $(\widehat{\nu}_n(h))_{h \in \mathbb{Z}}$ converge dans \mathbb{C} pour tout $h \in \mathbb{Z}$.

De plus, on a alors nécessairement $\widehat{\nu}(h) = \lim_{n \rightarrow \infty} \widehat{\nu}_n(h)$ pour tout $h \in \mathbb{Z}$.

Démonstration. Supposons d'abord (i), et fixons $h \in \mathbb{Z}$. Alors en prenant $g(u) = e^{ihu}$ dans (4.3), on conclut que $(\widehat{\nu}_n(h))_{n \geq 1}$ converge vers $\widehat{\nu}(h)$. Réciproquement, supposons (ii). Notons que la masse totale $\nu_n([-\pi, \pi]) = \widehat{\nu}_n(0)$ est convergente, donc bornée. Comme de plus $[-\pi, \pi]$ est compact, on en déduit que la suite $(\nu_n)_{n \geq 1}$ est tendue. Enfin, si $(\nu_{\varphi_1(n)})_{n \geq 1}$ et $(\nu_{\varphi_2(n)})_{n \geq 1}$ sont deux sous-suites qui admettent des limites ν_1 et ν_2 , alors la première partie de la preuve donne

$$\begin{aligned} \widehat{\nu}_1(h) &= \lim_{n \rightarrow \infty} \widehat{\nu}_{\varphi_1(n)}(h) \\ \widehat{\nu}_2(h) &= \lim_{n \rightarrow \infty} \widehat{\nu}_{\varphi_2(n)}(h). \end{aligned}$$

Mais l'hypothèse (ii) assure que ces deux quantités sont égales, et l'injectivité des coefficients de Fourier permet de conclure que $\nu_1 = \nu_2$. Cela garantit que $(\nu_n)_{n \geq 1}$ converge. \square

Attention, les coefficients de Fourier sont en général des nombres complexes. Le cas où ils sont réels est un cas particulier important, et très bien compris.

Proposition 4.1 (Mesures paires). Les propriétés suivantes sont équivalentes :

- (i) $\nu(B) = \nu(-B)$ pour tout $B \in \mathcal{B}([-\pi, \pi])$.
- (ii) $\int_{-\pi}^{\pi} g(u) \nu(du) = \int_{-\pi}^{\pi} g(-u) \nu(du)$ pour toute fonction $g: [-\pi, \pi] \rightarrow \mathbb{C}$ mesurable bornée.
- (iii) $\widehat{\nu}(h) = \widehat{\nu}(-h)$ pour tout $h \in \mathbb{Z}$.
- (iv) Pour tout $h \in \mathbb{Z}$,

$$\widehat{\nu}(h) = \int_{-\pi}^{\pi} \cos(hu) \nu(du).$$

- (v) La suite $\widehat{\nu}$ est réelle : $\widehat{\nu}(h) \in \mathbb{R}$ pour tout $h \in \mathbb{Z}$.

Dans ce cas, nous dirons que la mesure ν est **paire**.

Démonstration. L'implication (i) \implies (ii) découle du fait que toute fonction mesurable bornée $g: [-\pi, \pi] \rightarrow \mathbb{C}$ est la limite d'une suite bornée de fonctions étagées. L'implication (ii) \implies (iii) s'obtient en prenant $g(u) = e^{ihu}$. L'implication (iii) \implies (iv) découle du fait que \cos est paire et \sin impaire. L'implication (iv) \implies (v) est évidente. Enfin, l'implication (v) \implies (i) découle de l'injectivité des coefficients de Fourier appliquée aux mesures $B \mapsto \nu(B)$ et $B \mapsto \nu(-B)$. \square

Exercice 4.2. Soit $\rho \in (-1, 1)$. Trouver la mesure finie ν telle que $\widehat{\nu}(h) = \rho^{|h|}$ pour tout $h \in \mathbb{Z}$.

4.2 Le théorème de représentation spectrale d'Herglotz

La suite des coefficients de Fourier d'une mesure a des propriétés bien particulières. Lesquelles exactement? C'est précisément l'objet du résultat célèbre suivant.

Théorème 4.4 (Représentation spectrale d'Herglotz). *Soit $\gamma = (\gamma(h))_{h \in \mathbb{Z}}$ une suite de nombres complexes. Alors les conditions suivantes sont équivalentes :*

- (i) γ est la transformée de Fourier d'une mesure finie sur $[-\pi, \pi]$.
- (ii) γ est de **type positif** au sens complexe : pour tout $n \geq 1$ et tout $(z_1, \dots, z_n) \in \mathbb{C}^n$,

$$\sum_{j=1}^n \sum_{k=1}^n z_j \bar{z}_k \gamma(j-k) \in \mathbb{R}_+. \quad (4.4)$$

Démonstration. Supposons (i) : il existe une mesure finie ν sur $[-\pi, \pi]$ telle que pour tout $h \in \mathbb{Z}$,

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ihu} \nu(du).$$

Alors pour tout $n \geq 1$ et tout $\theta \in \mathbb{C}^n$, on a

$$\begin{aligned} \sum_{j=1}^n \sum_{k=1}^n z_j \bar{z}_k \gamma(j-k) &= \int_{-\pi}^{\pi} \left(\sum_{j=1}^n z_j e^{iju} \right) \left(\sum_{k=1}^n \bar{z}_k e^{-iku} \right) \nu(du) \\ &= \int_{-\pi}^{\pi} \left| \sum_{j=1}^n z_j e^{iju} \right|^2 \nu(du) \in \mathbb{R}_+. \end{aligned}$$

Ainsi, γ est bien de type positif. Réciproquement, supposons que γ soit de type positif. Pour tout $n \geq 1$ et tout $u \in [-\pi, \pi]$, posons

$$\begin{aligned} f_n(u) &= \frac{1}{2\pi n} \sum_{j=1}^n \sum_{k=1}^n e^{i(k-j)u} \gamma(j-k) \\ &= \frac{1}{2\pi n} \sum_{\ell=-n}^n \text{card} \{ (j, k) \in \{1, \dots, n\}^2 : j-k = \ell \} e^{-i\ell u} \gamma(\ell) \\ &= \frac{1}{2\pi} \sum_{\ell=-n}^n \left(1 - \frac{|\ell|}{n} \right) e^{-i\ell u} \gamma(\ell) \end{aligned}$$

En prenant $z_k = e^{-iuk}$ dans la définition de type positif (4.4), on voit que $f_n(u) \in \mathbb{R}_+$. De plus, f_n est continue (polynôme trigonométrique) donc intégrable sur $[-\pi, \pi]$. On peut donc considérer la mesure finie $\nu_n(u) := f_n(u) du$ sur $[-\pi, \pi]$. Calculons ses coefficients de Fourier :

$$\begin{aligned} \widehat{\nu}_n(h) &= \int_{-\pi}^{\pi} e^{ihu} f_n(u) du \\ &= \frac{1}{2\pi} \sum_{\ell=-n}^n \left(1 - \frac{|\ell|}{n} \right) \gamma(\ell) \int_{-\pi}^{\pi} e^{i(h-\ell)u} du \\ &= \left(1 - \frac{|h|}{n} \right) \gamma(h) \mathbf{1}_{|h| \leq n}. \end{aligned}$$

Ainsi, $\widehat{\nu}_n(h) \rightarrow \gamma(h)$ lorsque $n \rightarrow \infty$, et ce pour tout $h \in \mathbb{Z}$. D'après le théorème de continuité de Lévy, on conclut que la suite $(\nu_n)_{n \geq 1}$ admet une limite faible finie ν , et que $\widehat{\nu} = \gamma$. \square

Remarque 4.3 (Simplification dans le cas réel). Lorsque la suite γ est réelle, il est tentant de restreindre la condition (4.4) au cas où le vecteur θ est réel (on dit alors que γ est de type positif **au sens réel**). Attention cependant : cette simplification n'est valide que si la suite γ est paire. Plus précisément, une suite réelle est de type positif au sens complexe si et seulement si elle est paire et de type positif au sens réel.

Corollaire 4.1 (Cas sommable). Soit $\gamma = (\gamma(h))_{h \in \mathbb{Z}}$ une suite de nombres complexes absolument sommable. Alors γ est de type positif si et seulement si la fonction $f : [-\pi, \pi] \rightarrow \mathbb{C}$ définie par

$$f(u) := \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma(h) e^{-ihu} \quad (4.5)$$

est à valeurs dans \mathbb{R}_+ . De plus, dans ce cas, on a $\gamma = \widehat{f}$.

Démonstration. Si γ est de type positif, le théorème de Herglotz assure que l'on a $\gamma = \widehat{\nu}$ pour une certaine mesure finie ν . De plus, on a $\widehat{\nu} \in \ell^1(\mathbb{Z})$ par hypothèse, donc la formule d'inversion de Lévy assure que ν admet pour densité la fonction f définie par (4.5). En particulier, $f \geq 0$. Réciproquement, supposons que la fonction f définie par (4.5) soit positive. Notons que f est aussi continue. De plus, pour tout $k \in \mathbb{Z}$, on a

$$\int_{-\pi}^{\pi} e^{iku} f(u) du = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma(h) \int_{-\pi}^{\pi} e^{i(k-h)u} du = \gamma(k).$$

Ainsi γ est la transformée de Fourier de $\nu(du) := f(u)du$, donc γ est bien de type positif. \square

4.3 Mesure spectrale d'un processus stationnaire

Nous pouvons enfin revenir aux processus stationnaires et définir l'objet clé du chapitre.

Définition 4.2 (Mesure spectrale). Soit X un processus stationnaire. Sa fonction d'auto-covariance γ_X étant réelle et de type positif, le théorème de Herglotz fournit la représentation spectrale

$$\forall h \in \mathbb{Z}, \quad \gamma_X(h) = \int_{-\pi}^{\pi} \cos(hu) \nu_X(du),$$

pour une unique mesure finie ν_X sur $[-\pi, \pi]$, qui est paire et que l'on appellera désormais la **mesure spectrale** du processus X . Lorsque ν_X admet une densité f_X , on appellera f_X la **densité spectrale** de X :

$$\forall h \in \mathbb{Z}, \quad \gamma_X(h) = \int_{-\pi}^{\pi} \cos(hu) f_X(u) du.$$

Comme nous allons le voir, la mesure spectrale d'un processus X constitue une représentation extrêmement pratique de la fonction d'auto-covariance γ_X .

Remarque 4.4 (Masse totale). Attention ! ν_X n'est une mesure de probabilité que si la variance de X est 1. En effet, on a toujours $\nu_X([-\pi, \pi]) = \gamma_X(0) = \text{Var}(X_0)$.

Exemple 4.2 (Quelques mesures spectrales élémentaires). Déterminer ν_X dans les cas suivants :

1. $X = (X_t)_{t \in \mathbb{Z}}$ est défini par $X_t = A$ pour tout $t \in \mathbb{Z}$, où A est une v.a.r. centrée réduite.
2. X est un processus harmonique de fréquence $\theta \in [-\pi, \pi]$.
3. $X \sim \text{BB}(0, 1)$.

Pour des processus plus sophistiqués, la détermination pratique de la mesure spectrale pourra se faire grâce à la formule très utile suivante.

4.3. Mesure spectrale d'un processus stationnaire

Proposition 4.2 (Calcul pratique de la densité spectrale). *Soit X un processus stationnaire tel que γ_X est sommable. Alors X admet une densité spectrale, donnée par la formule*

$$f_X(u) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_X(h) e^{-ihu}.$$

Démonstration. C'est une conséquence de la formule d'inversion de Lévy ou du corollaire 4.1. \square

Exemple 4.3 (Quelques densités spectrales). *Dans chacun des cas suivants, calculer f_X .*

1. $X \sim \text{BB}(0, 1)$.
2. X est un processus $\text{MA}(1)$.
3. X est un processus $\text{AR}(1)$.

Le théorème fondamental suivant résume à lui seul tout l'intérêt d'utiliser la représentation spectrale ν_X plutôt que la fonction d'auto-covariance γ_X : vu sous l'angle spectral, l'effet d'un filtrage devient beaucoup plus simple à comprendre.

Théorème 4.5 (Filtrage spectral). *Soit X un processus stationnaire et soit $\alpha \in \ell^1(\mathbb{Z})$ un filtre. Alors le processus filtré $Y := F_\alpha(X)$ vérifie*

$$\nu_Y(\mathrm{d}u) = |P_\alpha(e^{iu})|^2 \nu_X(\mathrm{d}u).$$

En particulier, si X admet une densité spectrale, alors Y aussi et l'on a

$$f_Y(u) = |P_\alpha(e^{iu})|^2 f_X(u). \quad (4.6)$$

Démonstration. D'après le théorème de filtrage des processus stationnaires, on a pour tout $h \in \mathbb{Z}$,

$$\begin{aligned} \gamma_Y(h) &= \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \alpha_j \alpha_k \gamma_X(h + k - j) \\ &= \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \alpha_j \alpha_k \int_{-\pi}^{\pi} e^{i(h+k-j)u} \nu_X(\mathrm{d}u) \\ &= \int_{-\pi}^{\pi} e^{ihu} \left(\sum_{k \in \mathbb{Z}} \alpha_k e^{iku} \right) \left(\sum_{j \in \mathbb{Z}} \alpha_j e^{-iju} \right) \nu_X(\mathrm{d}u) \\ &= \int_{-\pi}^{\pi} e^{ihu} |P_\alpha(e^{iu})|^2 \nu_X(\mathrm{d}u). \end{aligned}$$

Par unicité de la représentation spectrale, on en déduit que $\nu_Y(\mathrm{d}u) = |P_\alpha(e^{iu})|^2 \nu_X(\mathrm{d}u)$. \square

Exemple 4.4 (Densité spectrale des processus linéaires). *On a vu que le bruit blanc $Z \sim \text{BB}(0, 1)$ admet pour densité spectrale la densité uniforme :*

$$f_Z(u) = \frac{1}{2\pi}.$$

On en déduit que pour $\alpha \in \ell^1(\mathbb{Z})$, le processus $Y = F_\alpha(Z)$ admet une densité spectrale donnée par

$$f_Y(u) = \frac{1}{2\pi} |P_\alpha(e^{iu})|^2.$$

Remarque 4.5 (Atomes). *Le fait que les processus linéaires admettent une densité spectrale montre en particulier que le processus harmonique ($\nu_X = \frac{\delta_\theta + \delta_{-\theta}}{2}$) ne pourra jamais s'obtenir par filtrage d'un bruit blanc. Il en est de même de processus dont la mesure spectrale a une partie atomique.*

4.4 Retour sur les processus ARMA

Pour illustrer la puissance de l'analyse spectrale, nous nous penchons à présent sur le cas des processus ARMA. Rappelons que ces processus vérifient une équation de la forme

$$X_t = Z_t + \sum_{k=1}^p \phi_k X_{t-k} + \sum_{k=1}^q \theta_k Z_{t-k},$$

où Z est un bruit blanc standard. Comme d'habitude, nous notons $\Phi(z) = 1 - \sum_{k=1}^p \phi_k z^k$ et $\Theta(z) = 1 + \sum_{k=1}^q \theta_k z^k$ les polynômes associés. La formule suivante généralise considérablement les exemples MA(1) et AR(1) étudiés plus haut.

Théorème 4.6 (Densité spectrale des processus ARMA). *Si Φ n'a pas de racines complexes de module 1, alors l'unique solution stationnaire X de l'équation ARMA(Φ, Θ) admet pour densité spectrale*

$$f_X(u) = \frac{1}{2\pi} \left| \frac{\Theta(e^{iu})}{\Phi(e^{iu})} \right|^2.$$

Démonstration. C'est une conséquence immédiate du théorème de filtrage spectral. \square

Théorème 4.7 (Retour sur les équations ARMA mal posées). *Si Φ a une racine de module 1 qui n'est pas compensée par Θ , alors l'équation ARMA(Φ, Θ) n'admet pas de solution stationnaire.*

Démonstration. Supposons que X soit une solution stationnaire de l'équation ARMA(Φ, Θ). En voyant cette équation comme une équation de filtrage, et en appliquant le théorème de filtrage spectral, on obtient l'égalité des deux mesures spectrales suivantes :

$$\left| \Phi(e^{i\theta}) \right|^2 \nu_X(d\theta) = \frac{1}{2\pi} \left| \Theta(e^{i\theta}) \right|^2 d\theta. \quad (4.7)$$

Fixons un point $e^{i\theta_0}$ sur le cercle unité, et développons $|\Phi(e^{i\theta})|^2$ et $|\Theta(e^{i\theta})|^2$ au voisinage de θ_0 :

$$\begin{aligned} \left| \Phi(e^{i\theta}) \right|^2 &\sim a |\theta - \theta_0|^{2m}; \\ \left| \Theta(e^{i\theta}) \right|^2 &\sim b |\theta - \theta_0|^{2n}, \end{aligned}$$

où m et n désignent les multiplicités respectives de $e^{i\theta_0}$ en tant que racine dans Φ et Θ , et où a, b sont des constantes strictement positives. En intégrant, on obtient lorsque $\varepsilon \rightarrow 0$,

$$\begin{aligned} \int_{\theta_0-\varepsilon}^{\theta_0+\varepsilon} \left| \Phi(e^{i\theta}) \right|^2 \nu_X(d\theta) &\sim a \int_{\theta_0-\varepsilon}^{\theta_0+\varepsilon} |\theta - \theta_0|^{2m} \nu_X(d\theta) \leq a \varepsilon^{2m} \nu_X(-\pi, \pi) = a \gamma_X(0) \varepsilon^{2m} \\ \frac{1}{2\pi} \int_{\theta_0-\varepsilon}^{\theta_0+\varepsilon} \left| \Theta(e^{i\theta}) \right|^2 d\theta &\sim \frac{b}{2\pi} \int_{\theta_0-\varepsilon}^{\theta_0+\varepsilon} |\theta - \theta_0|^{2n} d\theta = \frac{b}{\pi(2n+1)} \varepsilon^{2n+1}. \end{aligned}$$

Mais d'après (4.7), ces deux intégrales coïncident, donc on a forcément $2n+1 \geq 2m$, c'est à-dire $n-m \geq -\frac{1}{2}$, c'est-à-dire $n \geq m$ (car n et m sont entiers). Nous venons donc de démontrer que, s'il existe une solution stationnaire X , alors les racines de module 1 de Φ sont nécessairement compensées par Θ . C'est exactement la contraposée du résultat cherché. \square

Exercice 4.3 (Solutions stationnaires non-linéaires). *On suppose ici que le polynôme Φ admet des racines de module 1, mais que celles-ci sont toutes compensées par Θ . On note X l'unique solution linéaire de l'équation ARMA(Φ, Θ). On se donne également un paramètre θ et deux variables aléatoires U, V standard, décorréelées entre elles et décorréelées du bruit blanc Z . On introduit le processus H défini par :*

$$\forall t \in \mathbb{Z}, \quad H_t := U \cos(\theta t) + V \sin(\theta t). \quad (4.8)$$

4.4. Retour sur les processus ARMA

1. Rappeler la moyenne et la mesure spectrale du processus H .
2. Montrer que $X + \sigma H$ est un processus stationnaire non-linéaire pour tout $\sigma \in \mathbb{R} \setminus \{0\}$.
3. Soit $\alpha \in \ell^1(\mathbb{Z})$ un filtre. Trouver une condition nécessaire et suffisante sur $P_\alpha(e^{i\theta})$ pour que le processus filtré $F_\alpha(H)$ soit identiquement nul.
4. Construire une infinité de solutions stationnaires non-linéaires à l'équation $\text{ARMA}(\Phi, \Theta)$.

Une observation importante est que la densité spectrale (donc la fonction d'auto-covariance) d'un processus ARMA ne dépend que du module de la fraction rationnelle associée.

Théorème 4.8 (Représentation isospectrale causale et d'inverse causal). *Soit X la solution d'une équation $\text{ARMA}(p, q)$ dont les polynômes associés Φ et Θ n'admettent pas de racine de module 1 :*

$$\Phi(z) = \prod_{j=1}^p \left(1 - \frac{z}{a_j}\right) \quad \text{avec} \quad |a_1| \leq \dots \leq |a_r| < 1 < |a_{r+1}| \leq \dots \leq |a_p|;$$

$$\Theta(z) = \prod_{j=1}^q \left(1 - \frac{z}{b_j}\right) \quad \text{avec} \quad |b_1| \leq \dots \leq |b_r| < 1 < |b_{r+1}| \leq \dots \leq |b_q|.$$

Considérons une nouvelle équation $\text{ARMA}(p, q)$ en multipliant le bruit blanc $(Z_t)_{t \in \mathbb{Z}}$ par

$$\sigma := \frac{\prod_{j=1}^r |a_j|}{\prod_{j=1}^s |b_j|},$$

et en remplaçant chaque racine de module < 1 de Φ, Θ par l'inverse de son conjugué :

$$\tilde{\Phi}(z) = \prod_{j=1}^r (1 - \bar{a}_j z) \prod_{j=r+1}^p \left(1 - \frac{z}{a_j}\right);$$

$$\tilde{\Theta}(z) = \prod_{j=1}^s (1 - \bar{b}_j z) \prod_{j=s+1}^q \left(1 - \frac{z}{b_j}\right).$$

Alors la nouvelle équation $\text{ARMA}(p, q)$ admet une unique solution stationnaire \tilde{X} , qui est causale, inversible, d'inverse causal, et qui a la même auto-covariance que X .

Démonstration. D'abord, Φ, Θ sont à coefficients réels, ce qui signifie que leurs racines irréeselles vont par paires conjuguées a, \bar{a} . Comme $\overline{1/a} = 1/\bar{a}$, on en déduit que les racines irréeselles de $\tilde{\Phi}, \tilde{\Theta}$ vont elles aussi par paires conjuguées, et donc $\tilde{\Phi}, \tilde{\Theta}$ sont encore à coefficients réels. Comme de plus leur terme constant vaut 1, ils définissent bien une équation ARMA. Ensuite, l'opération $T: a \mapsto 1/\bar{a}$ inverse le module, donc $\tilde{\Phi}, \tilde{\Theta}$ n'ont que des racines de module > 1 : la nouvelle équation ARMA admet donc une unique solution stationnaire \tilde{X} , causale, inversible, et d'inverse causal. De plus, d'après le théorème de filtrage spectral, on a pour $u \in [-\pi, \pi]$,

$$f_X(u) = \frac{|\Theta(e^{iu})|}{|\Phi(e^{iu})|} f_Z(u) \quad \text{et} \quad f_{\tilde{X}}(u) = \frac{|\tilde{\Theta}(e^{iu})|}{|\tilde{\Phi}(e^{iu})|} \sigma^2 f_Z(u).$$

Nous allons maintenant montrer que ces deux quantités sont égales, ce qui impliquera que $\gamma_X = \gamma_{\tilde{X}}$ et terminera la preuve. Pour cela, observons que pour $a, z \in \mathbb{C}$ avec $|z| = 1$, on a

$$\left|1 - \frac{z}{T(a)}\right| = |1 - \bar{a}z| = |a\bar{z} - 1| = |a - z| = |a| \left|1 - \frac{z}{a}\right|$$

Ainsi, remplacer une racine a par $T(a)$ a pour effet de multiplier le module du polynôme par $|a|$. En conséquence, on a $\frac{|\tilde{\Theta}(z)|}{|\tilde{\Phi}(z)|} = \sigma \frac{|\Theta(z)|}{|\Phi(z)|}$ pour tout z de module 1, ce qui conclut la preuve. \square

Chapitre 5

Prédiction linéaire

5.1 Cadre théorique

Dans tout ce chapitre, on considère un processus stationnaire centré $X = (X_t)_{t \in \mathbb{Z}}$. On fixe un temps $t \in \mathbb{Z}$, et on cherche à prédire du mieux que l'on peut la variable aléatoire X_t , en utilisant uniquement les $p \in \mathbb{N}$ observations précédentes $X_{t-1}, X_{t-2}, \dots, X_{t-p}$. Autrement dit, on cherche une approximation de X_t sous la forme

$$\widehat{X}_t := f(X_{t-1}, \dots, X_{t-p}),$$

où $f: \mathbb{R}^p \rightarrow \mathbb{R}$ est une fonction déterministe aussi simple que possible. Le cadre le plus naturel (et le plus important en pratique) est celui où la fonction f est **linéaire**, ce qui explique le titre du chapitre. Autrement dit, on cherche des coefficients réels $\varphi_{p,1}, \dots, \varphi_{p,p}$ tels que la variable

$$\widehat{X}_t := \sum_{k=1}^p \varphi_{p,k} X_{t-k},$$

soit **aussi proche que possible** de la variable cible X_t . Nous allons donner un sens précis à cette expression vague, puis expliquer comment déterminer en pratique les coefficients $\varphi_{p,1}, \dots, \varphi_{p,p}$. Nous commençons par rappeler le résultat fondamental suivant (faire un dessin!).

Théorème 5.1 (Projection orthogonale). *Soit $(E, \langle \cdot, \cdot \rangle)$ un espace de Hilbert, et $H \subseteq E$ un sous-espace vectoriel fermé. Alors, tout point $x \in E$ se décompose de manière unique sous la forme*

$$x = x_H + x_{H^\perp}$$

avec $x_H \in H$ et $x_{H^\perp} \in H^\perp$. De plus, x_H est le point de H le plus proche de x pour la norme associée au produit scalaire $\langle \cdot, \cdot \rangle$. En effet, pour tout point $z \in H$, on a

$$\|x - z\|^2 = \|x_H - z\|^2 + \|x_{H^\perp}\|^2 \geq \|x_{H^\perp}\|^2,$$

avec égalité si et seulement si $z = x_H$. On appelle x_H le **projeté orthogonal** de x sur S , noté **proj**(x, H).

Remarque 5.1 (Linéarité et continuité). *L'application $x \mapsto \text{proj}(x, H)$ est clairement linéaire. Elle est en outre continue, puisque $\|\text{proj}(x, H)\|^2 = \|x\|^2 - \|x_{H^\perp}\|^2 \leq \|x\|^2$.*

Pour notre problème de prédiction linéaire, on se place bien-sûr sur l'espace de Hilbert $E = L^2(\Omega, \mathcal{F}, \mathbb{P})$ muni du produit scalaire $\langle X, Y \rangle = \mathbb{E}[XY]$. Pour $t \in \mathbb{Z}$ et $p \in \mathbb{N}$, on note

$$H_{t-1,p} := \text{Vect}(X_{t-1}, \dots, X_{t-p})$$

5.1. Cadre théorique

l'espace vectoriel engendré par les p observations passées. On cherche, dans cette espace, la variable aléatoire \widehat{X}_t qui soit la plus proche possible de X_t . Au vu du théorème 5.1, on a

$$\widehat{X}_t := \text{proj}(X_t, H_{t-1,p}). \quad (5.1)$$

Cette variable aléatoire est appelée **prédicteur progressif d'ordre p** de X_t . On peut mesurer la qualité de cette prédiction en calculant l'**erreur de prédiction**, définie par

$$\begin{aligned} \sigma_p^2 &:= \|X_t - \text{proj}(X_t, H_{t-1,p})\|_{L^2}^2 \\ &= \min_{\varphi \in \mathbb{R}^p} \left\| X_t - \sum_{k=1}^p \varphi_k X_{t-k} \right\|_{L^2}^2 \\ &= \min_{\varphi \in \mathbb{R}^p} \sum_{j=0}^p \sum_{k=0}^p \varphi_j \varphi_k \gamma_X(j-k), \end{aligned}$$

avec la convention $\varphi_0 = -1$. La dernière ligne montre que σ_p ne dépend pas du temps t , et il en est de même de l'ensemble des vecteurs $\varphi \in \mathbb{R}^p$ réalisant le minimum dans la formule ci-dessus.

Remarque 5.2 (Norme d'une somme). *La dernière ligne s'obtient en développant la norme : si X_1, \dots, X_n sont des variables aléatoires centrées et $\theta_1, \dots, \theta_n$ des réels, alors on a par bilinéarité*

$$\begin{aligned} \left\| \sum_{j=1}^n \theta_j X_j \right\|^2 &= \text{Var} \left(\sum_{j=1}^n \theta_j X_j \right) \\ &= \text{Cov} \left(\sum_{j=1}^n \theta_j X_j, \sum_{k=1}^n \theta_k X_k \right) \\ &= \sum_{j=1}^n \sum_{k=1}^n \theta_j \theta_k \text{Cov}(X_j, X_k). \end{aligned}$$

Remarque 5.3 (Monotonie). *On a $H_{t-1,p} \subseteq H_{t-1,p+1}$ pour tout $p \in \mathbb{N}$. On en déduit aussitôt que $(\sigma_p)_{p \geq 1}$ est décroissante : plus nous avons d'observations à notre disposition, meilleure est notre prédiction. Attention, cependant, la limite $\sigma_\infty = \lim_{p \rightarrow \infty} \sigma_p$ n'est pas nécessairement nulle : en général, on ne peut pas prédire parfaitement X_t , même en connaissant tout le passé X_{t-1}, X_{t-2}, \dots*

Exemple 5.1 (AR(m) causal). *Soit $Z \sim \text{BB}(0,1)$ et soit X l'unique solution stationnaire de*

$$X_t = Z_t + \varphi_1 X_{t-1} + \dots + \varphi_m X_{t-m},$$

où le polynôme $\Phi(z) = 1 - \sum_{k=1}^m \varphi_k z^k$ est supposé sans racines de module ≤ 1 . Alors X est causal, donc $\text{Cov}(Z_t, X_{t-p}) = 0$ pour tout $p \geq 1$. Autrement dit, pour tout $p \geq 1$,

$$X_t - \sum_{k=1}^m \varphi_k X_{t-k} \in H_{t-1,p}^\perp$$

Comme par ailleurs $\sum_{k=1}^m \varphi_k X_{t-k} \in H_{t-1,p}$ pour tout $p \geq m$, on conclut que pour tout $p \geq m$,

$$\begin{aligned} \text{proj}(X_t, H_{t-1,p}) &= \sum_{k=1}^m \varphi_k X_{t-k} \\ \sigma_p^2 &= 1. \end{aligned}$$

5.2 Équation de Yule-Walker

Par définition, la prédiction $\widehat{X}_t := \text{proj}(X_t, H_{t-1,p})$ est dans l'espace $H_{t-1,p}$. Cela signifie que

$$\text{proj}(X_t, H_{t-1,p}) = \sum_{k=1}^p \varphi_{p,k} X_{t-k}, \quad (5.2)$$

pour un certain vecteur $\varphi_p = (\varphi_{p,1}, \dots, \varphi_{p,p})^\top \in \mathbb{R}^p$, que nous cherchons ici à déterminer. Attention, comme nous allons le voir, il peut y avoir plusieurs choix possibles pour φ_p . En fait, ces choix sont exactement les solutions de l'équation de Yule-Walker, que nous allons maintenant expliciter. Pour cela, on introduit un vecteur γ_p et une matrice Γ_p :

$$\gamma_p := (\gamma_X(1), \dots, \gamma_X(p))^\top, \quad (5.3)$$

$$\Gamma_p := (\gamma_X(k-j))_{1 \leq k, j \leq p}. \quad (5.4)$$

Théorème 5.2 (Équation de Yule-Walker). *Un vecteur $\varphi_p = (\varphi_{p,1}, \dots, \varphi_{p,p})^\top \in \mathbb{R}^p$ satisfait (5.2) si et seulement si il est solution de l'équation de Yule-Walker :*

$$\Gamma_p \varphi_p = \gamma_p.$$

De plus on a l'expression suivante pour l'erreur de prédiction :

$$\sigma_p^2 = \gamma_X(0) - \varphi_p^\top \Gamma_p \varphi_p.$$

Démonstration. Soit $\varphi_p = (\varphi_{p,1}, \dots, \varphi_{p,p})^\top \in \mathbb{R}^p$, et soit $\widehat{X}_t := \sum_{k=1}^p \varphi_{p,k} X_{t-k}$. Alors il est clair que $\widehat{X}_t \in H_{t-1,p}$, et par définition du projecteur, on a les équivalences suivantes :

$$\begin{aligned} \text{proj}(X_t, H_{t-1,p}) = \widehat{X}_t &\iff \widehat{X}_t - X_t \in H_{t-1,p}^\perp \\ &\iff \forall h \in \{1, \dots, p\}, \text{Cov}(\widehat{X}_t - X_t, X_{t-h}) = 0 \\ &\iff \forall h \in \{1, \dots, p\}, \text{Cov}(\widehat{X}_t, X_{t-h}) = \text{Cov}(X_t, X_{t-h}) \\ &\iff \forall h \in \{1, \dots, p\}, \sum_{k=1}^p \varphi_{p,k} \text{Cov}(X_{t-k}, X_{t-h}) = \text{Cov}(X_t, X_{t-h}) \\ &\iff \forall h \in \{1, \dots, p\}, \sum_{k=1}^p \varphi_{p,k} \gamma_X(k-h) = \gamma_X(h) \\ &\iff \Gamma_p \varphi_p = \gamma_p. \end{aligned}$$

D'autre part, lorsque $\text{proj}(X_t, H_{t-1,p}) = \widehat{X}_t$, on a par orthogonalité,

$$\|X_t\|_{L^2}^2 = \|X_t - \widehat{X}_t\|_{L^2}^2 + \|\widehat{X}_t\|_{L^2}^2.$$

Comme $\sigma_p^2 = \|X_t - \widehat{X}_t\|_{L^2}^2$, on en déduit que

$$\begin{aligned} \sigma_p^2 &= \|X_t\|_{L^2}^2 - \|\widehat{X}_t\|_{L^2}^2 \\ &= \text{Var}(X_t) - \text{Var}\left(\sum_{k=1}^p \varphi_{p,k} X_{t-k}\right) \\ &= \gamma_X(0) - \varphi_p^\top \Gamma_p \varphi_p. \end{aligned}$$

□

5.3. Sur l'unicité dans l'équation de Yule-Walker

Exemple 5.2 (Bruit blanc). *Écrire l'équation de Yule-Walker lorsque $X \sim \text{BB}(0, 1)$, puis la résoudre.*

Exercice 5.1 (Un processus non-linéaire). *Soit X un processus stationnaire centré, d'auto-covariance*

$$\gamma_X(h) := \begin{cases} 2 & \text{si } h = 0 \\ 1 & \text{sinon.} \end{cases}$$

1. Si on observe les valeurs $X_0 = 1, X_1 = -2, X_2 = 5$, quelle est la meilleure prédiction pour X_3 ?
2. Déterminer le prédicteur progressif d'ordre p de X_t pour tout $p \geq 1$.
3. Calculer l'erreur de prédiction associée, et sa limite lorsque $p \rightarrow \infty$.
4. Construire un processus stationnaire X ayant la fonction d'auto-covariance ci-dessus.
5. Les matrices $(\Gamma_p)_{p \geq 1}$ sont-elles inversibles?

5.3 Sur l'unicité dans l'équation de Yule-Walker

L'existence du projeté orthogonal $\text{proj}(X_t, H_{t-1,p})$ garantit que l'équation de Yule-Walker $\Gamma_p \varphi_p = \gamma_p$ admet au moins une solution. En revanche, il n'y a pas forcément unicité, ce qui constitue une difficulté majeure pour la résolution pratique de l'équation de Yule-Walker. Le lemme suivant nous dit précisément quand cela se produit.

Lemme 5.1 (Unicité et invertibilité). *Les conditions suivantes sont équivalentes :*

- (i) *l'équation de Yule-Walker $\Gamma_p \varphi_p = \gamma_p$ admet une unique solution ;*
- (ii) *la matrice Γ_p est inversible ;*
- (iii) *la matrice Γ_p est définie positive ;*
- (iv) *la famille $(X_{t-1}, \dots, X_{t-p})$ est linéairement indépendante.*
- (v) $\sigma_{p-1}^2 > 0$.

Démonstration de (i) \implies (ii). Supposons que l'équation de Yule-Walker admet une unique solution φ_p , et considérons un vecteur $\phi \in \ker(\Gamma_p)$. Alors $\varphi_p + \phi$ est aussi solution, puisque

$$\Gamma_p(\varphi_p + \phi) = \Gamma_p \varphi_p + \Gamma_p \phi = \Gamma_p \varphi_p + 0 = \gamma_p.$$

Par unicité, on a forcément $\phi = 0$, ce qui prouve que $\ker(\Gamma_p) = \{0\}$. Ainsi, Γ_p est inversible. \square

Démonstration de (ii) \implies (iii). Par définition, Γ_p n'est rien d'autre que la matrice de covariance du vecteur $(X_{t-1}, \dots, X_{t-p})$. Elle est donc toujours symétrique et ses valeurs propres sont positives. Si elle est inversible, alors 0 n'est pas valeur propre et donc toutes les valeurs propres sont strictement positives, ce qui montre que Γ_p est définie positive. \square

Démonstration de (iii) \implies (iv). Comme X est centré, la remarque 5.2 nous permet d'écrire

$$\|\varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p}\|^2 = \varphi^\top \Gamma_p \varphi, \quad (5.5)$$

pour tout vecteur $\varphi = (\varphi_1, \dots, \varphi_p) \in \mathbb{R}^p$. Si la matrice Γ_p est définie positive, on en déduit que $\varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} \neq 0$ pour tout $\varphi \in \mathbb{R}^p \setminus \{0\}$, ce qui signifie précisément que la famille $(X_{t-1}, \dots, X_{t-p})$ est linéairement indépendante. \square

Démonstration de (iv) \implies (v). Si $\sigma_{p-1}^2 = 0$, alors $X_{t-1} = \text{proj}(X_{t-1}, H_{t-2,p-1})$ donc $X_{t-1} \in H_{t-2,p-1}$ donc la famille $(X_{t-1}, \dots, X_{t-p})$ ne peut pas être linéairement indépendante. \square

Démonstration de (v) \implies (iv). Si la famille $(X_{t-1}, \dots, X_{t-p})$ est linéairement liée, alors il existe $k \in \{1, \dots, p\}$ tel que $X_{t-k} \in \text{Vect}(X_{t-1}, \dots, X_{t-k-1})$. Mais alors $\sigma_{k-1}^2 = 0$, donc $\sigma_{p-1}^2 = 0$. \square

Démonstration de (iv) \implies (i). On a déjà vu que l'équation de Yule-Walker admet toujours au moins une solution $\varphi \in \mathbb{R}^p$, donc il reste à établir l'unicité. Or si $\phi \neq \varphi$ était une autre solution de l'équation de Yule-Walker, alors d'après le théorème 5.2, on aurait

$$(\varphi_1 - \phi_1)X_{t-1} + \dots + (\varphi_p - \phi_p)X_{t-p} = \text{proj}(X_t, H_{t-1,p}) - \text{proj}(X_t, H_{t-1,p}) = 0, \quad (5.6)$$

ce qui contredirait l'indépendance linéaire de la famille $(X_{t-1}, \dots, X_{t-p})$. \square

Remarque 5.4 (Formule variationnelle et invertibilité). *On rappelle que la plus petite valeur propre d'une matrice symétrique Γ est donnée par la formule*

$$\lambda_{\min}(\Gamma) = \min_{\theta \in \mathbb{R}^n \setminus \{0\}} \frac{\theta^\top \Gamma \theta}{\|\theta\|_2^2},$$

comme on le voit en décomposant θ sur une base orthonormée de vecteurs propres. En particulier, une matrice symétrique positive est inversible ($\lambda_{\min}(\Gamma) \neq 0$) si et seulement si elle est définie positive.

Voici une condition suffisante très simple pour garantir l'invertibilité de Γ_p .

Théorème 5.3 (Inversibilité). *Si $\gamma_X(h) \rightarrow 0$ lorsque $h \rightarrow \infty$, et si en outre on a $\gamma_X(0) > 0$ (i.e. le processus X n'est pas identiquement nul), alors les matrices $(\Gamma_p)_{p \geq 1}$ sont toutes inversibles.*

Démonstration. On raisonne par récurrence sur p . Si $p = 1$, on a $\Gamma_1 = \gamma_X(0) > 0$, qui est bien inversible. Supposons maintenant qu'il existe $p \geq 1$ tel que Γ_p soit inversible, mais pas Γ_{p+1} . D'après le Lemme ci-dessus, la famille $(X_{t-1}, X_{t-2}, \dots, X_{t-p})$ est linéairement indépendante, mais pas $(X_t, X_{t-1}, \dots, X_{t-p})$. Ainsi, X_t est une combinaison linéaire des p v.a.r. précédentes X_{t-1}, \dots, X_{t-p} . Comme ceci est vrai pour tout t , on peut itérer pour obtenir que pour tout $t \geq p$, X_t est combinaison linéaire des (X_1, \dots, X_p) . Autrement dit, on peut écrire

$$X_t = \sum_{k=1}^p \theta_{t,k} X_k,$$

pour un certain vecteur $\theta_t = (\theta_{t,1}, \dots, \theta_{t,p}) \in \mathbb{R}^p$. On en déduit les deux formules suivantes :

$$\begin{aligned} \text{Var}(X_t) &= \theta_t^\top \Gamma_p \theta_t \geq \lambda_{\min}(\Gamma_p) \|\theta_t\|^2 \\ \text{Var}(X_t) &= \text{Cov} \left(X_t, \sum_{k=1}^p \theta_{t,k} X_k \right) = \sum_{k=1}^p \theta_{t,k} \gamma_X(t-k) \leq \|\theta_t\| \sqrt{\sum_{k=1}^p |\gamma_X(t-k)|^2}. \end{aligned}$$

Comme Γ_p est inversible et que $\text{Var}(X_t)$ ne dépend pas de t , la première ligne montre que $\|\theta_t\|$ est borné indépendamment de t . Mais alors la seconde ligne montre que $\text{Var}(X_t) \rightarrow 0$ lorsque $t \rightarrow \infty$, ce qui est absurde puisque $\text{Var}(X_t) = \gamma_X(0) > 0$. \square

Remarque 5.5 (Contre-exemple). *Attention, la condition $\gamma_X(h) \rightarrow 0$ est suffisante, mais pas nécessaire pour que les matrices $(\Gamma_p)_{p \geq 1}$ soient toutes inversibles, comme le montre l'exercice 5.1.*

Remarque 5.6 (Ubiquité). *La condition $\gamma_X(h) \rightarrow 0$ est vérifiée en particulier par tout processus linéaire, et plus généralement par tout processus admettant une densité spectrale (c.f. exercices).*

Exercice 5.2 (Processus harmonique). *Soit X un processus harmonique de fréquence $\theta \neq 0$.*

1. Montrer que Γ_1 et Γ_2 sont inversibles, mais pas Γ_3 .
2. Trouver une relation linéaire non-triviale explicite entre X_t, X_{t-1}, X_{t-2} .
3. En déduire le prédicteur progressif d'ordre p de X_t pour tout $p \geq 3$.

5.4 Algorithme de Levinson-Durbin

Dans le cas où les matrices $(\Gamma_p)_{p \geq 1}$ sont toutes inversibles, l'équation de Yule-Walker d'ordre p admet une unique solution pour tout $p \geq 1$, donnée par $\varphi_p = \Gamma_p^{-1} \gamma_p$. Cependant, l'inversion de la matrice Γ_p peut-être coûteuse d'un point de vue calculatoire. L'**algorithme de Levinson-Durbin** permet de calculer récursivement $(\varphi_p)_{p \geq 1}$ et $(\sigma_p^2)_{p \geq 1}$ sans inversion de matrice.

Théorème 5.4 (Algorithme de Levinson-Durbin). *On suppose que les matrices $(\Gamma_p)_{p \geq 1}$ sont toutes inversibles. Alors, les suites $(\varphi_p)_{p \geq 1}$ et $(\sigma_p^2)_{p \geq 1}$ se calculent récursivement à partir de γ_X en posant $\sigma_0 := \gamma_X(0)$ puis en appliquant successivement les formules suivantes pour $p \geq 0$,*

$$\begin{aligned} \varphi_{p+1,p+1} &= \frac{1}{\sigma_p^2} \left(\gamma_X(p+1) - \sum_{k=1}^p \varphi_{p,k} \gamma_X(p+1-k) \right) \\ \forall k \in \{1, \dots, p\}, \quad \varphi_{p+1,k} &= \varphi_{p,k} - \varphi_{p+1,p+1} \times \varphi_{p,p+1-k} \\ \sigma_{p+1}^2 &= \sigma_p^2 \left(1 - \varphi_{p+1,p+1}^2 \right). \end{aligned}$$

La preuve du théorème repose sur l'observation simple mais profonde suivante :

Remarque 5.7 (Réversibilité temporelle). *Le processus \tilde{X} défini par $\tilde{X}_t = X_{-t}$ (retournement du temps) a la même fonction d'auto-covariance que le processus X :*

$$\forall h \in \mathbb{Z}, \quad \text{Cov}(\tilde{X}_t, \tilde{X}_{t+h}) = \text{Cov}(X_{-t}, X_{-t-h}) = \gamma_X(-h) = \gamma_X(h).$$

En particulier, les données Γ_p et γ_p sont les mêmes pour X et \tilde{X} , donc les équations de Yule-Walker sont invariantes par retournement du temps ! Ainsi, le vecteur solution $\varphi_p = \Gamma_p^{-1} \gamma_p$ permet aussi bien de prédire le futur que le passé, et l'erreur de prédiction $\sigma_p^2 = \varphi_p^\top \gamma_p$ est la même. Comme $H_{t-1,p}$ est le futur d'ordre p de la coordonnée $X_{t-(p+1)}$, on en déduit que

$$\text{proj}(X_{t-(p+1)}, H_{t-1,p}) = \sum_{k=1}^p \varphi_{p,k} X_{t-(p+1)+k} = \sum_{k=1}^p \varphi_{p,p+1-k} X_{t-k} \quad (5.7)$$

$$\|X_{t-(p+1)} - \text{proj}(X_{t-(p+1)}, H_{t-1,p})\|^2 = \sigma_p^2. \quad (5.8)$$

Démonstration du Théorème. Le point de départ est la décomposition orthogonale suivante :

$$H_{t-1,p+1} = H_{t-1,p} \oplus^\perp \text{Vect}(Z), \quad (5.9)$$

où l'on a posé

$$Z = X_{t-(p+1)} - \text{proj}(X_{t-(p+1)}, H_{t-1,p}) \quad (5.10)$$

$$= X_{t-(p+1)} - \sum_{k=1}^p \varphi_{p,p+1-k} X_{t-k}. \quad (5.11)$$

Notons que la seconde ligne utilise (5.8), et que l'on a par ailleurs $\|Z\|^2 = \sigma_p^2$ par (5.8). De la décomposition orthogonale (5.9), on déduit immédiatement que

$$\text{proj}(X_t, H_{t-1,p+1}) = \text{proj}(X_t, H_{t-1,p}) \oplus^\perp \kappa Z, \quad (5.12)$$

où κ est donné par la formule suivante (projection orthogonale sur un espace de dimension 1) :

$$\kappa = \frac{\langle X_t, Z \rangle}{\|Z\|^2} = \frac{\text{Cov}(X_t, Z)}{\sigma_p^2} = \frac{1}{\sigma_p^2} \left(\gamma_X(p+1) - \sum_{k=1}^p \varphi_{p,p+1-k} \gamma_X(k) \right). \quad (5.13)$$

Par unicité dans les équations de Yule-Walker, on peut identifier le coefficient devant X_{t-k} doit être le même de part et d'autre de (5.12), pour tout $k \in \{1, \dots, p+1\}$. Pour $k = p+1$ cela donne

$$\varphi_{p+1,p+1} = \kappa, \quad (5.14)$$

tandis que pour $k \in \{1, \dots, p\}$ on obtient

$$\varphi_{p+1,k} = \varphi_{p,k} - \kappa\varphi_{p,p+1-k}. \quad (5.15)$$

Enfin, pour l'erreur de prédiction, on a par orthogonalité

$$\begin{aligned} \sigma_{p+1}^2 &= \|X_t\|^2 - \|\text{proj}(X_t, H_{t-(p+1)})\|^2 \\ &= \|X_t\|^2 - \|\text{proj}(X_t, H_{t-p})\|^2 - \kappa^2 \|Z\|^2 \\ &= \sigma_p^2 - \kappa^2 \sigma_p^2 \end{aligned}$$

Cela conclut la preuve. □

Chapitre 6

Estimation statistique

Soit $X = (X_t)_{t \geq 0}$ un processus stationnaire. Dans le chapitre précédent, nous avons vu comment prédire au mieux la valeur de X_t par une combinaison linéaire des observations passées X_{t-1}, \dots, X_{t-p} : il suffit pour cela de résoudre l'équation de Yule-Walker associée (ce que l'on peut faire efficacement, par exemple par l'algorithme récursif de Levinson-Durbin). Mais cette équation fait intervenir l'auto-covariance γ_X , que nous avons jusque-là supposée connue. Dans ce chapitre, nous allons voir comment estimer μ_X et γ_X à partir des données.

6.1 Rappels sur les estimateurs

Commençons par un rappel élémentaire de terminologie statistique, sur un exemple trivial. Soit $(X_n)_{n \geq 1}$ une suite de variable aléatoires dans L^2 , indépendantes et identiquement distribuées, de moyenne inconnue μ . À partir de l'observation des n premières valeurs X_1, \dots, X_n , on aimerait **estimer** μ . On définit pour cela la **moyenne empirique** :

$$\bar{X}_n := \frac{1}{n} \sum_{k=1}^n X_k.$$

Cette variable aléatoire, construite à partir des données observées, fournit un excellent estimateur de l'espérance μ . D'abord, cet estimateur est **sans biais**, puisque l'on a par linéarité

$$\mathbb{E} [\bar{X}_n] = \mu. \quad (6.1)$$

Ensuite et surtout, il est **fortement convergent**, puisqu'il obéit à la loi forte des grands nombres :

$$\bar{X}_n \xrightarrow[n \rightarrow \infty]{p.s.} \mu. \quad (6.2)$$

Enfin, on peut quantifier précisément l'**erreur quadratique moyenne** de cet estimateur :

$$\mathbb{E} [(\bar{X}_n - \mu)^2] = \frac{\sigma^2}{n}, \quad (6.3)$$

où σ^2 désigne la variance des variables $(X_n)_{n \geq 1}$. Cette quantité est précieuse, puisqu'elle permet immédiatement d'obtenir des **intervalles de confiance** via l'inégalité de Chebychev :

$$\mathbb{P} (|X_n - \mu| \geq \varepsilon) \leq \frac{\mathbb{E} [(\bar{X}_n - \mu)^2]}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}. \quad (6.4)$$

Ainsi, à précision $\varepsilon > 0$ fixée, la longueur de l'intervalle de confiance tend vers 0 à une vitesse de l'ordre de $\frac{1}{\sqrt{n}}$ lorsque $n \rightarrow \infty$: c'est la **vitesse de convergence** de l'estimateur.

Notons qu'en faisant tendre $n \rightarrow \infty$ dans (6.3), on obtient que la convergence (6.2) a aussi lieu dans L^2 . En revanche, si l'on fait tendre $n \rightarrow \infty$ dans (6.4), on obtient seulement la convergence en probabilité de $\overline{X}_n \rightarrow \mu$ (on parle dans ce cas d'estimateur **faiblement convergent**). Voyons maintenant comment cela se généralise au cas de processus stationnaires plus réalistes.

6.2 Estimation de la moyenne

Soit X un processus stationnaire. Comment estimer sa moyenne μ_X à partir de l'observation des n premières données X_1, \dots, X_n ? L'idée naturelle est bien-sûr de considérer la moyenne empirique, comme dans le cas i.i.d. Cependant, l'exemple dégénéré suivant montre que des hypothèses nous seront nécessaires pour obtenir un estimateur convergent.

Exemple 6.1. Soit $U \in L^2$ de variance non nulle. On considère le processus stationnaire X défini par :

$$\forall t \in \mathbb{Z}, \quad X_t := U.$$

Alors même l'observation d'une infinité de données $(X_1(\omega), X_2(\omega), \dots)$ n'apporte pas plus d'information que la réalisation $U(\omega)$, et il nous sera bien difficile d'en tirer quoi que ce soit d'intéressant sur la loi de U ! En particulier, la moyenne empirique n'est pas ici un estimateur convergent de μ_X , même faiblement.

Théorème 6.1 (Moyenne empirique). Soit X un processus stationnaire. Alors la **moyenne empirique**

$$\overline{X}_n := \frac{1}{n} \sum_{k=1}^n X_k, \tag{6.5}$$

est un estimateur **sans biais** de l'espérance μ_X du processus, et l'**erreur quadratique moyenne** associée est

$$\left[(\overline{X}_n - \mu_X)^2 \right] = \frac{1}{n} \sum_{h=-n}^n \left(1 - \frac{|h|}{n} \right) \gamma_X(h). \tag{6.6}$$

En particulier, sous l'hypothèse supplémentaire que $\gamma_X(h) \rightarrow 0$ lorsque $h \rightarrow \infty$, la moyenne empirique est un **estimateur convergent** dans L^2 :

$$\overline{X}_n \xrightarrow[n \rightarrow \infty]{L^2} \mu_X. \tag{6.7}$$

Enfin, sous l'hypothèse plus forte que γ_X est sommable, on a même

$$n \left[(\overline{X}_n - \mu_X)^2 \right] \xrightarrow[n \rightarrow \infty]{} 2\pi f_X(0), \tag{6.8}$$

où f_X désigne la densité spectrale du processus X . En particulier, la **vitesse de convergence** est $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$.

Démonstration. Pour le biais, on a par linéarité et stationnarité :

$$\mathbb{E}[\overline{X}_n] = \frac{1}{n} \sum_{k=1}^n \mathbb{E}[X_k] = \mu_X.$$

Pour l'erreur quadratique moyenne, on a par bilinéarité de la covariance

$$\begin{aligned}
 \left[(\bar{X}_n - \mu_X)^2 \right] &= \text{Var}(\bar{X}_n - \mu_X) \\
 &= \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n \gamma_X(j-k) \\
 &= \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n \sum_{h=-n}^n \gamma_X(h) \mathbf{1}_{j-k=h} \\
 &= \frac{1}{n^2} \sum_{h=-n}^n \gamma_X(h) \sum_{j=1}^n \sum_{k=1}^n \mathbf{1}_{j-k=h} \\
 &= \frac{1}{n^2} \sum_{h=-n}^n \gamma_X(h) (n - |h|),
 \end{aligned}$$

où le calcul de la double somme à la dernière ligne a déjà été vu dans la preuve du théorème de Herglotz. En particulier, on déduit de cette formule (et de la parité de γ_X) que

$$\left[(\bar{X}_n - \mu_X)^2 \right] \leq \frac{\gamma_X(0)}{n} + \frac{2}{n} \sum_{h=1}^n |\gamma_X(h)|,$$

et le membre droit tend bien vers 0 lorsque $n \rightarrow \infty$ sous l'hypothèse $\gamma_X(h) \rightarrow 0$ (Cesarro). Finalement, notons que l'identité précédente se réécrit

$$n \left[(\bar{X}_n - \mu_X)^2 \right] = \sum_{h \in \mathbb{Z}} \gamma_X(h) \left(1 - \frac{|h|}{n} \right) \mathbf{1}_{|h| \leq n}.$$

Or, pour $h \in \mathbb{Z}$ fixé, le h -ème terme de cette somme tend vers $\gamma_X(h)$ lorsque $n \rightarrow \infty$, et il est borné par $\gamma_X(h)$ indépendamment de n . Sous l'hypothèse de sommabilité de γ_X , on peut donc invoquer le théorème de convergence dominée pour conclure que

$$n \left[(\bar{X}_n - \mu_X)^2 \right] \xrightarrow[n \rightarrow \infty]{} \sum_{h \in \mathbb{Z}} \gamma_X(h).$$

L'expression en termes de la densité spectrale découle de la Proposition 4.2. □

Notons que les hypothèses faites sur γ_X dans le théorème ci-dessus sont toujours vérifiées pour les processus linéaires. Elles ne sont pas optimales, comme le montre l'exercice suivante.

Exercice 6.1 (Processus harmonique). Soit X un processus harmonique de fréquence $\theta \in [-\pi, \pi]$.

1. La condition $\gamma_X(h) \rightarrow 0$ lorsque $h \rightarrow \infty$ est-elle vérifiée ici ?
2. Calculer l'erreur quadratique moyenne associée à l'estimateur \bar{X}_n , pour tout $n \geq 1$.
3. En déduire que si $\theta \neq 0$, \bar{X}_n est un estimateur faiblement convergent de μ_X et préciser sa vitesse.
4. Que se passe-t-il dans le cas $\theta = 0$?

6.3 Estimation de l'auto-covariance

Passons maintenant à l'estimation de l'auto-covariance d'un processus stationnaire X . On fixe $h \geq 0$, et on cherche à estimer $\gamma_X(h) = \mathbb{E}[X_0 X_h] - \mu_X^2$ à partir d'un grand nombre $n > h$

d'observations X_1, \dots, X_n . Nous savons déjà estimer μ_X , donc notre tâche se résume à estimer $\mathbb{E}[X_0 X_h]$. La moyenne empirique correspondant à cette quantité est

$$\gamma_h(X_1, \dots, X_n) := \frac{1}{n-h} \sum_{t=1}^{n-h} X_t X_{t+h}. \quad (6.9)$$

Le résultat qui suit montre que cet estimateur est effectivement un excellent choix, pour une classe très générale de processus linéaires, à savoir les processus de la forme

$$X = F_\alpha(Z), \quad (6.10)$$

où $\alpha \in \ell_1(\mathbb{Z})$ est un filtre quelconque, et où $Z = (Z_t)_{t \in \mathbb{Z}}$ est un **bruit blanc fort**, c'est-à-dire une suite de variables aléatoires i.i.d. que nous supposons centrées de variance 1, et dans L^4 :

$$\mathbb{E}[Z_0^4] < \infty. \quad (6.11)$$

Théorème 6.2 (Estimation de l'auto-covariance). *Soit X un processus linéaire construit à partir d'un bruit blanc fort standard dans L^4 . Alors, pour tout $h \geq 0$, la variable aléatoire $\gamma_h(X_1, \dots, X_n)$ définit un estimateur sans biais de $\gamma_X(h)$ convergent dans L^2 à vitesse $\mathcal{O}(\frac{1}{\sqrt{n}})$.*

Démonstration. Fixons $h \geq 0$ une fois pour toutes et définissons un processus $W = (W_t)_{t \in \mathbb{Z}}$ par :

$$\forall t \in \mathbb{Z}, \quad W_t := X_t X_{t+h}. \quad (6.12)$$

Notons que $\mathbb{E}[W_t] = \gamma_X(h)$ puisque X est centré. De plus, pour tout $n > h$,

$$\gamma_h(X_1, \dots, X_n) = \frac{1}{n-h} \sum_{t=1}^{n-h} W_t = \bar{W}_{n-h}. \quad (6.13)$$

Ainsi, toutes les conclusions annoncées découleront du Théorème 6.1, une fois que nous aurons vérifié que le processus W est stationnaire, et que γ_W est sommable! \square

Démonstration de la stationnarité de W . Le processus Z étant borné dans L^4 , le Théorème 2.1 assure que $X = F_\alpha(Z)$ aussi L^4 , et que l'on a pour tout $t \in \mathbb{Z}$,

$$\sum_{k=-n}^n \alpha_k Z_{t-k} \xrightarrow[n \rightarrow \infty]{p.s., L^4} X_t.$$

En particulier, on a $W_t \in L^2$ (produit de deux variables dans L^4), ce qui montre que W est du second-ordre. On a déjà calculé son espérance $\mathbb{E}[W_t] = \gamma_X(h)$, qui ne dépend pas de $t \in \mathbb{Z}$. Il reste donc à montrer que $\mathbb{E}[W_t W_{t+s}]$ ne dépend pas de $t \in \mathbb{Z}$. Pour cela, on note que $\mathbb{E}[W_t W_{t+s}] = \mathbb{E}[X_t X_{t+h} X_{t+s} X_{t+s+h}]$, et on utilise la convergence dans L^4 ci-dessus pour intervertir limite et intégrales et écrire

$$\mathbb{E}[X_t X_{t+h} X_{t+s} X_{t+s+h}] = \sum_{i,j,k,\ell \in \mathbb{Z}} \alpha_i \alpha_j \alpha_k \alpha_\ell \mathbb{E}[Z_{t-i} Z_{t+h-j} Z_{t+s-k} Z_{t+s+h-\ell}] \quad (6.14)$$

Comme Z est fortement stationnaire, la quantité $\mathbb{E}[Z_{t-i} Z_{t+h-j} Z_{t+s-k} Z_{t+s+h-\ell}]$ ne dépend pas de $t \in \mathbb{Z}$, ce qui conclut la preuve de la stationnarité de W . \square

Démonstration de la sommabilité de γ_W . Poursuivons le calcul de γ_W entamé ci-dessus. Comme Z est un bruit blanc fort standard, on a pour $(i, j, k, \ell) \in \mathbb{Z}^4$ in

$$\mathbb{E}[Z_i Z_j Z_k Z_\ell] = \begin{cases} \mathbb{E}[Z_0^4] & \text{si } i = j = k = \ell \\ 1 & \text{si } i = j \neq k = \ell \text{ ou } i = k \neq j = \ell \text{ ou } i = \ell \neq j = k \\ 0 & \text{sinon.} \end{cases}$$

En réinjectant dans (6.15), on en déduit que $\mathbb{E}[W_0 W_s] = \mathbb{E}[Z_0^4] a_s + b_s + b_s + d_s$, avec

$$\begin{aligned} a_s &= \sum_{i \in \mathbb{Z}} \alpha_i \alpha_{i+h} \alpha_{i+s} \alpha_{i+s+h} \\ b_s &= \sum_{i \neq j} \alpha_i \alpha_{i+h} \alpha_{j+s} \alpha_{j+s+h} \\ c_s &= \sum_{i \neq j} \alpha_i \alpha_{j+h} \alpha_{i+s} \alpha_{j+s+h} \\ d_s &= \sum_{i \neq j} \alpha_i \alpha_{j+h} \alpha_{j+s} \alpha_{i+s+h}. \end{aligned}$$

Mais en utilisant la formule $\gamma_X(k) = \sum_i \alpha_i \alpha_{i+k}$ valable pour tout $k \in \mathbb{Z}$, on peut ré-écrire

$$b_s = \gamma_X^2(h) - a_s, \quad c_s = \gamma_X^2(s) - a_s, \quad d_s = \gamma_X^2(s+h) - a_s,$$

et on conclut donc que l'auto-covariance de W est donnée par

$$\gamma_W(s) = \gamma_X^2(s) + \gamma_X^2(s+h) + (\mathbb{E}[Z_0^4] - 3) a_s.$$

Pour conclure, il nous suffit de vérifier que chacun de ces termes est sommable en $s \in \mathbb{Z}$. Or

$$\sum_{s \in \mathbb{Z}} |\gamma_X(s)| \leq \sum_{s \in \mathbb{Z}} \sum_{i \in \mathbb{Z}} |\alpha_i \alpha_{i+s}| = \|\alpha\|_1^2,$$

ce qui montre que γ_X est dans $\ell^1(\mathbb{Z})$, donc dans $\ell^2(\mathbb{Z})$. De même, on a

$$\sum_{s \in \mathbb{Z}} |a_s| \leq \sum_{s \in \mathbb{Z}} \sum_{i \in \mathbb{Z}} |\alpha_i \alpha_{i+h} \alpha_{i+s} \alpha_{i+s+h}| = \left(\sum_{i \in \mathbb{Z}} |\alpha_i \alpha_{i+h}| \right)^2 \leq \|\alpha\|_1^4,$$

où la dernière égalité utilise la majoration triviale $|\alpha_{i+h}| \leq \|\alpha\|_1$. □