

Analyse des Séries Temporelles et Applications

Céline Lévy-Leduc, Eric Moulines, François Roueff
Télécom ParisTech

9 mai 2019

Table des matières

I	Processus au second ordre, représentation spectrale et prédiction	3
1	Processus aléatoires et stationnarité	4
1.1	Quelques exemples	4
1.2	Définition et construction de la loi d'un processus aléatoire	6
1.2.1	Processus aléatoire	6
1.2.2	Répartitions finies	6
1.2.3	Stationnarité stricte d'un processus à temps discret	9
1.2.4	Processus gaussiens	10
2	Rappels sur la transformée de Fourier	12
3	Processus stationnaires au second ordre	14
3.1	Processus du second ordre	14
3.2	Covariance d'un processus stationnaire au second ordre	15
3.2.1	Propriétés	15
3.2.2	Interprétation de la fonction d'autocovariance	18
3.3	Mesure spectrale d'un processus stationnaire au second ordre à temps discret	21
3.4	Représentation harmonique des processus	25
3.4.1	Champ aléatoire non corrélé et intégrale stochastique	25
3.4.2	Intégrale stochastique par rapport à un champ	27
3.4.3	Champ spectral associé à un processus et représentation spectrale	28
4	Filtrage des processus stationnaires au second ordre	30
4.1	Filtrages linéaires de processus	30
4.2	Processus ARMA	32
4.2.1	Processus MA(q)	32
4.2.2	Processus AR(p)	33
4.2.3	Cas général	36
4.3	Preuves des théorèmes 4.1.1 et 4.1.2	40
5	Prédiction des processus stationnaires au second ordre	43
5.1	Prédiction linéaire	43
5.1.1	Estimation linéaire en moyenne quadratique	43
5.1.2	Prédiction linéaire d'un processus stationnaire au second-ordre	44

5.2	Algorithme de Levinson-Durbin	49
5.3	Algorithme de Schur	52
5.4	Algorithme des innovations	55
5.5	Décomposition de Wold	56
5.6	Preuves des théorèmes 5.1.5 et 5.5.6	61
II	Estimation spectrale	64
6	Estimation de la moyenne et de la fonction d'autocovariance	65
6.1	Estimation de la moyenne	65
6.2	Estimation des coefficients d'autocovariance et d'autocorrélation	67
6.3	Preuve du Théorème 6.1.2	74
6.4	Preuves des Théorèmes 6.2.4 et 6.2.3	77
7	Estimation de la densité spectrale	84
7.1	Le périodogramme	84
7.2	Estimateur à noyau	90
7.3	Preuves des théorèmes 7.1.2, 7.1.4	94
III	Annexes	99
A	Eléments d'Analyse Hilbertienne	100
A.1	Définitions	100
A.2	Projection et principe d'orthogonalité	104
A.3	Bases Hilbertiennes	107
A.3.1	Isométries et isomorphismes d'espaces de Hilbert	111
B	Statistique Asymptotique	113
B.1	Notions de convergence	113
B.2	Suites tendues	115
B.3	Caractérisations de la convergence en loi	117
B.4	théorème de continuité	120
B.5	Loi des grands nombres	121
B.6	théorème de la limite centrale	121
B.7	Symboles o et O stochastiques	125
B.8	Delta-méthode	127
B.9	Convergence des moments	128

Première partie

Processus au second ordre,
représentation spectrale et prédiction

Chapitre 1

Processus aléatoires et stationnarité

1.1 Quelques exemples

Le paragraphe 1.2 définit le formalisme probabiliste permettant de décrire les *processus aléatoires*. Les quelques exemples qui suivent illustrent la diversité des situations dans lesquelles la modélisation stochastique (ou aléatoire) des séries temporelles joue un rôle important.

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

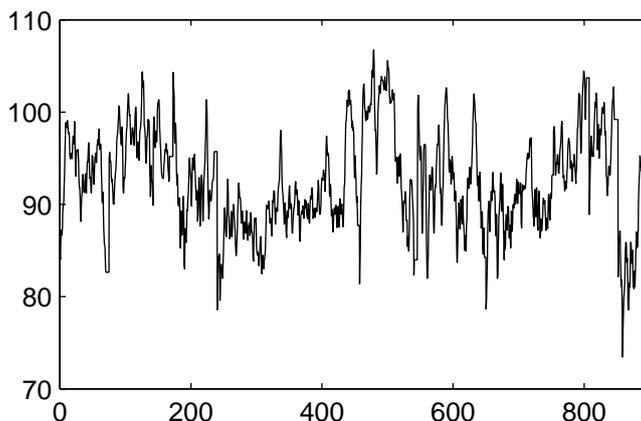


FIGURE 1.1 – *Battements cardiaques : évolution du nombre de battements par minute en fonction du temps mesuré en seconde.*

Exemple 1.1.2 (Trafic internet). *La figure 1.2 représente les temps d'inter-arrivées de paquets TCP, mesurés en secondes, sur la passerelle du laboratoire Lawrence Livermore. La trace représentée a été obtenue en enregistrant 2 heures de trafic. Pendant cette durée, environ 1.3 millions de paquets TCP, UDP, etc. ont été enregistrés, en utilisant la procédure tcpdump sur une station Sun. D'autres séries de ce type peuvent être obtenues sur The Internet Traffic Archive, <http://ita.ee.lbl.gov/>.*

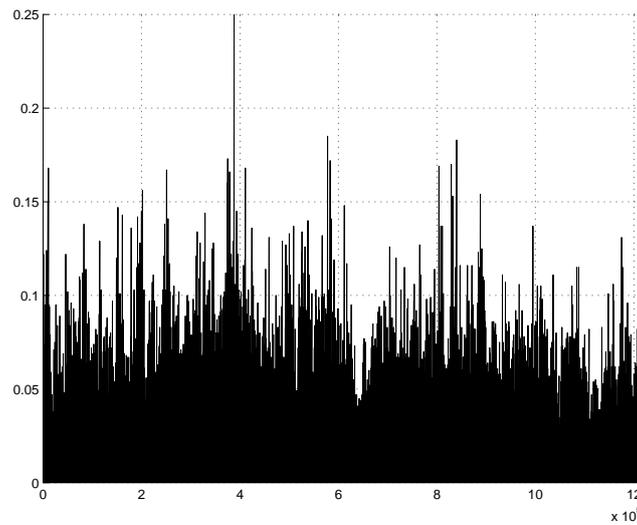


FIGURE 1.2 – Trace de trafic Internet : temps d'inter-arrivées de paquets TCP.

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

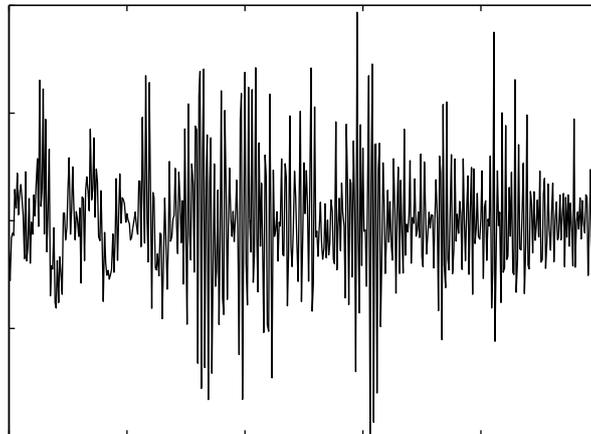


FIGURE 1.3 – Signal de parole échantillonné à 8000 Hz : son non voisé *ch*.

Exemple 1.1.4 (Indice financier). La figure 1.4 représente les cours d'ouverture journaliers de l'indice Standard and Poor 500, du 2 Janvier 1990 au 25 Août 2000. L'indice S&P500 est calculé à partir de 500 actions choisies parmi les valeurs cotées au New York Stock Exchange (NYSE) et au NASDAQ en fonction de leur capitalisation, leur liquidité, leur représentativité dans différents secteurs d'activité.

Cet indice est obtenu en pondérant le prix des actions par le nombre total d'actions, le poids de chaque valeur dans l'indice composite étant proportionnel à la capitalisation.

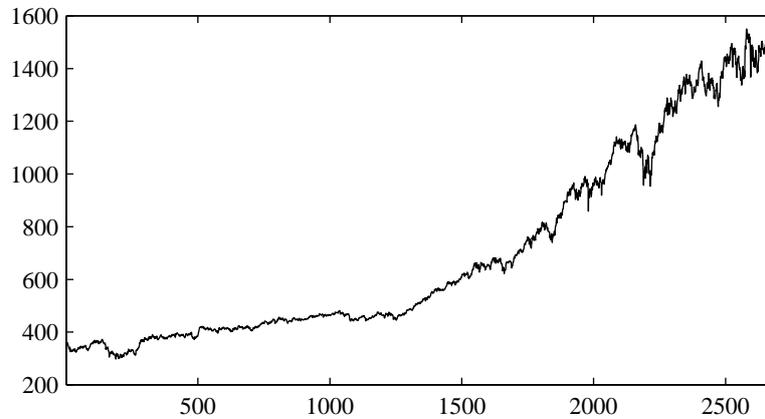


FIGURE 1.4 – Cours quotidien d'ouverture de l'indice S&P500 : entre Janvier 1990 et Août 2000.

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

1.2.1 Processus aléatoire

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

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

Notons qu'en fait un processus est une application $X : \Omega \times T \rightarrow E$, $(\omega, t) \mapsto X_t(\omega)$ telle que :

- à chaque instant $t \in T$, l'application $\omega \mapsto X_t(\omega) \in (E, \mathcal{E})$ est une variable aléatoire,
- pour chaque épreuve $\omega \in \Omega$, l'application $t \mapsto X_t(\omega)$ est une fonction de $T \rightarrow E$ qui s'appelle la *trajectoire* associée à l'épreuve ω .

1.2.2 Répartitions finies

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

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

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

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

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

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

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

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

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

Définition 1.2.2. On appelle famille des répartitions finies l'ensemble des répartitions finies $(\mathbb{P}_I, I \in \mathcal{I})$.

La spécification de la mesure \mathbb{P}_I permet de calculer la probabilité d'événements de la forme $\mathbb{P}(\cap_{t \in I} \{X_t \in A_t\})$ où $(A_t, t \in I)$ sont des éléments de la tribu \mathcal{E} , ou de manière équivalente, de calculer l'espérance $\mathbb{E} \left[\prod_{t \in I} f_t(X_t) \right]$ où $(f_t, t \in I)$ sont des fonctions boréliennes positives. Il est important de noter que, la donnée des répartitions finies ne permet pas *directement* d'évaluer la probabilité d'un événement faisant intervenir un nombre infini d'indices de temps ; par exemple, pour un processus à temps discret indexé par $T = \mathbb{Z}$, les répartitions finies ne permettent pas d'évaluer directement la probabilité d'un événement de la forme $\{\sup_{t \in T} X_t \geq a\}$. Soit $J \subset I$ deux parties finies ordonnées. Soit $\Pi_{I,J}$ la projection canonique de E^I sur E^J définie par

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

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

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

Cette relation formalise le résultat intuitif que la distribution fini-dimensionnelle d'un sous-ensemble $J \subset I$ se déduit de la distribution fini-dimensionnelle \mathbb{P}_I en "intégrant" par rapport aux variables X_t sur l'ensemble des t appartenant au complémentaire de J dans I . Cette propriété montre que la famille des répartitions finies d'un processus est fortement structurée. En particulier, les répartitions

finies doivent, au moins, vérifier les conditions de compatibilité (1.3). Nous allons voir dans la suite que cette condition est en fait aussi *suffisante*.

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

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

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

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

Théorème 1.2.3 (théorème de Kolmogorov). *On pose $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ pour $d \geq 1$. Soit $\{\nu_I, I \in \mathcal{I}\}$ une famille de probabilités indexées par l'ensemble des parties finies ordonnées de T telle, que pour tout $I \in \mathcal{I}$, ν_I est une probabilité sur $(E^I, \mathcal{E}^{\otimes I})$. Supposons de plus que la famille $\{\nu_I, I \in \mathcal{I}\}$ vérifie les conditions de compatibilité (1.3), pour tout $I, J \in \mathcal{I}$, tel que $I \subset J$, $\nu_I \circ \Pi_{I,J}^{-1} = \nu_J$. Il existe une probabilité unique \mathbb{P} sur l'espace mesurable $(E^T, \mathcal{E}^{\otimes T})$ telle que, pour tout $I \in \mathcal{I}$, $\nu_I = \mathbb{P} \circ \Pi_I^{-1}$.*

Démonstration. Comme la classe des cylindres est stable par intersection et engendre la tribu \mathcal{E}^T , il est clair que la relation $\nu_I = \mathbb{P} \circ \Pi_I^{-1}$ implique l'unicité de \mathbb{P} . On admet l'existence (voir [Kallenberg, 2002, Theorem 6.16] pour une preuve complète) sous l'hypothèse où (E, \mathcal{E}) un espace mesurable borélien c'est-à-dire pour lequel il existe une bijection mesurable d'inverse mesurable de E dans un borélien inclus dans $[0, 1]$ (en particulier $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ convient pour tout $d \geq 1$). \square

Définition 1.2.4 (Processus canonique). *Soit (E, \mathcal{E}) un espace mesurable et (E^T, \mathcal{E}^T) l'espace mesurable des trajectoires correspondants. La famille canonique sur (E^T, \mathcal{E}^T) est la famille des fonctions mesurables $\{\xi_t, t \in T\}$ définies sur (E^T, \mathcal{E}^T) à valeurs dans (E, \mathcal{E}) par $\xi_t(\omega) = \omega_t$ pour tout $\omega = (\omega_t)_{t \in T} \in E^T$.*

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

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

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

L'existence et l'unicité de \mathbb{P}_X est donnée par le théorème 1.2.3. On l'appellera aussi plus simplement la *loi du processus X* . Cette loi est donc *entièrement* déterminée par la donnée des répartitions finies.

Exemple 1.2.5 (Suite de v.a. indépendantes). *Soit $(\nu_n, n \in \mathbb{N})$ une suite de probabilités sur (E, \mathcal{E}) . Pour $I \in \mathcal{I}$, on pose*

$$\nu_I = \bigotimes_{n \in I} \nu_n, \quad (1.6)$$

où \otimes désigne le produit tensoriel sur les probabilités (loi du vecteur à composantes indépendantes et de lois marginales données par les $\nu_n, n \in I$). Il est clair que l'on définit ainsi une famille $(\nu_I, I \in \mathcal{I})$ compatible, c'est-à-dire, vérifiant la condition donnée par l'équation (1.3). Donc, si $\Omega = E^{\mathbb{N}}$, $X_n(\omega) = \omega_n$ et $\mathcal{F} = \sigma(X_n, n \in \mathbb{N})$, il existe une unique probabilité \mathbb{P} sur (Ω, \mathcal{F}) telle que $(X_n, n \in \mathbb{N})$ soit une suite de v.a. indépendantes de loi ν_n .

1.2.3 Stationnarité stricte d'un processus à temps discret

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

Définition 1.2.6 (Opérateurs de décalage et de retard). *On pose $T = \mathbb{Z}$ ou $T = \mathbb{N}$. On note S et l'on appelle opérateur de décalage (Shift) l'application $E^T \rightarrow E^T$ définie par*

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

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

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

Définition 1.2.7 (Stationnarité stricte). *On pose $T = \mathbb{Z}$ ou $T = \mathbb{N}$. Un processus aléatoire $\{X_t, t \in T\}$ est stationnaire au sens strict si X et $S \circ X$ ont même loi, i.e. $\mathbb{P}_{S \circ X} = \mathbb{P}_X$.*

Par définition de la loi image on a $\mathbb{P}_{S \circ X} = \mathbb{P}_X$ si et seulement si

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

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

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

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

Exemple 1.2.8 (Processus i.i.d et transformations). *Soit $\{Z(t)\}$ une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d). $\{Z(t)\}$ est un processus stationnaire au sens strict, car, pour toute partie finie ordonnée $I = \{t_1, < t_2 < \dots < t_n\}$ nous avons :*

$$\mathbb{P}(Z(t_1) \in A_1, \dots, Z(t_n) \in A_n) = \prod_{j=1}^n \mathbb{P}(Z(0) \in A_j)$$

Soient k un entier et g une fonction borélienne de \mathbb{R}^k dans \mathbb{R} . Il est facile de vérifier que le processus aléatoire $\{X_t\}$ défini par

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

est encore un processus aléatoire stationnaire au sens strict. Par contre, ce processus obtenu par transformation n'est plus i.i.d dans la mesure où, dès que $k \geq 1$, $X_t, X_{t+1}, \dots, X_{t+k-1}$ bien qu'ils aient la même distribution marginale sont, en général, dépendants car fonctions de variables aléatoires communes. Un tel processus est dit k -dépendant dans la mesure où, par contre, $\tau \geq k$ implique que X_t et $X_{t+\tau}$ sont indépendants (ils dépendent de deux groupes indépendants de k variables aléatoires).

1.2.4 Processus gaussiens

Définition 1.2.9 (Variable aléatoire gaussienne réelle). *On dit que X est une variable aléatoire réelle gaussienne si sa loi de probabilité a pour fonction caractéristique :*

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

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

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

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

Définition 1.2.10 (Vecteur gaussien réel). *Un vecteur aléatoire réel de dimension n (X_1, \dots, X_n) est un vecteur gaussien si toute combinaison linéaire de X_1, \dots, X_n est une variable aléatoire gaussienne réelle.*

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

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

On en déduit l'expression, en fonction de μ et de Γ , de la fonction caractéristique de la loi de probabilité d'un vecteur gaussien $X(1), \dots, X(n)$:

$$\phi_X(u) = \mathbb{E} [\exp(iu^T X)] = \mathbb{E} [\exp(iY)] = \exp\left(iu^T \mu - \frac{1}{2}u^T \Gamma u\right) \quad (1.7)$$

De plus si Γ est de rang plein n , alors la loi de probabilité de X possède une densité dont l'expression est :

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

Dans le cas où Γ est de rang $r < n$, c'est à dire où Γ possède $n - r$ valeurs propres nulles, X se trouve, avec probabilité 1, dans un sous espace de dimension r de \mathbb{R}^n , dans la mesure où il existe $r - n$ combinaisons linéaires indépendantes a_i telles que $\text{cov}(a_i^T X) = 0$.

Définition 1.2.11 (Processus gaussien réel). *On dit qu'un processus réel $X = \{X_t, t \in T\}$ est gaussien si, pour toute suite finie d'instantes $\{t_1 < t_2 < \dots < t_n\}$, $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ est un vecteur gaussien.*

1. Dans cet ouvrage, les vecteurs sont par convention identifiés sous forme matricielle à des vecteurs colonnes et l'exposant T indique l'opérateur de transposition des matrices.

D'après (1.7), la famille des répartitions finies est donc caractérisée par la donnée de la fonction moyenne $\mu : t \in T \mapsto \mu(t) \in \mathbb{R}$ et de la fonction de covariance $\gamma : (t, s) \in (T \times T) \mapsto \gamma(t, s) \in \mathbb{R}$. Réciproquement, donnons nous une fonction $\mu : t \in T \mapsto \mu(t) \in \mathbb{R}$ et une fonction de covariance $\gamma : (t, s) \in (T \times T) \mapsto \gamma(t, s) \in \mathbb{R}$ de type positif, c'est-à-dire telle que, pour tout n , toute suite (u_1, \dots, u_n) et toute suite (t_1, \dots, t_n) on ait :

$$\sum_{j=1}^n \sum_{k=1}^n u_j u_k \gamma(t_j, t_k) \geq 0 \quad (1.8)$$

On peut alors définir, pour $I = \{t_1 < \dots < t_n\}$, une probabilité gaussienne ν_I sur \mathbb{R}^n par :

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

où $\mu_I = (\mu(t_1), \dots, \mu(t_n))$ et Γ_I est la matrice positive d'éléments $\gamma_I(m, k) = \gamma(t_m, t_k)$, où $1 \leq m, k \leq n$. La famille $(\nu_I, I \in \mathcal{I})$, ainsi définie, vérifie les conditions de compatibilité et l'on a ainsi établi, d'après le théorème 1.2.3, le résultat suivant :

Théorème 1.2.12. *Soit $r \mapsto \mu(r)$ une fonction et $(s, t) \mapsto \gamma(s, t)$ une fonction de type positif (vérifiant l'équation (1.8)). Il existe un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et un processus aléatoire $\{X_t, t \in T\}$ gaussien défini sur cet espace vérifiant*

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

Chapitre 2

Rappels sur la transformée de Fourier

Dans toute la suite, I désigne l'intervalle $I = [-\pi, \pi]$ et $\mathcal{B}(I)$ la tribu de Borel de I construite sur les ouverts de I .

Proposition 2.0.1 (Transformée de Fourier discrète d'une suite sommable). *Soit $R(n)$ une suite complexes de module sommable. Alors :*

$$R(n) = \int_I e^{in\lambda} f(\lambda) d\lambda \quad \text{où} \quad f(\lambda) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} R(n) e^{-in\lambda}$$

Une application directe du théorème de Fubini donne :

$$\int_I e^{in\lambda} f(\lambda) d\lambda = \int_I e^{in\lambda} \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} R(k) e^{-ik\lambda} d\lambda = \sum_{k=-\infty}^{\infty} R(k) \frac{1}{2\pi} \int_I e^{i(n-k)\lambda} d\lambda = R(n)$$

Proposition 2.0.2 (Coefficients de Fourier d'une mesure finie). *Soit ν une mesure non-négative, définie sur $\{I, \mathcal{B}(I)\}$, finie (i.e. telle que $\int_I \nu(d\lambda) < +\infty$) et soit $k \in \mathbb{Z}$. On appelle k -ième coefficient de Fourier de ν :*

$$\hat{\nu}(k) = \int_I e^{ik\lambda} \nu(d\lambda)$$

1. L'application $\nu \rightarrow \hat{\nu}$ est injective.
 2. La suite $\{\hat{\nu}(k)\}$ est de type non-négatif.
 3. Soit $\{\nu_n\}_{n \geq 0}$ et ν des mesures finies. La suite de mesures $\{\nu_n\}$ converge étroitement vers la mesure ν (quand n tend vers l'infini), si et seulement si, pour tout $k \in \mathbb{Z}$, $\hat{\nu}_n(k)$ converge vers $\hat{\nu}(k)$ (quand n tend vers l'infini).
1. $\mathcal{C}_b(I)$ désigne l'ensemble des fonctions complexes, continues et bornées, définies sur $I = [-\pi, \pi]$, muni de la topologie associée à la norme uniforme $\|f\|_\infty = \sup_{\lambda \in [-\pi, \pi]} |f(\lambda)|$. Précisons que l'égalité $\nu_1 = \nu_2$ doit être comprise dans le sens où $\int_I f(\lambda) \nu_1(d\lambda) = \int_I f(\lambda) \nu_2(d\lambda)$ pour toute fonction $f \in \mathcal{C}_b(I)$. Le point 1 est alors une conséquence directe du fait que les combinaisons linéaires d'exponentielles complexes, de la forme $e^{in\lambda}$, sont denses dans $\mathcal{C}_b(I)$. L'application qui, à tout $f \in \mathcal{C}_b(I)$ fait correspondre le nombre complexe $c_\nu(f) = \int f(\lambda) \nu(d\lambda) \in \mathbb{C}$ est une

forme linéaire continue sur $\mathcal{C}_b(I)$, qui associe aux exponentielles complexes de la forme $e^{in\lambda}$ les coefficients de Fourier $c_\nu(e^{in\bullet}) = \hat{\nu}(n)$. Par conséquent, si pour deux mesures ν et ν , les formes linéaires associées, c_ν et c_ν , coïncident pour les exponentielles complexes (i.e. $\hat{\nu}(n) = \hat{\nu}(n)$), alors elles coïncident pour toute fonction de $\mathcal{C}_b(I)$. Ce qui démontre le point 1.

2. Soit (z_1, z_2, \dots, z_n) des nombres complexes. On a :

$$\sum_{r,s=1}^d z_r z_s^* \hat{\nu}(r-s) = \int_I \sum_{r,s=1}^d z_s z_r^* e^{i(r-s)\lambda} \nu(d\lambda) = \int_I \left| \sum_{r=1}^d z_r e^{-ir\lambda} \right|^2 \nu(d\lambda) \geq 0$$

3. Par définition, la suite de mesure ν_n converge étroitement vers ν si pour toute fonction $f \in \mathcal{C}_b(I)$, $\lim_n c_{\nu_n}(f) = c_\nu(f)$. En particulier, si on prend $f = e^{-ik\bullet}$ (qui est continue et bornée), nous avons $c_{\nu_n}(e^{ik\bullet}) = \hat{\nu}_n(k) \rightarrow \hat{\nu}(k)$. Réciproquement, soit $\{\nu_n\}$ une suite de mesures finies sur I telles que, pour tout $k \in \mathbb{Z}$, $\lim_n \hat{\nu}_n(k) = \hat{\nu}(k)$. Cette propriété implique en particulier que la suite $\hat{\nu}_n(0) = \nu_n(I)$ est convergente, et est donc bornée, $\sup_{n \geq 0} \hat{\nu}_n(0) < \infty$. Remarquons aussi que $|\hat{\nu}_n(k)| \leq \nu_n(0)$. Pour $f \in L^2(I, d\lambda)$ (où $d\lambda$ désigne la mesure de Lebesgue), définissons :

$$\hat{f}(k) = \int_I f(t) e^{-ikt} dt$$

Considérons la classe \mathcal{F} de fonctions f vérifiant $\sum_{k \in \mathbb{Z}} |\hat{f}(k)| < \infty$. La classe \mathcal{F} est dense dans $\mathcal{C}_b(I)$. Notons que, pour toute fonction $f \in \mathcal{F}$, nous avons :

$$f(\lambda) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \hat{f}(k) e^{-ik\lambda}$$

Par conséquent, en appliquant le théorème de Fubini, on a :

$$c_{\nu_n}(f) = \int_I f(\lambda) \nu_n(d\lambda) = \frac{1}{2\pi} \int_I \sum_{k \in \mathbb{Z}} \hat{f}(k) e^{-ik\lambda} \nu_n(d\lambda) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \hat{f}(k) \hat{\nu}_n(k)$$

Comme $\sup_k \sup_n |\hat{\nu}_n(k)| < \infty$, le théorème de convergence dominée et le théorème de Fubini impliquent que :

$$\lim_n c_{\nu_n}(f) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \hat{f}(k) \lim_{n \rightarrow +\infty} \hat{\nu}_n(k) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \hat{f}(k) \hat{\nu}(k) = c_\nu(f)$$

Soit maintenant f une fonction continue. Pour tout $\epsilon > 0$, il existe $f_\epsilon \in \mathcal{F}$ tel que $\|f - f_\epsilon\|_\infty \leq \epsilon$ et nous avons :

$$\begin{aligned} |\nu_n(f) - \nu(f)| &\leq |\nu_n(f_\epsilon) - \nu(f_\epsilon)| + |\nu(f_\epsilon) - \nu(f)| \\ &\leq |\nu_n(f_\epsilon) - \nu(f_\epsilon)| + \|f - f_\epsilon\|_\infty (|\hat{\nu}_n(0)| + |\hat{\nu}(0)|) \end{aligned}$$

et donc puisque $f_\epsilon \in \mathcal{F}$ la limite du premier terme est 0 et on a :

$$\limsup_n |\nu_n(f) - \nu(f)| \leq 2\epsilon |\hat{\nu}(0)|$$

Comme ϵ est arbitraire, nous avons donc $\lim_n \nu_n(f) = \nu(f)$, ce qui conclut la preuve.

Chapitre 3

Processus stationnaires au second ordre

3.1 Processus du second ordre

On a vu dans l'exemple A.1.5 que l'espace $L^2(\Omega)$ des v.a. aléatoires de variance finie est un espace de Hilbert. Pour profiter des propriétés de ces espaces il est donc naturel de travailler sur des processus faisant intervenir des v.a. de cet espace.

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

Notons que la *moyenne* $\mu(t) = \mathbb{E}[X_t]$ est un vecteur de dimension d dépendant de t et que la *fonction d'autocovariance* définie en utilisant l'exposant H pour indiquer l'opération de transposition et conjugaison par

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

est une matrice de dimension $d \times d$ dépendant à la fois de s et de t .

Proposition 3.1.2. *Pour un processus du second ordre on a :*

1. $\Gamma(s, s) \geq 0$, l'égalité ayant lieu si et seulement si X_s est presque sûrement égale à sa moyenne.
2. *Symétrie hermitienne*

$$\Gamma(s, t) = \Gamma(t, s)^H \tag{3.1}$$

3. *Type positif*

Pour tout n , pour toute suite d'instants $(t_1 < t_2 < \dots < t_n)$ et pour toute suite de vecteurs complexes (a_1, \dots, a_n) de dimension d , on a :

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

Démonstration. Formons la combinaison linéaire $Y = \sum_{k=1}^n a_k^H X_{t_k}$. Y est une variable aléatoire complexe. Sa variance, qui est positive, s'écrit

$$\text{var}(Y) = \mathbb{E}[|Y - \mathbb{E}[Y]|^2] \geq 0$$

On note $X_t^c = X_t - \mathbb{E}[X_t]$ le processus centré. En développant $\text{var}(Y)$ en fonction de $X_{t_k}^c$, il vient :

$$\text{var}(Y) = \mathbb{E} \left[\sum_{k=1}^n \lambda_k^H X_{t_k}^c \sum_{m=1}^n X_{t_m}^{cT} \lambda_k \right] = \sum_{1 \leq k, m \leq n} \lambda_k^H \Gamma(t_k, t_m) \lambda_m$$

ce qui établit (3.2). □

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

3.2 Covariance d'un processus stationnaire au second ordre

On définit la stationnarité au second ordre en ne retenant que les propriétés du second ordre (moyenne et covariance) d'un processus stationnaire au sens stricte indexé par \mathbb{Z} . Cela donne la définition suivante.

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

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

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

3.2.1 Propriétés

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

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

$$\gamma(-h) = \gamma^*(h)$$

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

$$\sum_{k=1}^n \sum_{j=1}^n a_k^* \gamma(k - j) a_j \geq 0$$

Ces propriétés découlent immédiatement des propriétés de la fonction d'autocovariance d'un processus. La matrice de covariance de n valeurs consécutives X_1, \dots, X_n du processus est donc hermitienne positive. Elle possède de plus une structure particulière, dite de *Toeplitz*, caractérisée par le fait que $(\Gamma_n)_{ij} = \gamma(i - j)$. On obtient une matrice de la forme

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

Définition 3.2.3 (Fonction d'autocorrélation). *Pour un processus stationnaire, on appelle fonction d'autocorrélation $\rho(h) = \gamma(h)/\gamma(0)$. Il s'agit d'une quantité normalisée dans le sens où $\rho(1) = 1$ et $|\rho(k)| \leq 1$.*

En effet, l'inégalité de Cauchy-Schwarz (voir le théorème A.1.6) appliquée à $\gamma(k)$ implique

$$|\gamma(h)| = |\mathbb{E}[(X_{t+h} - \mu_X)(X_t - \mu_X)^*]| \leq \sqrt{\mathbb{E}[|X_{t+h} - \mu_X|^2] \mathbb{E}[|X_t - \mu_X|^2]} = \gamma(0)$$

la dernière inégalité découlant de l'hypothèse de stationnarité. Attention, certaines références (livres et publications), en général anciennes, utilisent (incorrectement) le terme de "fonction d'autocorrélation" pour $\gamma(h)$. Dans la suite de ce document, le terme autocorrélation est réservée à la quantité normalisée $\rho(h)$.

Exemple 3.2.4 (Processus retourné temporel). *Soit X_t un processus aléatoire stationnaire au second ordre à valeurs réelles de moyenne μ_X et de fonction d'autocovariance $\gamma_X(h)$. On note $X_t^r = X_{-t}$ le processus retourné temporel. Alors X_t^r est un processus stationnaire au second ordre de même moyenne et de même fonction d'autocovariance que le processus X_t . En effet on a :*

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

Définition 3.2.5 (Bruit blanc). *On appelle bruit blanc un processus aléatoire stationnaire au second ordre à valeurs réelles, centré, de fonction d'autocovariance, $\gamma(s, t) = \gamma(t - s) = \sigma^2 \delta_{t,s}$. On le notera $\{X_t\} \sim \text{BB}(0, \sigma^2)$.*

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

Par définition si $\{X_t\} \sim \text{IID}(0, \sigma^2)$, $\mathbb{E}[X_t] = 0$, $\mathbb{E}[X_t^2] = \sigma^2$ et pour tout $h \neq 0$, $\mathbb{E}[X_{t+h}X_t] = \mathbb{E}[X_{t+h}]\mathbb{E}[X_t] = 0$. $\{X_t\}$ est donc également stationnaire au second ordre, de fonction d'autocovariance $\gamma(s, t) = \sigma^2 \delta(t - s)$. La structure de bruit blanc fort est clairement plus contraignante que celle de simple bruit blanc. En général, il est tout à fait inutile de faire une telle hypothèse lorsque l'on s'intéresse à des processus stationnaires au second ordre. Il arrivera cependant dans la suite que nous adoptions cette hypothèse plus forte afin de simplifier les développements mathématiques. Notons que dans le cas d'un processus gaussienne, ces deux notions sont confondues puisque la loi gaussienne est complètement caractérisée par les moments du premier et du second ordre (un bruit blanc gaussien est donc également un bruit blanc fort).

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

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

où $\{Z_t\} \sim \text{BB}(0, \sigma^2)$ et $\theta \in \mathbb{R}$. On vérifie aisément que $\mathbb{E}[X_t] = 0$ et que :

$$\gamma(t, s) = \begin{cases} \sigma^2(1 + \theta^2) & t = s \\ \sigma^2\theta & |t - s| = 1 \\ 0 & |t - s| > 1 \end{cases}$$

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

Exemple 3.2.8 (Processus harmonique). Soient $\{A_k\}_{1 \leq k \leq N}$ N variables aléatoires vérifiant $\text{cov}(A_k, A_l) = \sigma_k^2 \delta(k - l)$ et $\{\Phi_k\}_{1 \leq k \leq N}$, N variables aléatoires indépendantes et identiquement distribuées (i.i.d), de loi uniforme sur $[-\pi, \pi]$, et indépendantes de $\{A_k\}_{1 \leq k \leq N}$. On définit :

$$X_t = \sum_{k=1}^N A_k \cos(\lambda_k t + \Phi_k) \quad (3.5)$$

où $\{\lambda_k\} \in [-\pi, \pi]$ sont N pulsations. Le processus X_t est appelé processus harmonique. On vérifie aisément que $\mathbb{E}[X_t] = 0$ et que sa fonction d'autocovariance est donnée par :

$$\gamma(h) = \mathbb{E}[X_{t+h}X_t] = \frac{1}{2} \sum_{k=1}^N \sigma_k^2 \cos(\lambda_k h)$$

Le processus harmonique est donc stationnaire au second ordre.

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

$$\gamma(t+h, t) = \mathbb{E}[(S_t + X_{t+1} + \dots + X_{t+h})S_t] = t\sigma^2$$

Le processus $\{S_t\}$ n'est donc pas stationnaire au second ordre.

Exemple 3.2.10. Nous allons montrer que la suite définie, pour $h \in \mathbb{Z}$, par :

$$\chi(h) = \begin{cases} 1 & h = 0, \\ \rho & |h| = 1 \\ 0 & |h| \geq 2 \end{cases}$$

est la fonction d'autocovariance d'un processus stationnaire au second ordre si et seulement si $|\rho| \leq 1/2$. Nous avons déjà montré exemple 3.2.7 que la fonction d'autocovariance d'un processus MA(1) est donnée par :

$$\gamma(h) = \begin{cases} \sigma^2(1 + \theta^2) & \text{pour } h = 0 \\ \sigma^2\theta & \text{pour } |h| = 1 \\ 0 & \text{pour } |h| > 1 \end{cases}$$

La suite $(\chi(h))$ est donc la fonction d'autocovariance d'un processus $MA(1)$ si et seulement si $\sigma^2(1 + \theta^2) = 1$ et $\sigma^2\theta = \rho$. Lorsque $|\rho| \leq 1/2$, ce système d'équations admet comme solution :

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

Lorsque $|\rho| > 1/2$, ce système d'équations n'admet pas de solution réelles et la suite $(\chi(h))$ n'est donc pas la fonction d'autocovariance d'un processus $MA(1)$. On vérifie facilement que $(\chi(h))$ ne vérifie pas dans ce cas la condition de positivité (en prenant $a_k = (-1)^k$ pour $\rho > 1/2$ et $a_k = 1$ dans le cas opposé). Pour $|\rho| > 1/2$, $(\chi(h))$ n'est donc pas une séquence d'autocovariance.

3.2.2 Interprétation de la fonction d'autocovariance

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

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

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

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

Lorsqu'il est *a priori* raisonnable de penser que la série considérée est stationnaire au second ordre, la moyenne empirique, la fonction d'autocovariance empirique et la fonction d'autocorrélation empirique sont de "bons" estimateurs, dans un sens que nous préciserons chapitre 6. L'analyse de la fonction d'autocovariance empirique est un élément permettant de guider le choix d'un modèle approprié pour les observations. Par exemple, le fait que la fonction d'autocovariance empirique soit *proche* de zéro pour tout $h \neq 0$ (proximité qu'il faudra définir dans un sens statistique précis) indique par exemple qu'un bruit blanc est un modèle adéquat pour les données. La figure 3.1 représente les 100 premières valeurs de la fonction d'autocorrélation empirique de la série des battements cardiaques représentés figure 1.1. On observe que cette série est *positivement corrélée* c'est-à-dire que les fonctions coefficients d'autocorrélation sont positifs et significativement non nuls. Nous avons, à titre de comparaison, représenté aussi la fonction d'autocorrélation empirique d'une trajectoire de même longueur d'un bruit blanc gaussien. Une forte corrélation peut être interprétée comme l'indice d'une dépendance linéaire. Ainsi la figure 3.2 montre que le fait que $\hat{\rho}(1) = 0.966$ pour la série des battements cardiaques se traduit par une très forte prédictabilité de X_{t+1} en fonction de X_t (les couples de points successifs s'alignent quasiment sur une droite). Nous montrerons au chapitre 5, que dans un tel contexte, $\mathbb{E}[(X_{t+1} - \mu) - \rho(1)(X_t - \mu)] = (1 - \rho^2)\text{cov}(X_t)$, c'est à dire, compte tenu de la valeur estimée pour

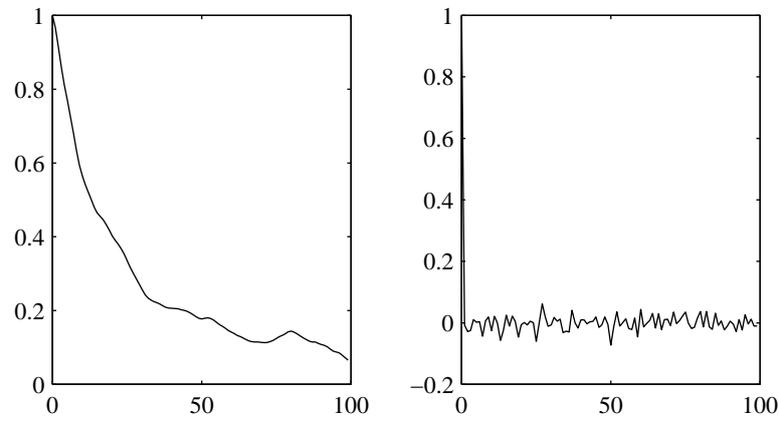


FIGURE 3.1 – Courbe de gauche : fonction d'autocorrélation empirique de la série des battements cardiaques (figure 1.1). Courbe de droite : fonction d'autocorrélation empirique d'une trajectoire de même longueur d'un bruit blanc gaussien.

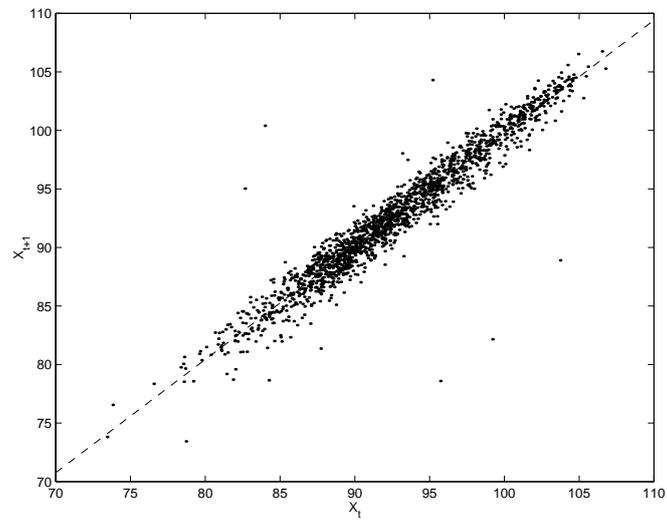


FIGURE 3.2 – X_{t+1} en fonction de X_t pour la série des battements cardiaques de la figure 1.1). Les tirets figurent la meilleure droite de régression linéaire de X_{t+1} sur X_t .

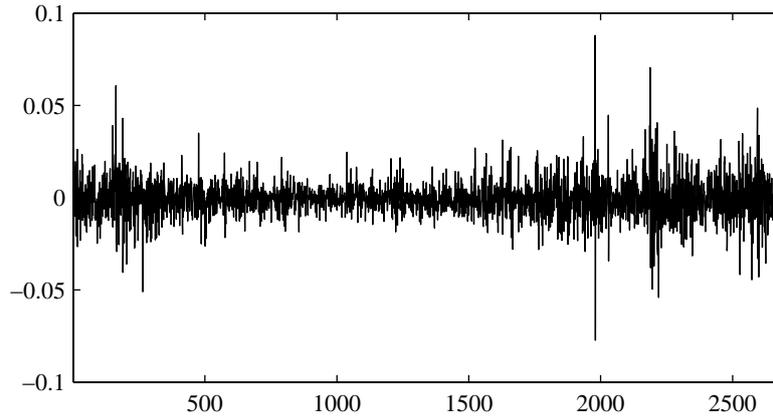


FIGURE 3.3 – *Log-Retour de la série S&P 500 (figure 1.4).*

$\rho(1)$, que la variance de “l’erreur de prédiction” $X_{t+1} - [\mu + \rho(1)(X_t - \mu)]$ est 15 fois plus faible que celle du signal original. L’indice S&P500 tracé (fig. 1.4) présente un cas de figure plus difficile, d’une part parce que la série de départ ne saurait être tenue pour stationnaire et qu’il nous faudra considérer la série des évolutions journalières; d’autre part, parce que selon le choix de la transformation des données considérées, la série transformée présente ou non des effets de corrélation. On définit tout d’abord les *log-retours* de l’indice S&P500 comme les différences des logarithmes de l’indice à deux dates successives :

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

La série des log-retours de la série S&P 500 est représentée figure 3.3. Les coefficients d’autocorrélation

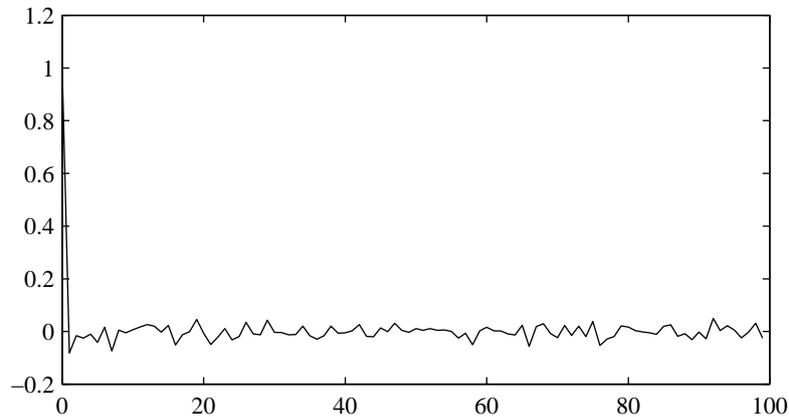


FIGURE 3.4 – *Fonction d’autocorrélation empirique de la série des log-retours de l’indice S&P 500.*

empiriques de la série des log-retours sont représentés figure 3.4. On remarque qu’ils sont approximati-

vement nuls pour $h \neq 0$ ce qui suggère de modéliser la série des log-retours par un bruit blanc (une suite de variables décorréelées). Il est intéressant d'étudier aussi la série des log-retours absolus, $A(t) = |X_t|$.

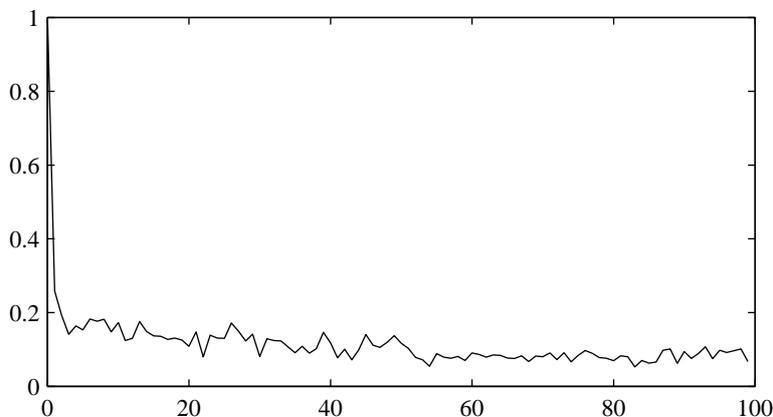


FIGURE 3.5 – Fonction d'autocorrélation empirique de la série des valeurs absolues des log-retours de l'indice S&P 500.

On peut, de la même façon, déterminer la suite des coefficients d'autocorrélation empirique de cette série, qui est représentée dans la figure 3.5. On voit, qu'à l'inverse de la série des log-retours, la série des valeurs absolues des log-retours est positivement corrélée, les valeurs d'autocorrélation étant significativement non nuls pour $|h| \leq 100$. On en déduit, en particulier, que la suite des log-retours peut être modélisée comme un bruit blanc, mais pas un bruit blanc fort : en effet, pour un bruit blanc fort X_t , nous avons, pour toute fonction f telle que $\mathbb{E}[f(X_t)^2] = \sigma_f^2 < \infty$, $\text{cov}(f(X_{t+h}), f(X_t)) = 0$ pour $h \neq 0$ (les variables $f(X_{t+h})$ et $f(X_t)$ étant indépendantes, elles sont a fortiori non corrélées). Nous reviendrons dans la suite du cours sur des modèles possibles pour de telles séries.

3.3 Mesure spectrale d'un processus stationnaire au second ordre à temps discret

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

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

$$\gamma(h) = \int_{\mathbb{T}} e^{ih\lambda} \nu(d\lambda). \quad (3.6)$$

Si la suite $(\gamma(h))$ est de carré sommable (i.e. $\sum_{h \in \mathbb{Z}} \gamma^2(h) < \infty$), la mesure ν possède une densité f

(fonction positive) par rapport à la mesure de Lebesgue sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ et s'écrit donc

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

où f est donnée par la série de Fourier (convergente dans $L^2(\mathbb{T}, \lambda^{\text{Leb}})$ ¹)

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

Remarque 3.3.2. Lorsque γ est la fonction d'autocovariance d'un processus stationnaire au second ordre, on sait d'après la proposition 3.2.2 que $\{\gamma(h)\}_{h \in \mathbb{Z}}$ est de type positif. Les hypothèses du théorème de Herglotz sont donc vérifiées et dans ce cas la mesure ν est appelée la mesure spectrale du processus et la fonction f , lorsque qu'elle existe (i.e. lorsque $\sum_{h \in \mathbb{Z}} |\gamma(h)| < \infty$), est appelée la densité spectrale de puissance du processus.

Démonstration. Si $\gamma(n)$ a la représentation (3.6), montrons que $\gamma(n)$ est de type positif. En effet, pour tout n et toute suite $\{a_k \in \mathbb{C}\}_{1 \leq k \leq n}$,

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

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

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

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

$$\begin{aligned} \int_{\mathbb{T}} e^{ih\lambda} \mu_n(d\lambda) &= \int_{\mathbb{T}} e^{ih\lambda} f_n(\lambda) d\lambda = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) \gamma(k) \int_{\mathbb{T}} e^{i(h-k)\lambda} d\lambda \\ &= \begin{cases} \left(1 - \frac{|h|}{n}\right) \gamma(h), & \text{si } |h| < n , \\ 0, & \text{sinon .} \end{cases} \end{aligned} \quad (3.7)$$

D'après le théorème de Prohorov, il existe une mesure positive μ et une sous-suite μ_{n_k} de μ_n telle que

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

En remplaçant n par n_k dans (3.7) et en faisant tendre k vers l'infini, on a

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

1. voir Théorème A.3.3

Supposons maintenant que $\sum_h |\gamma(h)| < \infty$. Calculons $\int_{\mathbb{T}} e^{ih\lambda} f(\lambda) d\lambda$. Notons que

$$\int_{\mathbb{T}} e^{ih\lambda} f(\lambda) d\lambda = \frac{1}{2\pi} \int_{\mathbb{T}} \sum_{k \in \mathbb{Z}} \gamma(k) e^{i(h-k)\lambda} d\lambda = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \gamma(k) \int_{\mathbb{T}} e^{i(h-k)\lambda} d\lambda = \gamma(h), \quad (3.8)$$

où on a pu intervertir la somme et l'intégrale d'après le théorème de Fubini puisque $\sum_h |\gamma(h)| < \infty$. \square

Proposition 3.3.3 (Corollaire du théorème d'Herglotz). *Une suite $\{\gamma(h), h \in \mathbb{Z}\}$ à valeurs complexes absolument sommable est de type positif si et seulement si la fonction définie par*

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

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

Démonstration. Supposons tout d'abord que γ est absolument sommable et montrons que si $f(\lambda)$ définie dans la proposition est positive sur \mathbb{T} alors $\{\gamma(h)\}$ est de type positif. D'après (3.8), $\gamma(h) = \int_{\mathbb{T}} e^{ih\lambda} f(\lambda) d\lambda$. Comme $f(\lambda) \geq 0$, $f(\lambda) d\lambda$ définit bien une mesure positive sur \mathbb{T} et donc d'après le théorème de Herglotz, $\{\gamma(h)\}$ est de type positif.

Supposons à présent que $\{\gamma(h)\}$ est de type positif et absolument sommable et montrons que $f(\lambda)$ définie dans la proposition est positive sur \mathbb{T} . On a

$$\begin{aligned} 0 \leq f_n(\lambda) &= \frac{1}{2\pi n} \sum_{1 \leq r, s \leq n} e^{-ir\lambda} \gamma(r-s) e^{is\lambda} \\ &= \frac{1}{2\pi} \sum_{|m| < n} \left(1 - \frac{|m|}{n}\right) e^{-im\lambda} \gamma(m) \rightarrow \frac{1}{2\pi} \sum_{m \in \mathbb{Z}} \gamma(m) e^{-im\lambda} = f(\lambda), \text{ lorsque } n \rightarrow \infty, \end{aligned}$$

d'après le théorème de convergence dominée que l'on peut appliquer puisque $\sum_h |\gamma(h)| < \infty$. Ainsi $f(\lambda) \geq 0$ comme limite de fonctions positives. \square

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

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

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

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

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

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

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

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

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

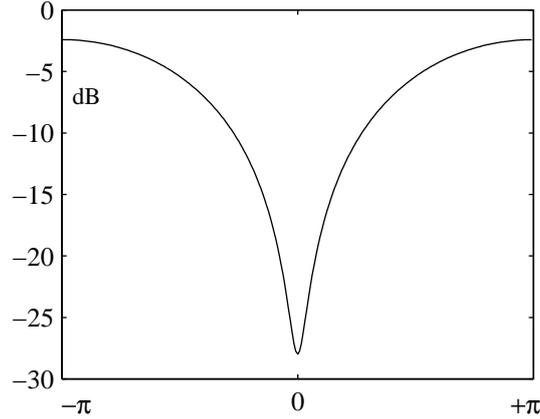


FIGURE 3.6 – Densité spectrale (en dB) d'un processus MA-1, défini par l'équation (3.4) pour $\sigma = 1$ et $\theta = -0.9$.

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

$$\gamma(h) = \frac{1}{2} \sum_{k=1}^N \sigma_k^2 \cos(\lambda_k h) \quad (3.9)$$

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

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

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

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

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

Contrairement aux autres exemples étudiés, le processus harmonique possède une fonction d'autocovariance, donnée par 3.9, non absolument sommable ($\gamma(h)$ ne tend pas même vers 0 pour les grandes valeurs de h). Par suite, il admet une mesure spectrale mais pas une densité spectrale. La propriété suivante, à démontrer à titre d'exercice, implique que le processus harmonique est en fait entièrement prédictible à partir de quelques-unes de ses valeurs passées.

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

L'expression de la fonction d'autocovariance, obtenue en (3.9) pour le processus harmonique, montre que les matrices de covariances associées s'écrivent comme la somme de $2N$ matrices complexes de rang 1. Par conséquent, les matrices Γ_n ne sont pas inversibles dès que $n > 2N$, ce qui implique que le processus harmonique est prédictible dès lors que l'on en a observé $2N$ valeurs. Ce résultat est sans surprise compte tenu du fait que les trajectoires de ce processus sont des sommes de sinusoides de fréquences $\lambda_1, \dots, \lambda_N$ dont seules les amplitudes et les phases sont aléatoires. La propriété suivante donne une condition suffisante simple pour éviter ce type de comportements "extrêmes". Cette propriété implique en particulier que, pour une fonction d'autocovariance absolument sommable (tous les exemples vus ci-dessus en dehors du processus harmoniques), les valeurs futures du processus correspondant ne sont pas prédictibles sans erreur à partir d'un ensemble fini de valeurs passées du processus. Nous reviendrons en détail sur ces problèmes de prédiction au chapitre 5.

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

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

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

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

3.4 Représentation harmonique des processus

3.4.1 Champ aléatoire non corrélé et intégrale stochastique

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

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

Lemme 3.4.2. Soit Z un champ aléatoire non corrélé. Pour $A \in \mathcal{B}(\mathbb{T})$, posons $\eta(A) = \mathbb{E}[Z_A^2]$. Alors η est une mesure finie sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$. De plus, pour tout $A, B \in \mathcal{B}(\mathbb{T})$, $\text{Cov}(Z_A, Z_B) = \eta(A \cap B)$.

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

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

$$\text{Cov}(Z_A, Z_B) = \text{Var}(Z_{A \cap B}) = \eta(A \cap B).$$

□

La mesure η est appelée la *base du champ aléatoire* Z . Remarquons que le Lemme précédent admet la réciproque suivante.

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

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

$$\text{Var}(Z_{A \cup B} - Z_A - Z_B) = \eta(A \cup B) + \eta(A) + \eta(B) - 2\eta(A) - 2\eta(B) + 2\eta(A \cap B) = 0,$$

en utilisant $\eta(A \cup B) = \eta(A) + \eta(B)$ et $\eta(A \cap B) = 0$. Nous avons donc $Z_{A \cup B} = Z_A + Z_B$ et la propriété d'additivité est satisfaite.

Soit $(A_n)_{n \geq 0}$ une suite décroissantes d'éléments de $\mathcal{B}(\mathbb{T})$ telle que $\bigcap_{n=0}^{\infty} A_n = \emptyset$. Comme η est une mesure, nous avons $\lim_{n \rightarrow \infty} \eta(A_n) = 0$ et donc $\lim_{n \rightarrow \infty} \text{Var}(Z_{A_n}) = 0$. La propriété de continuité est aussi satisfaite. □

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

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

où δ_x est la mesure de Dirac. Nous avons

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

où

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

Le lemme 3.4.3 montre que Z est un champ aléatoire décorrélé de mesure de base η .

3.4.2 Intégrale stochastique par rapport à un champ

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

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

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

$$\langle \mathbb{1}_A, \mathbb{1}_B \rangle_{\mathcal{H}} = \int \mathbb{1}_A \mathbb{1}_B d\eta = \langle Z_A, Z_B \rangle_{\mathcal{I}}.$$

Comme les fonctions étagées sont denses dans $L^2(\mathbb{T}, \mathcal{B}(\mathbb{T}), \eta)$, $\overline{\text{Vect}(\mathbb{1}_A, A \in \mathcal{B}(\mathbb{T}))} = L^2(\mathbb{T}, \mathcal{B}(\mathbb{T}), \eta)$, le résultat découle du Théorème A.3.16. \square

Nous noterons, pour $f \in L^2(\mathbb{T}, \mathcal{B}(\mathbb{T}), \eta)$,

$$Z_f \stackrel{\text{def}}{=} \int_{\mathbb{T}} f dZ = \int_{\mathbb{T}} f(\lambda) dZ(\lambda). \quad (3.10)$$

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

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

et

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

On remarque aussi que, comme pour toute variable aléatoire $Y \in \overline{\text{Vect}(Z_A, A \in \mathcal{B}(\mathbb{T}))}$, $\mathbb{E}[Y] = 0$, nous avons $\mathbb{E} \left[\int_{\mathbb{T}} f dZ \right] = 0$.

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

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

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

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

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

3.4.3 Champ spectral associé à un processus et représentation spectrale

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

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

Démonstration. Posons $f_n(\lambda) = e^{-in\lambda}$ pour $n \in \mathbb{N}$. Nous avons $f_n \in L^2(\mathbb{T}, \mathcal{B}(\mathbb{T}), \mu)$. Comme $Z : f_n \mapsto Z_{f_n}$ est une isométrie, nous avons pour tout $(n, m) \in \mathbb{N}^2$,

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

\square

Nous allons maintenant montrer que tout processus stationnaire $\{X_k\}_{k \geq 0}$ est la transformée de Fourier d'un champ spectral décorrélé.

Théorème 3.4.8. *Soit $\{X_k\}_{k \geq 0}$ un processus stationnaire au second-ordre centré de mesure spectrale μ . Alors il existe un unique champ aléatoire décorrélé Z sur $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ de base μ tel que, pour tout $n \in \mathbb{N}$, $X_n = \int_{\mathbb{T}} e^{-in\lambda} dZ(\lambda)$.*

On appellera Z le *champ spectral* de X .

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

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

D'après le théorème A.3.16 il existe une unique isométrie S de $\overline{\text{Vect}(f_n, n \in \mathbb{Z})} = L^2(\mathbb{T}, \mathcal{B}(\mathbb{T}), \mu)$ à valeurs dans $\overline{\text{Vect}(X_n, n \in \mathbb{Z})} = \mathcal{H}^X$, l'enveloppe linéaire du processus $\{X_k\}_{k \geq 0}$. Nous notons, pour $A \in \mathcal{B}(\mathbb{T})$, considérons le champ aléatoire $Z_A = S\mathbb{1}_A \in \mathcal{H}^X$. Comme $\{X_k\}_{k \geq 0}$ est un processus centré,

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

$$\mu(A \cap B) = \langle 1_A, 1_B \rangle_{L^2(\mathbb{T}, \mathcal{B}(\mathbb{T}), \mu)} = \langle Z_A, Z_B \rangle_{L^2(\Omega, \mathcal{A}, \mathbb{P})},$$

ce qui montre, en utilisant le Lemme 3.4.3 que Z est un champ aléatoire décorrélé. Le théorème 3.4.5 montre que $Z : f \mapsto \int f dZ$ définit une isométrie de $L^2(\mathbb{T}, \mathcal{B}(\mathbb{T}), \mu)$ à valeurs dans \mathcal{H}^X . Mais comme, pour tout $A \in \mathcal{B}(\mathbb{T})$, nous avons, $\int \mathbb{1}_A dZ = S\mathbb{1}_A$, nous avons $\int f dZ = Sf$ pour tout $f \in L^2(\mathbb{T}, \mathcal{B}(\mathbb{T}), \mu)$ et par conséquent,

$$X_n = Z_{f_n} = \int_{\mathbb{T}} e^{-in\lambda} dZ(\lambda).$$

□

Chapitre 4

Filtrage des processus stationnaires au second ordre

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

4.1 Filtrages linéaires de processus

On appellera dans la suite l'*opérateur de filtrage linéaire* l'opérateur qui à un processus (X_t) associe le processus (Y_t) défini par

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

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

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

$$Y_{n,t} = \sum_{s=-n}^n \psi_s X_{t-s}$$

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

$$Y_t = \sum_{s=-\infty}^{\infty} \psi_s X_{t-s}.$$

De plus, la variable aléatoire Y_t est intégrable, i.e. $\mathbb{E}[|Y_t|] < \infty$ et la suite $\{Y_{n,t}\}_{n \geq 0}$ converge vers Y_t en norme $\|\cdot\|_1$, i.e.

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

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

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

Démonstration. Voir le paragraphe 4.3 en fin de chapitre. \square

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

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

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

de fonction d'autocovariance :

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

et de mesure spectrale :

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

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

Démonstration. Voir le paragraphe 4.3 à la fin de ce chapitre. \square

La relation (4.4) qui donne la mesure spectrale du processus filtré en fonction de la fonction de transfert du filtre et de la mesure spectrale du processus de départ est particulièrement simple. Elle montre par exemple que la mise en série de deux filtres $\alpha(B)$, $\beta(B)$ de réponses impulsionnelles absolument sommables conduit à une mesure spectrale $|\alpha(e^{-i\lambda})|^2 |\beta(e^{-i\lambda})|^2 \nu_X(d\lambda)$ pour le processus filtré, ce qui montre au passage que l'ordre d'application des filtres est indifférent.

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

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

$$X_t = \mu + \sum_{k=-\infty}^{\infty} \psi_k Z_{t-k} \quad (4.5)$$

où μ désigne une valeur arbitraire.

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

$$\gamma_X(h) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h},$$

et dont la mesure spectrale admet une densité donnée par :

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} |\psi(e^{-i\lambda})|^2, \quad (4.6)$$

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

4.2 Processus ARMA

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

4.2.1 Processus MA(q)

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

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \quad (4.7)$$

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

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

$$\gamma_X(h) = \begin{cases} \sigma^2 \sum_{t=0}^{q-|h|} \theta_t \theta_{t+|h|}, & \text{si } 0 \leq |h| \leq q, \\ 0, & \text{sinon.} \end{cases} \quad (4.8)$$

Enfin, d'après la formule (4.6), le processus admet une densité spectrale dont l'expression est :

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

Un exemple de densité spectrale pour le processus MA(1) est représenté figure 3.6. De manière générale, la densité spectrale d'un processus MA(q) possède des anti-résonnances au voisinage des pulsations correspondant aux arguments des racines du polynôme $\theta(z) = \sum_{k=1}^q \theta_k z^k$.

4.2.2 Processus AR(p)

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

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t, \quad (4.9)$$

où $Z_t \sim \text{BB}(0, \sigma^2)$ est un bruit blanc.

Le terme "autorégressif" provient de la forme de l'équation (4.9) dans laquelle la valeur courante du processus s'exprime sous la forme d'une régression des p valeurs précédentes du processus plus un bruit additif.

L'existence et l'unicité d'une solution stationnaire au second ordre de l'équation (4.9) constituent des questions délicates (qui ne se posaient pas lorsque nous avons défini les modèles MA). Nous détaillons ci-dessous la réponse à cette question dans le cas le cas $p = 1$.

Cas : $|\phi_1| < 1$

L'équation de récurrence s'écrit :

$$X_t = \phi_1 X_{t-1} + Z_t \quad (4.10)$$

Puisque $|\phi_1| < 1$, la fraction rationnelle $\psi(z) = (1 - \phi_1 z)^{-1}$ a un développement en série entière de la forme :

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

qui converge sur le disque $\{z \in \mathbb{C} : |z| < |\phi_1|^{-1}\}$. Considérons alors le filtre linéaire de réponse impulsionnelle $\psi_k = \phi_1^k$ pour $k \geq 0$ et $\psi_k = 0$ sinon. Comme (ψ_k) est absolument sommable, le processus

$$Y_t = \sum_{k=0}^{\infty} \psi_k Z_{t-k} = \sum_{k=0}^{\infty} \phi_1^k Z_{t-k}$$

est bien défini et est stationnaire au second ordre d'après le théorème 4.1.2. Par construction Y_t est solution de (4.10) ce que l'on peut également vérifier directement en notant que :

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

L'unicité de la solution est garantie par l'hypothèse de stationnarité au second ordre. Supposons en effet que $\{X_t\}$ et $\{Y_t\}$ soient deux processus stationnaires au second-ordre et que ces deux processus soient solutions de l'équation de récurrence (4.10). On a alors par différence $(X_t - Y_t) = \phi_1(X_{t-1} - Y_{t-1})$, relation qui itérée k fois implique

$$(X_t - Y_t) = \phi_1^k (X_{t-k} - Y_{t-k}).$$

Par suite, en utilisant l'inégalité triangulaire et l'inégalité de Cauchy-Schwarz,

$$\begin{aligned} \mathbb{E}[|X_t - Y_t|] &= |\phi_1|^k \mathbb{E}[|X_{t-k} - Y_{t-k}|] \\ &\leq |\phi_1|^k (\mathbb{E}[|X_{t-k}|] + \mathbb{E}[|Y_{t-k}|]) \leq |\phi_1|^k (\mathbb{E}[X_0^2]^{\frac{1}{2}} + \mathbb{E}[Y_0^2]^{\frac{1}{2}}), \end{aligned}$$

où k est un entier quelconque. Comme ϕ_1 est en module plus petit que 1, on en déduit que $\mathbb{E}[|X_t - Y_t|] = 0$ et donc que $X_t = Y_t$ presque sûrement. La fonction d'autocovariance de X_t so-

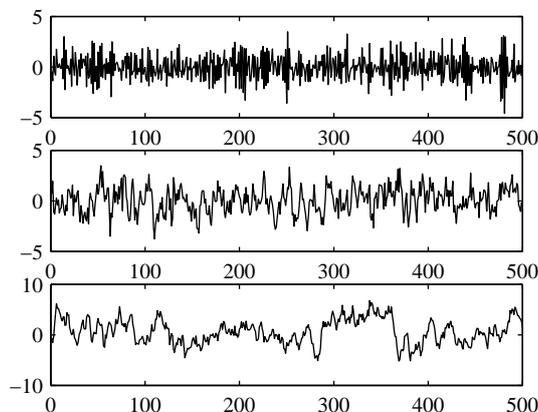


FIGURE 4.1 – Trajectoires de longueur 500 d'un processus AR(1) gaussien. Courbe du haut : $\phi_1 = -0.7$. Courbe du milieu : $\phi_1 = 0.5$. Courbe du bas : $\phi_1 = 0.9$

lution stationnaire de (4.10) est donnée par la formule (4.3) qui s'écrit ;

$$\gamma_X(h) = \sigma^2 \sum_{k=0}^{\infty} \phi_1^k \phi_1^{k+|h|} = \sigma^2 \frac{\phi_1^{|h|}}{1 - \phi_1^2} \quad (4.11)$$

Lorsque $\phi_1 > 0$, le processus X_t est positivement corrélé, dans le sens où tous ses coefficients d'autocovariance sont positifs. Les exemples de trajectoires représentées sur la figure 4.1 montrent que des valeurs de ϕ_1 proches de 1 correspondent à des trajectoires "persistantes" (dont, par exemple, les temps successifs de passage par zéro sont relativement espacés). Inversement, des valeurs de ϕ_1 négatives conduisent à des trajectoires où une valeur positive a tendance à être suivie par une valeur négative. La densité spectrale de (X_t) est donnée par

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{k=0}^{\infty} \phi_1^k e^{-ik\lambda} \right|^2 = \frac{\sigma^2}{2\pi} \frac{1}{|1 - \phi_1 e^{-i\lambda}|^2} . \quad (4.12)$$

La figure 4.2 donne la forme de cette densité spectrale pour $\phi_1 = 0.7$.

Cas $|\phi_1| > 1$

Nous allons montrer que le processus retourné temporellement vérifie une équation récurrente qui nous ramène au cas précédent. Pour cela posons $X_t^r = X_{-t}$. En portant X_t^r dans l'équation (4.10), on obtient

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

qui peut encore s'écrire :

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

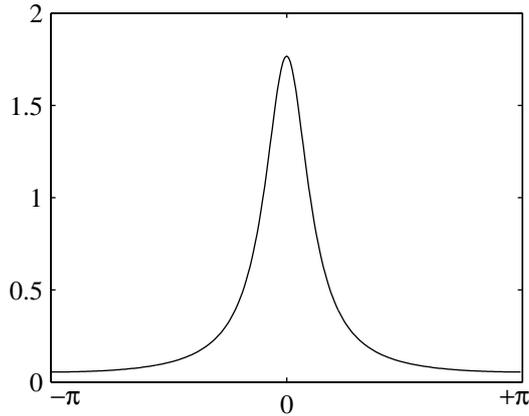


FIGURE 4.2 – Densité spectrale d'un processus $AR(1)$, défini par (4.10) pour $\sigma = 1$ et $\phi_1 = 0.7$.

où $W_t = -\phi_1^{-1}Z_{-t-1}$ est un bruit blanc de variance $\sigma_W^2 = \sigma^2/\phi_1^2$. L'équation (4.13) est maintenant du type que (4.9) puisque $|\phi_1^{-1}| < 1$. Par conséquent il existe un unique processus stationnaire solution de l'équation (4.13) donné par

$$X_t^r = \sum_{k=0}^{\infty} \phi_1^{-k} W_{t-k} . \quad (4.14)$$

Comme $\{X_t^r\}$ est stationnaire au second ordre, le processus

$$X_t = X_{-t}^r = \sum_{k=0}^{\infty} \phi_1^{-k} W_{-t+k} = - \sum_{k=1}^{\infty} \phi_1^{-k} Z_{t+k} \quad (4.15)$$

l'est également (cf. exemple 3.2.4) avec la même moyenne et la même fonction d'autocovariance. Les expressions de la fonction d'autocovariance et de la densité spectrale du processus sont donc données respectivement par (4.11) et (4.12) à condition de substituer ϕ_1 par $1/\phi_1$. Un point remarquable à propos de l'expression de la solution stationnaire donnée par (4.15) est que celle ci est entièrement anti-causale, dans le sens où elle ne dépend que des valeurs futures du bruit Z_t . Cette remarque montre qu'il ne faut pas se laisser tromper par l'apparence de la relation de récurrence (4.13) : la solution stationnaire ne s'exprime par forcément comme un filtrage causal du bruit Z_t , dans le sens où elle ne dépend que des valeurs passées du bruit Z_t . Nous définirons la notion de causalité dans le paragraphe 4.2.3.

Cas $|\phi_1| = 1$

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

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

Or, le terme de gauche est aussi égal à

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

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

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

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

Une remarque intéressante est que dans le cas où $\phi_1 = 1$, le processus $Z_t = X_t - X_{t-1}$ est par hypothèse stationnaire. On peut donc utiliser le modèle $X_t - X_{t-1} = Z_t$ pour un processus $\{X_t, t \in \mathbb{Z}\}$ non-stationnaire dont les incréments sont supposés stationnaires. C'est implicitement la stratégie que nous avons adoptée pour analyser la série de l'indice S&P500 représentée figure 1.4 au paragraphe 3.2.2 (en utilisant en plus une transformation logarithmique des données).

4.2.3 Cas général

Le théorème 4.2.3 donne une condition nécessaire et suffisante d'existence d'une solution stationnaire à l'équation récurrente (4.16) définissant un processus ARMA(p, q).

Théorème 4.2.3 (Existence des processus ARMA(p, q)). *Soit l'équation récurrente :*

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \quad (4.16)$$

où $Z_t \sim \text{BB}(0, \sigma^2)$. On pose $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ et $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$. On suppose que $\phi(z)$ et $\theta(z)$ n'ont pas de zéros communs. Alors l'équation (4.16) admet une solution stationnaire au second ordre si et seulement si le polynôme $\phi(z) \neq 0$ pour $|z| = 1$. Cette solution est unique et a pour expression :

$$X_t = \sum_{k=-\infty}^{\infty} \psi_k Z_{t-k}, \quad (4.17)$$

où ψ_k est la suite des coefficients du développement en série de Laurent de $\theta(z)/\phi(z)$ au voisinage du cercle unité.

Démonstration. Comme $\phi(z) \neq 0$ pour $|z| = 1$, $1/\phi(z)$ est développable en série de Laurent au voisinage du cercle unité, suivant :

$$\xi(z) = \frac{1}{\phi(z)} = \sum_{k=-\infty}^{\infty} \xi_k z^k$$

où la suite $\{\xi_k\}$ est de module sommable et vérifie $\xi_0 = 1$. D'après le théorème 4.1.2, nous pouvons donc appliquer le filtre de réponse impulsionnelle $\{\xi_k\}$ aux deux membres de l'équation récurrente $\phi(B)X_t = \theta(B)Z_t$. Nous obtenons $(\xi(B)\phi(B))X_t = X_t = \psi(B)Z_t$ où $\psi(B) = \xi(B)\theta(B)$. On en déduit que $\psi(z) = \sum_k \psi_k z^k$ avec :

$$\psi_k = \xi_k + \sum_{j=1}^q \theta_j \xi_{k-j}$$

où $\{\psi_k\}$ est absolument sommable. □

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

- Les zéros communs ne sont pas sur le cercle unité. Dans ce cas on se ramène au cas sans zéro commun en annulant les facteurs communs.
- Certains des zéros communs se trouvent sur le cercle unité. L'équation (4.16) admet une infinité de solutions stationnaires au second ordre.

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

ARMA(p, q) causal

Comme dans le cas d'un processus AR(p), on peut distinguer trois cas, suivant que les zéros de $\phi(z)$ sont à l'extérieur, à l'intérieur ou de part et d'autre du cercle unité. Dans le cas où les zéros de $\phi(z)$ sont à l'extérieur du cercle unité, la suite ξ_k est causale ($\xi_k = 0$ pour $k < 0$) et donc $\psi_k = \xi_k + \sum_{j=1}^q \theta_j \xi_{k-j}$ est aussi causale. Par conséquent le processus X_t s'exprime causalement en fonction de Z_t .

Théorème 4.2.4 (ARMA(p, q) causal).

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \quad (4.18)$$

où $Z_t \sim \text{BB}(0, \sigma^2)$. On pose $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ et $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$. On suppose que $\phi(z)$ et $\theta(z)$ n'ont pas de zéros communs. Alors l'équation (4.18) admet une solution stationnaire causale au second ordre si et seulement si le polynôme $\phi(z) \neq 0$ pour $|z| \leq 1$. Cette solution est unique et a pour expression :

$$X_t = \sum_{k=0}^{\infty} \psi_k Z_{t-k} \quad (4.19)$$

où ψ_k est la suite des coefficients du développement en série de Laurent de $\theta(z)/\phi(z)$ dans le disque $\{z : |z| \leq 1\}$.

Démonstration. Il suffit de remarquer que la condition sur $\phi(z)$ implique que $1/\phi(z)$ possède un développement causal au voisinage du cercle unité. $\xi(B)$ correspond donc à une opération de filtrage causal (voir preuve du théorème 4.2.3 pour les notations), ce qui implique qu'il en va de même pour $\xi(B)\phi(B)$. □

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

Une première méthode consiste à utiliser l'expression (4.3) qui s'écrit, compte tenu du fait que $\{Z_t\}$ est un bruit blanc,

$$\gamma(h) = \sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{k+|h|}$$

où la suite $\{\psi_k\}$ se détermine de façon récurrente à partir de l'égalité $\psi(z)\theta(z) = \phi(z)$ par identification du terme en z^k . Pour les premiers termes on trouve :

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

La seconde méthode utilise une formule de récurrence, vérifiée par la fonction d'autocovariance d'un processus ARMA(p, q), qui s'obtient en multipliant les deux membres de (4.16) par X_{t-k} et en prenant l'espérance. On obtient :

$$\gamma(k) - \phi_1\gamma(k-1) - \dots - \phi_p\gamma(k-p) = \sigma^2 \sum_{k \leq j \leq q} \theta_j \psi_{j-k} \quad \text{pour } 0 \leq k < \max(p, q+1) \quad (4.20)$$

$$\gamma(k) - \phi_1\gamma(k-1) - \dots - \phi_p\gamma(k-p) = 0 \quad \text{pour } k \geq \max(p, q+1) \quad (4.21)$$

où nous avons utilisé la causalité du processus pour écrire que $\mathbb{E}[Z_t X_{t-k}] = 0$ pour tout $k \geq 1$. Le calcul de la suite $\{\psi_k\}$ pour $k = 1, \dots, p$ se fait comme précédemment. En reportant ces valeurs dans (4.20) pour $0 \leq k \leq p$, on obtient $(p+1)$ équations linéaires aux $(p+1)$ inconnues $(\gamma(0), \dots, \gamma(p))$ que l'on peut résoudre. Pour déterminer les valeurs suivantes on utilise l'expression (4.21).

Inversibilité d'un processus ARMA(p, q)

Théorème 4.2.5 (ARMA(p, q) inversible). *Soit X_t un processus ARMA(p, q). On suppose que $\phi(z)$ et $\theta(z)$ n'ont pas de zéros communs. Alors il existe une suite $\{\pi_k\}$ causale absolument sommable telle que :*

$$Z_t = \sum_{k=0}^{\infty} \pi_k X_{t-k} \quad (4.22)$$

si et seulement si $\theta(z) \neq 0$ pour $z \leq 1$. On dit alors que le modèle ARMA(p, q) est inversible. La suite π_k est la suite des coefficients du développement en série de $\phi(z)/\theta(z)$ dans le disque $\{z : |z| \leq 1\}$.

La preuve de ce théorème est tout à fait analogue à celle du théorème 4.2.4. Remarquons que la notion d'inversibilité, comme celle de causalité, est bien relative au modèle ARMA(p, q) lui-même et pas uniquement au processus X_t . Un modèle ARMA(p, q) est causal et inversible lorsque les racines des polynômes $\phi(z)$ et $\theta(z)$ sont toutes situées à l'extérieur du filtre unité. Dans ce cas, X_t et Z_t se déduisent mutuellement l'un de l'autre par des opérations de filtrage causal, la réponse impulsionnelle de chacun de ces filtres étant à *phase minimale* (c'est à dire inversible causalement).

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

Théorème 4.2.6 (Densité spectrale d'un processus ARMA(p, q)). *Soit X_t un processus ARMA(p, q) (pas nécessairement causal ou inversible) défini par $\phi(B)X_t = \theta(B)Z_t$ où $Z_t \sim BB(0, \sigma^2)$ et où $\theta(z)$ et $\phi(z)$ sont des polynômes de degré q et p n'ayant pas de zéros communs. Alors X_t possède une densité spectrale qui a pour expression :*

$$f(\lambda) = \frac{\sigma^2 \left| 1 + \sum_{k=1}^q \theta_k e^{-ik\lambda} \right|^2}{2\pi \left| 1 - \sum_{k=1}^p \phi_k e^{-ik\lambda} \right|^2} \quad (4.23)$$

4.3 Preuves des théorèmes 4.1.1 et 4.1.2

Preuve du théorème 4.1.1. Notons pour tout $t \in \mathbb{Z}$ et $n \in \mathbb{N}$, $|Y|_{n,t} = \sum_{s=-n}^{+n} |\psi_s| |X_{t-s}|$. La suite $\{|Y|_{n,t}\}_{n \geq 0}$ est une suite de variables aléatoires intégrables. Le théorème de convergence dominé montre que

$$\lim_{n \rightarrow \infty} \mathbb{E}[|Y|_{n,t}] = \mathbb{E}[|Y|_t]$$

où $|Y|_t = \sum_{s=-\infty}^{\infty} |\psi_s| |X_{t-s}|$. Comme,

$$\mathbb{E}[|Y|_{n,t}] = \sum_{s=-n}^{+n} |\psi_s| \mathbb{E}[|X_{t-s}|] \leq \sup_{t \in \mathbb{Z}} \mathbb{E}[|X_t|] \sum_{s=-\infty}^{\infty} |\psi_s|,$$

on a donc

$$\mathbb{E} \left[\sum_{s=-\infty}^{\infty} |\psi_s| |X_{t-s}| \right] < \infty.$$

Par conséquent, il existe un ensemble $A \in \mathcal{F}$, vérifiant $\mathbb{P}A = 1$ tel que, pour tout $\omega \in A$, nous ayons

$$\sum_{s=-\infty}^{\infty} |\psi_s| |X_{t-s}(\omega)| < \infty$$

Pour $\omega \in A$, la série de terme générique $s \mapsto \psi_s X_{t-s}(\omega)$ est normalement sommable, ce qui implique que, pour tout $\omega \in A$, la suite $n \mapsto Y_{n,t}(\omega)$ converge.

Notons, pour tout $\omega \in \Omega$, $Y_t(\omega) = \limsup Y_{n,t}(\omega)$. $\omega \mapsto Y_t(\omega)$ est une variable aléatoire comme limite supérieure de variables aléatoires et pour tout $\omega \in A$, nous avons $\lim_{n \rightarrow \infty} Y_{n,t}(\omega) = Y_t(\omega)$ et donc la suite $n \mapsto Y_{n,t}$ converge \mathbb{P} -p.s vers Y_t .

Remarquons également que la suite $n \mapsto Y_{n,t}$ est une suite de Cauchy dans $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. En effet, pour tout $p \geq q$, nous avons :

$$\mathbb{E}[|Y_{p,t} - Y_{q,t}|] \leq \sup_{t \in \mathbb{Z}} \mathbb{E}[|X_t|] \sum_{s=q+1}^p |\psi_s| \xrightarrow{q,p \rightarrow \infty} 0$$

Fixons $\epsilon > 0$ et choisissons n tel que

$$\sup_{p,q \geq n} \mathbb{E}[|Y_{p,t} - Y_{q,t}|] \leq \epsilon$$

Par application du lemme de Fatou nous avons alors, pour tout $q \geq n$,

$$\mathbb{E} \left[\liminf_{p \rightarrow \infty} |Y_{p,t} - Y_{q,t}| \right] = \mathbb{E}[|Y_t - Y_{q,t}|] \leq \liminf_{p \rightarrow \infty} \mathbb{E}[|Y_{p,t} - Y_{q,t}|] \leq \epsilon$$

et donc $\limsup_{q \rightarrow \infty} \mathbb{E}[|Y_{q,t} - Y_t|] \leq \epsilon$. Comme ϵ est arbitraire, nous avons donc $\lim_{q \rightarrow \infty} \mathbb{E}[|Y_{q,t} - Y_t|] = 0$. L'inégalité triangulaire

$$\mathbb{E}[|Y_t|] \leq \mathbb{E}[|Y_t - Y_{n,t}|] + \mathbb{E}[|Y_{n,t}|]$$

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

$$\begin{aligned} \mathbb{E} [(Y_{p,t} - Y_{q,t})^2] &= \mathbb{E} \left[\sum_{s=q+1}^p \psi_s X_{t-s} \right]^2 = \sum_{j,k=q+1}^p \psi_j \psi_k \mathbb{E} [X_{t-j} X_{t-k}] \\ &\leq \sum_{j,k=q+1}^p |\psi_j| |\psi_k| \sup_{t \in \mathbb{Z}} \mathbb{E} [X_t^2] = \sup_{t \in \mathbb{Z}} \mathbb{E} [X_t^2] \left(\sum_{j=q+1}^p |\psi_j| \right)^2 \end{aligned}$$

Comme précédemment fixons $\epsilon > 0$ et choisissons n tel que :

$$\sup_{p,q \geq n} \mathbb{E} [|Y_{p,t} - Y_{q,t}|^2] \leq \epsilon.$$

Par application du lemme de Fatou, nous avons :

$$\mathbb{E} \left[\liminf_{p \rightarrow \infty} (Y_{p,t} - Y_{q,t})^2 \right] = \mathbb{E} [(Y_t - Y_{q,t})^2] \leq \liminf_{p \rightarrow \infty} \mathbb{E} [(Y_{p,t} - Y_{q,t})^2] \leq \epsilon$$

et donc : $\limsup_{q \rightarrow \infty} \mathbb{E} [(Y_t - Y_{q,t})^2] \leq \epsilon$. Comme ϵ est arbitraire, $\limsup_{q \rightarrow \infty} \mathbb{E} [(Y_t - Y_{q,t})^2] = 0$, en d'autres termes, la suite $\{Y_{q,t}\}_{q \geq 0}$ converge en moyenne quadratique vers Y_t . Finalement, nous avons :

$$\mathbb{E} [Y_t^2] \leq 2(\mathbb{E} [(Y_t - Y_{q,t})^2] + \mathbb{E} [Y_{q,t}^2]) < \infty$$

et Y_t est donc une variable de carré intégrable. □

Preuve du théorème 4.1.2. Comme $\mathbb{E} [\sum_{s=-\infty}^{\infty} |\psi_s| \mathbb{E} [|X_{t-s}|]] < \infty$, le théorème de Fubini implique

$$\mathbb{E} \left[\sum_{s=-\infty}^{\infty} \psi_s X_{t-s} \right] = \sum_{s=-\infty}^{\infty} \psi_s \mathbb{E} [X_{t-s}]$$

ce qui établit (4.2). Pour la fonction d'autocovariance, notons tout d'abord que, pour tout n , le processus $Y_{n,t} = \sum_{s=-n}^n \psi_s X_{t-s}$ est stationnaire au second ordre et que nous avons

$$\text{cov}(Y_{n,t}, Y_{n,t+h}) = \sum_{j=-n}^n \sum_{k=-n}^n \psi_j \psi_k \gamma_X(h+k-j)$$

Remarquons ensuite que

$$\begin{aligned} \text{cov}(Y_t, Y_{t+h}) &= \text{cov}(Y_{n,t} + (Y_t - Y_{n,t}), Y_{n,t+h} + (Y_{t+h} - Y_{n,t+h})) \\ &= \text{cov}(Y_{n,t}, Y_{n,t+h}) + \text{cov}(Y_t - Y_{n,t}, Y_{n,t+h}) \\ &\quad + \text{cov}(Y_{n,t}, Y_{t+h} - Y_{n,t+h}) + \text{cov}(Y_t - Y_{n,t}, Y_{t+h} - Y_{n,t+h}) \\ &= A + B + C + D \end{aligned}$$

L'inégalité :

$$\text{var}(Y_{n,t} - Y_t) = \lim_{p \rightarrow \infty} \text{var}(Y_{n,t} - Y_{p,t}) \leq \left(\sum_{j=n+1}^{\infty} |\psi_j| \right)^2 \gamma_X(0)$$

permet ensuite de déduire, quand n tend vers l'infini, les limites suivantes

$$\begin{aligned} |B| &\leq (\text{var}(Y_t - Y_{n,t}))^{1/2} (\text{var}(Y_{n,t+h}))^{1/2} \rightarrow 0 \\ |C| &\leq (\text{var}(Y_{t+h} - Y_{n,t+h}))^{1/2} (\text{var}(Y_{n,t}))^{1/2} \rightarrow 0 \\ |D| &\leq (\text{var}(Y_{t+h} - Y_{n,t+h}))^{1/2} (\text{var}(Y_t - Y_{n,t}))^{1/2} \rightarrow 0 \end{aligned}$$

et donc $\text{cov}(Y_t, Y_{t+h}) = \lim_{n \rightarrow \infty} \text{cov}(Y_{n,t}, Y_{n,t+h})$, ce qui démontre l'expression (4.3)¹. En reportant dans cette expression $\gamma_X(h) = \int_{\mathbb{T}} e^{ih\lambda} \nu_X(d\lambda)$ où ν_X désigne la mesure spectrale du processus $\{X_t\}$, nous obtenons

$$\gamma_Y(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \int_{\mathbb{T}} e^{i(h+k-j)\lambda} \nu_X(d\lambda)$$

En remarquant ensuite que

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

nous pouvons appliquer le théorème de Fubini et permuter les signes somme et intégrale dans l'expression de $\gamma_Y(h)$. Ce qui donne :

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

On en déduit que $\nu_Y(d\lambda) = |\psi(e^{-i\lambda})|^2 \nu_X(d\lambda)$. Pour déterminer l'expression de l'intercovariance entre les processus entre les processus Y_t et X_t , il suffit de noter $|\text{cov}(Y_{t+h}, X_t)|^2 \leq \gamma_Y(0) \gamma_X(0) < +\infty$ et que :

$$\begin{aligned} \mathbb{E}[(Y_{t+h} - \mu_Y)(X_t - \mu_X)] &= \lim_{n \rightarrow \infty} \text{cov}(Y_{n,t+h}, X_t) = \lim_{n \rightarrow \infty} \sum_{k=-n}^n \psi_k \text{cov}(X_{t+h-k}, X_t) \\ &= \sum_{k=-\infty}^{\infty} \psi_k \gamma_X(h-k) \end{aligned}$$

Ce qui conclut la preuve. □

1. Nous venons ici de démontrer directement la propriété de continuité de la covariance dans L^2 que nous verrons comme une conséquence de la structure d'espace de Hilbert au chapitre 5.

Chapitre 5

Prédiction des processus stationnaires au second ordre

5.1 Prédiction linéaire

5.1.1 Estimation linéaire en moyenne quadratique

Soient X et $\{Y_1, \dots, Y_p\}$ des variables aléatoires réelles de $L^2(\Omega, \mathcal{F}, \mathbb{P})$. On cherche à déterminer la meilleure approximation de X par une combinaison linéaire des variables Y_k . Nous supposons ici que nous connaissons les quantités $\mu = \mathbb{E}[X]$, $\nu_k = \mathbb{E}[Y_k]$ ainsi que les coefficients de covariance $\text{cov}(X, Y_k)$ et $\text{cov}(Y_k, Y_\ell)$, pour tout $1 \leq k, \ell \leq p$. En pratique, nous verrons au chapitre 6 comment il est possible, sous certaines hypothèses, de construire des estimateurs consistants et asymptotiquement normaux de ces quantités à partir d'une suite d'observations.

On considère l'espace fermé de dimension finie $\mathcal{Y} = \text{Vect}(1, Y_1, \dots, Y_p)$ et on cherche l'élément $Y \in \mathcal{Y}$ qui minimise la norme de le risque quadratique $\|X - Y\|^2$. Il découle immédiatement du théorème de projection que le prédicteur linéaire optimal est la projection orthogonale $\text{proj}(X|\mathcal{Y})$ de X sur \mathcal{Y} qui vérifie $(X - \text{proj}(X|\mathcal{Y})) \perp \mathcal{Y}$. On en déduit que :

$$\begin{cases} \langle X - \text{proj}(X|\mathcal{Y}), 1 \rangle = 0 \\ \langle X - \text{proj}(X|\mathcal{Y}), Y_k \rangle = 0 \quad \text{pour } k \in \{1, \dots, p\} \end{cases} \quad (5.1)$$

Ce sont ces $(p + 1)$ équations qui vont nous donner la solution cherchée. En effet la condition $\text{proj}(X|\mathcal{Y}) \in \mathcal{Y}$ implique (comme \mathcal{Y} est de dimension finie) que $\text{proj}(X|\mathcal{Y}) = a_0 + \sum_{k=1}^p a_k(Y_k - \nu_k)$. Il suffit donc de calculer a_0, a_1, \dots, a_p . Partant de la première expression de (5.1), on obtient :

$$\left\langle X - a_0 - \sum_{k=1}^p a_k(Y_k - \nu_k), 1 \right\rangle = \langle X, 1 \rangle - a_0 = 0, \quad (5.2)$$

qui donne $a_0 = \mu$. En posant $a_0 = \mu$ dans la seconde expression de (5.1), on a obtenu alors $k \in$

$\{1, \dots, p\}$:

$$\left\langle X - \mu - \sum_{j=1}^p a_j (Y_j - \nu_j), Y_k - \nu_k \right\rangle = \langle X - \mu, Y_k - \nu_k \rangle - \sum_{j=1}^p a_j \langle Y_j - \nu_j, Y_k - \nu_k \rangle = 0, \quad (5.3)$$

qui montrent que $\{a_1, \dots, a_p\}$ sont solution d'un système de p équations linéaires à p inconnues.

Ce système d'équations peut se mettre sous forme plus compacte en utilisant la matrice $\Gamma = [\text{cov}(Y_k, Y_\ell)]_{1 \leq k, \ell \leq p}$ des coefficients de covariance de (Y_1, \dots, Y_p) et le vecteur $\gamma = [\text{cov}(X, Y_1), \dots, \text{cov}(X, Y_p)]^T$ des coefficients de covariance entre X et les composantes Y_k . Avec ces notations, le vecteur $\alpha = [a_1, \dots, a_p]^T$ est solution de l'équation :

$$\Gamma \alpha = \gamma \quad (5.4)$$

Ce système linéaire admet une unique solution si la matrice Γ est inversible. Notons enfin qu'en vertu de l'identité de Pythagore, nous avons :

$$\|X\|^2 = \|\text{proj}(X|\mathcal{Y})\|^2 + \|X - \text{proj}(X|\mathcal{Y})\|^2$$

et donc la norme minimale de l'erreur de prédiction a pour expression :

$$\|X - \text{proj}(X|\mathcal{Y})\|^2 = \|X\|^2 - \|\text{proj}(X|\mathcal{Y})\|^2.$$

Nous allons à présent appliquer ce résultat à la prédiction d'un processus stationnaire au second-ordre à partir de son passé immédiat en prenant $X = X_t$ et $Y_k = X_{t-k}$ avec $k = \{1, \dots, p\}$.

5.1.2 Prédiction linéaire d'un processus stationnaire au second-ordre

Soit $\{X_t, t \in \mathbb{Z}\}$ un processus stationnaire au second-ordre, de moyenne $\mathbb{E}[X_0] = \mu$ et de fonction d'autocovariance $\gamma(h) = \text{cov}(X_h, X_0)$. On cherche à *prédire* la valeur du processus à la date t à partir d'une combinaison linéaire des p derniers échantillons du passé $\{X_{t-1}, \dots, X_{t-p}\}$. Ce problème est bien entendu un cas particulier du problème précédent où nous avons $X = X_t$ et $Y_k = X_{t-k}$, pour $k \in \{1, \dots, p\}$ et où :

$$\mathcal{H}_{t-1,p} = \text{Vect}(1, X_{t-1}, X_{t-2}, \dots, X_{t-p}) \quad (5.5)$$

Formons la matrice de covariance Γ_p du vecteur $[X_{t-1}, \dots, X_{t-p}]$:

$$\Gamma_p = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \vdots \\ \vdots & \ddots & \ddots & \ddots \\ \vdots & & & \gamma(1) \\ \gamma(p-1) & \gamma(p-2) & \cdots & \gamma(1) & \gamma(0) \end{bmatrix} \quad (5.6)$$

Cette matrice est dite de Toeplitz, ses éléments étant égaux le long de ses diagonales. Notons γ_p le vecteur $[\gamma(1), \gamma(2), \dots, \gamma(p)]^T$ le vecteur des coefficients de corrélation. D'après l'équation (5.4), les coefficients $\{\phi_{k,p}\}_{1 \leq k \leq p}$ du prédicteur linéaire optimal défini par :

$$\text{proj}(X_t | \mathcal{H}_{t-1,p}) - \mu = \sum_{k=1}^p \phi_{k,p}(X_{t-k} - \mu) \quad (5.7)$$

sont solutions du système d'équations :

$$\Gamma_p \phi_p = \gamma_p \quad (5.8)$$

D'autre part l'erreur de prédiction minimale a pour expression :

$$\begin{aligned} \sigma_p^2 &= \|X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p})\|^2 = \langle X_t - \mu, X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p}) \rangle \\ &= \gamma(0) - \sum_{k=1}^p \phi_{k,p} \gamma(k) = \gamma(0) - \phi_p^T \gamma_p \end{aligned} \quad (5.9)$$

Les équations (5.8) et (5.9) sont appelées *équations de Yule-Walker*. Notons la propriété importante suivante : pour p fixé, la suite des coefficients $\{\phi_{k,p}\}_{1 \leq k \leq p}$ du prédicteur linéaire optimal et la variance de l'erreur minimale de prédiction *ne dépendent pas de t* . Les équations (5.8) et (5.9) peuvent encore être réécrites à partir des coefficients de corrélation $\rho(h) = \gamma(h)/\gamma(0)$. Il vient :

$$\begin{bmatrix} \rho(0) & \rho(1) & \cdots & \rho(p-1) \\ \rho(1) & \rho(0) & \rho(1) & \vdots \\ \vdots & \ddots & \ddots & \ddots \\ \vdots & & & \rho(1) \\ \rho(p-1) & \rho(p-2) & \cdots & \rho(1) & \rho(0) \end{bmatrix} \begin{bmatrix} \phi_{1,p} \\ \phi_{2,p} \\ \vdots \\ \vdots \\ \phi_{p,p} \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \vdots \\ \rho(p) \end{bmatrix} \quad (5.10)$$

Exemple 5.1.1 (Prédiction avant/arrière). Soit $X_t = Z_t + \theta_1 Z_{t-1}$ où $Z_t \sim \text{BB}(0, \sigma^2)$. On note $\rho(h)$ la fonction d'autocorrélation de X_t .

1. $\rho(0) = (1 + \theta_1^2)$, $\rho(\pm 1) = \theta_1$ et $\rho(h) = 0$ pour $|h| \geq 2$.
2. Déterminons la prédiction de X_3 en fonction de X_2 et X_1 . D'après le théorème de projection $\text{proj}(X_3 | \text{Vect}(X_2, X_1)) = \alpha_1 X_1 + \alpha_2 X_2$ vérifie $\langle X_3 - \alpha_2 X_2 - \alpha_1 X_1, X_j \rangle = 0$ pour $j = 1, 2$. On en déduit que :

$$\begin{bmatrix} 1 + \theta_1^2 & \theta_1 \\ \theta_1 & 1 + \theta_1^2 \end{bmatrix} \begin{bmatrix} \alpha_2 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} \theta_1 \\ 0 \end{bmatrix}$$

3. Déterminons la prédiction de X_3 en fonction de X_4 et X_5 . D'après le théorème de projection $\text{proj}(X_3 | \text{Vect}(X_4, X_5)) = \alpha_4 X_4 + \alpha_5 X_5$ vérifie $\langle X_3 - \alpha_4 X_4 - \alpha_5 X_5, X_j \rangle = 0$ pour $j = 4, 5$. On en déduit que :

$$\begin{bmatrix} 1 + \theta_1^2 & \theta_1 \\ \theta_1 & 1 + \theta_1^2 \end{bmatrix} \begin{bmatrix} \alpha_4 \\ \alpha_5 \end{bmatrix} = \begin{bmatrix} \theta_1 \\ 0 \end{bmatrix}$$

Par conséquent $\alpha_1 = \alpha_5$ et $\alpha_2 = \alpha_4$.

4. Déterminons la prédiction de X_3 en fonction de X_1, X_2, X_4 et X_5 . Pour déterminer $\text{proj}(X_3 | \text{Vect}(X_1, X_2, X_4, X_5)) = \beta_1 X_1 + \beta_2 X_2 + \beta_4 X_4 + \beta_5 X_5$ Il suffit de remarquer que $\text{Vect}(X_1, X_2) \perp \text{Vect}(X_4, X_5)$ et donc :

$$\text{proj}(X_3 | \text{Vect}(X_1, X_2, X_4, X_5)) = \text{proj}(X_3 | \text{Vect}(X_1, X_2)) + \text{proj}(X_3 | \text{Vect}(X_4, X_5))$$

Exemple 5.1.2 (Cas d'un processus AR(m) causal). Soit le processus AR(m) causal solution stationnaire de l'équation récurrente :

$$X_t = \phi_1 X_{t-1} + \dots + \phi_m X_{t-m} + Z_t$$

où $Z_t \sim B(0, \sigma^2)$ et où $\phi(z) = 1 - \sum_{k=1}^m \phi_k z^k \neq 0$ pour $|z| \leq 1$. Comme la solution est causale on a, pour tout $h \geq 1$, $\mathbb{E}[Z_t X_{t-h}] = 0$ et donc $\mathbb{E}[(X_t - \sum_{k=1}^m \phi_k X_{t-k}) X_{t-h}] = 0$ qui signifie que, pour tout $p \geq m$,

1. $(X_t - \sum_{k=1}^m \phi_k X_{t-k}) \perp \mathcal{H}_{t-1,p}$

2. $\sum_{k=1}^m \phi_k X_{t-k} \in \mathcal{H}_{t-1,p}$.

Par conséquent, d'après le théorème de projection, $\sum_{k=1}^m \phi_k X_{t-k} = \text{proj}(X_t | \mathcal{H}_{t-1,p})$ et donc, pour tout $p \geq m$:

$$\phi_{k,p} = \begin{cases} \phi_k & \text{pour } 1 \leq k \leq m \\ 0 & \text{pour } k > m \end{cases} .$$

La projection orthogonale d'un AR(m) causal sur son passé de longueur $p \geq m$ coïncide avec la projection orthogonale sur les m dernières valeurs et les coefficients de prédiction sont précisément les coefficients de l'équation récurrente.

Dans le cas où la matrice de covariance Γ_p , supposée connue, est inversible, le problème de la détermination des coefficients de prédiction ϕ_p et de la variance de l'erreur de prédiction σ_p^2 a une solution unique. Rappelons que, d'après la propriété 3.3.9, si $\gamma(0) > 0$ et si $\lim_{n \rightarrow \infty} \gamma(n) = 0$, alors la matrice Γ_p est inversible à tout ordre.

Il est facile de démontrer que :

$$\text{proj}(X_t | \text{Vect}(1, X_{t-1}, \dots, X_{t-p})) = \mu + \text{proj}(X_t - \mu | \text{Vect}(X_{t-1} - \mu, \dots, X_{t-p} - \mu)) . \quad (5.11)$$

Par conséquent, dans le problème de la prédiction, il n'y a aucune perte de généralité à considérer que le processus est centré. S'il ne l'était pas, il suffirait, d'après l'équation (5.11), d'effectuer le calcul des prédicteurs sur le processus centré $X_t^c = X_t - \mu$ puis d'ajouter μ . Dans la suite, sauf indication contraire, les processus sont supposés centrés.

Les coefficients de prédiction d'un processus stationnaire au second ordre fournissent une décomposition particulière de la matrice de covariance Γ_{p+1} sous la forme d'un produit de matrice triangulaire.

Théorème 5.1.3. Soit $\{X_t\}$ un processus stationnaire au second ordre, centré, de fonction d'autocovariance $\gamma(h)$. On note :

$$A_{p+1} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -\phi_{1,1} & 1 & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & 0 \\ -\phi_{p,p} & -\phi_{p-1,p} & \cdots & -\phi_{1,p} & 1 \end{bmatrix} \quad D_{p+1} = \begin{bmatrix} \sigma_0^2 & 0 & \cdots & 0 \\ 0 & \sigma_1^2 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & & \sigma_p^2 \end{bmatrix}$$

On a alors :

$$\Gamma_{p+1} = A_{p+1}^{-1} D_{p+1} A_{p+1}^{-T} \quad (5.12)$$

Démonstration. Posons $\mathcal{F}_k = \text{Vect}(X_k, \dots, X_1)$ et montrons tout d'abord que, pour $k \neq \ell$, nous avons :

$$\langle X_k - \text{proj}(X_k | \mathcal{F}_{k-1}), X_\ell - \text{proj}(X_\ell | \mathcal{F}_{\ell-1}) \rangle = 0. \quad (5.13)$$

En effet, pour $k < \ell$, on a $X_k - \text{proj}(X_k | \mathcal{F}_{k-1}) \in \mathcal{F}_k \subseteq \mathcal{F}_{\ell-1}$. On a aussi $X_\ell - \text{proj}(X_\ell | \mathcal{F}_{\ell-1}) \perp \mathcal{F}_{\ell-1}$ et donc $X_\ell - \text{proj}(X_\ell | \mathcal{F}_{\ell-1}) \perp X_k - \text{proj}(X_k | \mathcal{F}_{k-1})$, ce qui démontre (5.13). D'autre part, par définition des coefficients de prédiction, on peut écrire successivement :

$$A_{p+1} \mathbf{X}_{p+1} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -\phi_{1,1} & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ -\phi_{p,p} & -\phi_{p-1,p} & \cdots & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{p+1} \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 - \text{proj}(X_2 | \mathcal{F}_1) \\ \vdots \\ X_{p+1} - \text{proj}(X_{p+1} | \mathcal{F}_p) \end{bmatrix}$$

qui donne :

$$\mathbb{E} [A_{p+1} \mathbf{X}_{p+1} \mathbf{X}_{p+1}^T A_{p+1}^T] = A_{p+1} \Gamma_{p+1} A_{p+1}^T = D_{p+1}$$

où, par définition, $\sigma_k^2 = \|X_k - \text{proj}(X_k | \mathcal{F}_{k-1})\|^2$, ce qui démontre (5.12) puisque la matrice A_{p+1} est inversible, son déterminant étant égal à 1. Ajoutons que l'inverse d'une matrice triangulaire supérieure est elle-même triangulaire supérieure. \square

Dans la suite nous notons $\mathcal{H}_{t-1,p} = \text{Vect}(X_{t-1}, \dots, X_{t-p})$ et nous appelons *erreur de prédiction directe* d'ordre p ou *innovation partielle* d'ordre p le processus :

$$\epsilon_{t,p}^+ = X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p}) = X_t - \sum_{k=1}^p \phi_{k,p} X_{t-k} \quad (5.14)$$

D'après l'équation (5.12) lorsque la matrice Γ_{p+1} est inversible, la variance $\sigma_p^2 = \|\epsilon_{t,p}^+\|^2$ est strictement positive. Il est clair, d'autre part, que la suite σ_p^2 est décroissante et donc que σ_p^2 possède une limite quand p tend vers l'infini. Cela conduit à la définition suivante, dont nous verrons paragraphe 5.5 qu'elle joue un rôle fondamental dans la décomposition des processus stationnaires au second ordre.

Définition 5.1.4 (Processus régulier/déterministe). Soit $\{X_t, t \in \mathbb{Z}\}$ un processus aléatoire stationnaire au second ordre. On note σ_p^2 la variance de l'innovation partielle d'ordre p et $\sigma^2 = \lim_{p \rightarrow +\infty} \sigma_p^2$. On dit que le processus $\{X_t, t \in \mathbb{Z}\}$ est régulier si $\sigma^2 > 0$ et déterministe si $\sigma^2 = 0$.

Nous avons déjà noté (voir équation (5.8)) que, pour p fixé, la suite $\{\phi_{k,p}\}$ ne dépend pas de t et donc que le processus $\epsilon_{t,p}^+$ (relativement à l'indice t) est stationnaire au second ordre, centré. On a aussi la formule suivante :

$$\langle \epsilon_{t,p}^+, \epsilon_{t,q}^+ \rangle = \sigma_{\max(p,q)}^2. \quad (5.15)$$

En effet soit $q > p$. Par construction, nous avons $\epsilon_{t,q}^+ \perp \mathcal{H}_{t-1,q}$, et comme $\mathcal{H}_{t-1,p} \subseteq \mathcal{H}_{t-1,q}$, $\epsilon_{t,q}^+ \perp \mathcal{H}_{t-1,p}$ et en particulier $\epsilon_{t,q}^+ \perp \text{proj}(X_t | \mathcal{H}_{t-1,p})$ puisque $\text{proj}(X_t | \mathcal{H}_{t-1,p}) \in \mathcal{H}_{t-1,p}$. Par conséquent, pour $q > p$, on a :

$$\begin{aligned} (\epsilon_{t,p}^+, \epsilon_{t,q}^+) &= \langle X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p}), \epsilon_{t,q}^+ \rangle \\ &= \langle X_t, X_t - \text{proj}(X_t | \mathcal{H}_{t-1,q}) \rangle = \langle X_t, X_t - \text{proj}(X_t | \mathcal{H}_{t-1,q}) \rangle = \sigma_q^2, \end{aligned}$$

ce qui démontre (5.15).

Notons ici que le problème de la recherche des coefficients de prédiction pour un processus stationnaire au second ordre se ramène à celui de la minimisation de l'intégrale :

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |\psi(e^{-i\lambda})|^2 \nu_X(d\lambda)$$

sur l'ensemble \mathcal{P}_p des polynômes à coefficients réels de degré p de la forme $\psi(z) = 1 + \psi_1 z + \dots + \psi_p z^p$. En effet, en utilisant la relation (4.4) de filtrage des mesures spectrales, on peut écrire que la variance de $\|\epsilon_{t,p}^+\|^2$, qui minimise de l'erreur de prédiction, a pour expression :

$$\sigma_p^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\phi_p(e^{-i\lambda})|^2 \nu_X(d\lambda) \quad (5.16)$$

où :

$$\phi_p(z) = 1 - \sum_{k=1}^p \phi_{k,p} z^k$$

désigne le *polynôme prédicteur d'ordre p* .

Théorème 5.1.5. *Si $\{X_t\}$ est un processus régulier, alors, pour tout p , $\phi_p(z) \neq 0$ pour $|z| \leq 1$. Tous les zéros des polynômes prédicteurs sont à l'extérieur du cercle unité.*

Démonstration. Elle est donnée en fin de chapitre. □

Une conséquence directe du théorème 5.1.5 est qu'à toute matrice de covariance de type défini positif, de dimension $(p+1) \times (p+1)$, on peut associer un processus AR(p) causal dont les $(p+1)$ premiers coefficients de covariance sont précisément la première ligne de cette matrice. Ce résultat n'est pas général. Ainsi il existe bien un processus AR(2) causal ayant $\gamma(0) = 1$ et $\gamma(1) = \rho$, comme premiers coefficients de covariance, à condition toutefois que la matrice de covariance soit positive c'est-à-dire que $|\rho| < 1$, tandis qu'il n'existe pas, pour cette même matrice de processus MA(2). Il faut en effet, en plus du caractère positif, que $|\rho| \geq 1/2$ (voir exemple 3.2.10).

5.2 Algorithme de Levinson-Durbin

La solution directe du système des équations de Yule-Walker requiert de l'ordre de p^3 opérations : la résolution classique de ce système implique en effet la décomposition de la matrice Γ_p sous la forme du produit d'une matrice triangulaire inférieure et de sa transposée, $\Gamma_p = L_p L_p^T$ (décomposition de Choleski) et la résolution par substitution de deux systèmes triangulaires. Cette procédure peut s'avérer coûteuse lorsque l'ordre de prédiction est grand (on utilise généralement des ordres de prédiction de l'ordre de quelques dizaines à quelques centaines), ou lorsque, à des fins de modélisation, on est amené à évaluer la qualité de prédiction pour différents horizons de prédiction. L'algorithme de Levinson-Durbin exploite la structure géométrique particulière des processus stationnaires au second ordre pour établir une formule de récurrence donnant les coefficients de prédiction à l'ordre $(p+1)$ à partir des coefficients de prédiction obtenus à l'ordre p . Supposons que nous connaissions les coefficients de prédiction linéaire et la variance de l'erreur de prédiction à l'ordre p , pour $p \geq 0$:

$$\text{proj}(X_t | \mathcal{H}_{t-1,p}) = \sum_{k=1}^p \phi_{k,p} X_{t-k} \quad \text{et} \quad \sigma_p^2 = \|X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p})\|^2 \quad (5.17)$$

Nous avons besoin ici d'introduire l'erreur de prédiction rétrograde à l'ordre p définie par :

$$\epsilon_{t,p}^- = X_t - \text{proj}(X_t | \mathcal{H}_{t+p,p}) = X_t - \text{proj}(X_t | \text{Vect}(X_{t+1}, \dots, X_{t+p}))$$

Elle représente la différence entre l'échantillon courant X_t et la projection orthogonale de X_t sur les p échantillons $\{X_{t+1}, \dots, X_{t+p}\}$ qui suivent l'instant courant. Le qualificatif *rétrograde* est clair : il traduit le fait que l'on cherche à prédire la valeur courante en fonction des valeurs futures. Indiquons que l'erreur rétrograde joue un rôle absolument essentiel dans tous les algorithmes rapides de résolution des équations de Yule-Walker. Remarquer tout d'abord que les coefficients de prédiction rétrograde coïncident avec les coefficients de prédiction directe. Cette propriété, que nous avons rencontrée exemple 5.1.1, est fondamentalement due à la *propriété de réversibilité* des processus stationnaires au second ordre. En effet, si $Y_t = X_{-t}$, alors Y_t a même moyenne et même fonction de covariance que X_t (voir exemple 3.2.4 chapitre 1) et par conséquent, en utilisant aussi l'hypothèse de stationnarité, on a simultanément pour tout $u, v \in \mathbb{Z}$:

$$\text{proj}(X_{t+u} | \mathcal{H}_{t+u-1,p}) = \sum_{k=1}^p \phi_{k,p} X_{t+u-k} \quad \text{et} \quad \text{proj}(X_{t+v} | \mathcal{H}_{t+v+p,p}) = \sum_{k=1}^p \phi_{k,p} X_{t+v+k}$$

ainsi que :

$$\sigma_p^2 = \|\epsilon_{t+u,p}^+\|^2 = \|\epsilon_{t+v,p}^-\|^2 \quad (5.18)$$

En particulier on a :

$$\begin{cases} \text{proj}(X_t | \mathcal{H}_{t-1,p}) = \sum_{k=1}^p \phi_{k,p} X_{t-k} \\ \text{proj}(X_{t-p-1} | \mathcal{H}_{t-1,p}) = \sum_{k=1}^p \phi_{k,p} X_{t-p-1+k} = \sum_{k=1}^p \phi_{p+1-k,p} X_{t-p-1+k} \end{cases} \quad (5.19)$$

Cherchons maintenant à déterminer, à partir de ces projections à l'ordre p , la projection de X_t à l'ordre $p+1$ sur le sous-espace $\mathcal{H}_{t-1,p+1} = \text{Vect}(X_{t-1}, \dots, X_{t-p-1})$. Pour cela décomposons cet espace en somme directe de la façon suivante :

$$\mathcal{H}_{t-1,p+1} = \mathcal{H}_{t-1,p} \oplus \text{Vect}(X_{t-p-1} - \text{proj}(X_{t-p-1} | \mathcal{H}_{t-1,p})) = \mathcal{H}_{t-1,p} \oplus \text{Vect}(\epsilon_{t-p-1,p}^-)$$

Un calcul simple montre (voir exemple A.2.3) que

$$\text{proj} \left(X_t | \epsilon_{t-p-1,p}^- \right) = \alpha \epsilon_{t-p-1,p}^- \quad \text{avec} \quad \alpha = (X_t, \epsilon_{t-p-1,p}^-) / \|\epsilon_{t-p-1,p}^-\|^2$$

et donc que

$$\text{proj} (X_t | \mathcal{H}_{t-1,p+1}) = \text{proj} (X_t | \mathcal{H}_{t-1,p}) + k_{p+1} [X_{t-p-1} - \text{proj} (X_{t-p-1} | \mathcal{H}_{t-1,p})], \quad (5.20)$$

où, en utilisant aussi (5.18), on peut écrire :

$$k_{p+1} = \frac{\langle X_t, \epsilon_{t-p-1,p}^- \rangle}{\sigma_p^2} = \frac{\langle X_t, \epsilon_{t-p-1,p}^- \rangle}{\|\epsilon_{t+u,p}^+ \| \|\epsilon_{t+v,p}^-\|}. \quad (5.21)$$

En portant à présent (5.19) dans (5.20), on obtient l'expression :

$$\text{proj} (X_t | \mathcal{H}_{t-1,p+1}) = \sum_{k=1}^{p+1} \phi_{k,p+1} X_{t-k} = \sum_{k=1}^p (\phi_{k,p} - k_{p+1} \phi_{p+1-k,p}) X_{t-k} + k_{p+1} X_{t-p-1}$$

On en déduit les formules de récurrence donnant les coefficients de prédiction à l'ordre $p+1$ à partir de ceux à l'ordre p :

$$\begin{cases} \phi_{k,p+1} = \phi_{k,p} - k_{p+1} \phi_{p+1-k,p} & \text{pour } k \in \{1, \dots, p\} \\ \phi_{p+1,p+1} = k_{p+1} \end{cases} \quad (5.22)$$

Déterminons maintenant la formule de récurrence donnant k_{p+1} . En utilisant encore (5.19) et (5.20), on obtient :

$$\langle X_t, \text{proj} (X_{t-p-1} | \mathcal{H}_{t-1,p}) \rangle = \sum_{k=1}^p \phi_{k,p} \mathbb{E} [X_t X_{t-p-1+k}] = \sum_{k=1}^p \phi_{k,p} \gamma(p+1-k)$$

Partant de l'expression de $\langle X_t, \epsilon_{t-p-1,p}^- \rangle$ on en déduit que :

$$\begin{aligned} \langle X_t, \epsilon_{t-p-1,p}^- \rangle &= \langle X_t, X_{t-p-1} - \text{proj} (X_{t-p-1} | \mathcal{H}_{t-1,p}) \rangle \\ &= \gamma(p+1) - \sum_{k=1}^p \phi_{k,p} \gamma(p+1-k) \end{aligned}$$

et donc d'après (5.21) :

$$k_{p+1} = \frac{\gamma(p+1) - \sum_{k=1}^p \phi_{k,p} \gamma(p+1-k)}{\sigma_p^2}$$

Il nous reste maintenant à déterminer l'erreur de prédiction σ_{p+1}^2 à l'ordre $(p+1)$. En utilisant l'équation (5.20), on a

$$\begin{aligned} \epsilon_{t,p+1}^+ &= X_t - \text{proj} (X_t | \mathcal{H}_{t-1,p+1}) \\ &= X_t - \text{proj} (X_t | \mathcal{H}_{t-1,p}) - k_{p+1} (X_{t-p-1} - \text{proj} (X_{t-p-1} | \mathcal{H}_{t-1,p})) \end{aligned}$$

dont on déduit d'après (5.21) :

$$\begin{aligned}\sigma_{p+1}^2 &= \|\epsilon_{t,p+1}^+\|^2 = \sigma_p^2 + k_{p+1}^2 \sigma_p^2 - 2k_{p+1} \langle X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p}), X_{t-p-1} - \text{proj}(X_{t-p-1} | \mathcal{H}_{t-1,p}) \rangle \\ &= \sigma_p^2 (1 - k_{p+1}^2)\end{aligned}$$

Pour initialiser l'algorithme, nous faisons $p = 0$. Dans ce cas la meilleure prédiction de X_t est $\mathbb{E}[X_t] = 0$ et la variance de l'erreur de prédiction est alors donnée par $\sigma_0^2 = \mathbb{E}[(X_t - 0)^2] = \gamma(0)$. Au pas suivant on a $k_1 = \gamma(1)/\gamma(0)$, $\phi_{1,1} = \gamma(1)/\gamma(0)$ et $\sigma_1^2 = \gamma(0)(1 - k_1^2)$.

Partant d'une suite de $(K + 1)$ coefficients de covariance $\gamma(0), \dots, \gamma(K)$, l'*algorithme de Levinson-Durbin* permet de déterminer les coefficients de prédiction $\{\phi_{m,p}\}_{1 \leq m \leq p, 1 \leq p \leq K}$:

Initialisation $k_1 = \gamma(1)/\gamma(0)$, $\phi_{1,1} = \gamma(1)/\gamma(0)$ et $\sigma_1^2 = \gamma(0)(1 - k_1^2)$

Récursion Pour $p = \{2, \dots, K\}$ répéter :

— Calculer

$$\begin{aligned}k_p &= \sigma_{p-1}^{-2} \left(\gamma(p) - \sum_{k=1}^{p-1} \phi_{k,p-1} \gamma(p-k) \right) \\ \phi_{p,p} &= k_p \\ \sigma_p^2 &= \sigma_{p-1}^2 (1 - k_p^2)\end{aligned}$$

— Pour $m \in \{1, \dots, p-1\}$ calculer :

$$\phi_{m,p} = \phi_{m,p-1} - k_p \phi_{p-m,p-1}$$

Le coefficient k_p possède la propriété remarquable d'être de module inférieur à 1. Notons tout d'abord que $\text{proj}(X_t | \mathcal{H}_{t-1,p}) \perp \epsilon_{t-p-1,p}^-$ puisque $\text{proj}(X_t | \mathcal{H}_{t-1,p}) \in \mathcal{H}_{t-1,p}$ et que $\epsilon_{t-p-1,p}^- \perp \mathcal{H}_{t-1,p}$. Partant de (5.21) on peut écrire que :

$$k_{p+1} = \frac{\langle X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p}), X_{t-p-1} - \text{proj}(X_{t-p-1} | \mathcal{H}_{t-1,p}) \rangle}{\|\epsilon_{t,p}^+\| \|\epsilon_{t-p-1,p}^-\|} = \frac{\langle \epsilon_{t,p}^+, \epsilon_{t-p-1,p}^- \rangle}{\|\epsilon_{t,p}^+\| \|\epsilon_{t-p-1,p}^-\|} \quad (5.23)$$

En utilisant l'inégalité de Schwarz, on montre que $|k_{p+1}| \leq 1$. Remarquons aussi que k_{p+1} apparaît comme le coefficient de corrélation entre l'erreur de prédiction directe et l'erreur de prédiction rétrograde. Dans la littérature ce coefficient est appelé coefficient d'autocorrélation partielle.

Définition 5.2.1 (Fonction d'autocorrélation partielle). *Soit X_t un processus aléatoire, stationnaire au second ordre, de fonction de covariance $\gamma(h)$. On appelle fonction d'autocorrélation partielle la suite $\{k_p, p \geq 1\}$ définie par :*

$$k_p = \begin{cases} \text{corr}(X_t, X_{t-1}) = \frac{\langle X_t, X_{t-1} \rangle}{\|X_t\| \|X_{t-1}\|} & \text{pour } p = 1 \\ \text{corr}(\epsilon_{t,p-1}^+, \epsilon_{t-p,p-1}^-) = \frac{\langle X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p-1}), X_{t-p} - \text{proj}(X_{t-p} | \mathcal{H}_{t-1,p-1}) \rangle}{\|X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p-1})\| \|X_{t-p} - \text{proj}(X_{t-p} | \mathcal{H}_{t-1,p-1})\|} & \text{pour } p \geq 2 \end{cases} \quad (5.24)$$

Dans (5.24), l'expression pour $p = 1$ est en accord avec celle pour $p \geq 2$ dans la mesure où on peut noter que $\epsilon_{t,0}^+ = X_t$ et que $\epsilon_{t-1,0}^- = X_{t-1}$. Notons aussi que, dans l'expression de k_p , X_t et X_{t-p} sont projetés sur le même sous-espace $\text{Vect}(X_{t-1}, \dots, X_{t-p+1})$. Le résultat remarquable est que la suite des coefficients de corrélation partielle est donnée par :

$$k_p = \phi_{p,p} \quad (5.25)$$

où $\phi_{p,p}$ est défini au moyen des équations de Yule-Walker (5.10). Dans le cas particulier d'un processus AR(m) causal, on a alors :

$$k_p = \begin{cases} \phi_{p,p} & \text{pour } 1 \leq p < m \\ \phi_m & \text{pour } p = m \\ 0 & \text{pour } p > m \end{cases}$$

Notons enfin que contrairement à la fonction d'autocorrélation partielle d'un processus AR(m) qui vérifie $k_p = 0$ pour tout $p \geq m$, nous avons pour un processus MA(q), $k_p \neq 0$ pour un nombre infini de termes. Il est toutefois possible de montrer qu'il existe un réel ρ , $0 < \rho < 1$, et une constante C , telle que, pour tout $p \geq 1$, $|k_p| \leq C\rho^k$.

5.3 Algorithme de Schur

Partant des coefficients d'autocorrélation, l'algorithme de Levinson-Durbin évalue à la fois les coefficients des prédicteurs linéaires optimaux et les coefficients d'autocorrélation partielle. Dans certains cas, seuls les coefficients d'autocorrélation partielle sont nécessaires. Il en est ainsi, par exemple, lorsque l'on cherche à calculer les erreurs de prédiction directe et rétrograde à partir du processus X_t . Montrons, en effet, que les erreurs de prédiction à l'ordre $(p+1)$ s'expriment, en fonction des erreurs de prédictions à l'ordre p , à l'aide d'une formule de récurrence ne faisant intervenir que la valeur du coefficient de corrélation partielle :

$$\begin{cases} \epsilon_{t,p+1}^+ = \epsilon_{t,p}^+ - k_{p+1}\epsilon_{(t-1)-p,p}^- \\ \epsilon_{t-(p+1),p+1}^- = \epsilon_{(t-1)-p,p}^- - k_{p+1}\epsilon_{t,p}^+ \end{cases} \quad (5.26)$$

Reprenons les expressions de l'erreur de prédiction directe et de l'erreur de prédiction rétrograde :

$$\epsilon_{t,p}^+ = X_t - \sum_{k=1}^p \phi_{k,p} X_{t-k} \quad \text{et} \quad \epsilon_{t-p-1,p}^- = X_{t-p-1} - \sum_{k=1}^p \phi_{k,p} X_{t-p-1+k}$$

En utilisant directement la récursion de Levinson-Durbin, équations (5.22), dans l'expression de l'erreur de prédiction directe à l'ordre $p+1$, nous obtenons :

$$\begin{aligned} \epsilon_{t,p+1}^+ &= X_t - \sum_{k=1}^{p+1} \phi_{k,p+1} X_{t-k} \\ &= \left(X_t - \sum_{k=1}^p \phi_{k,p} X_{t-k} \right) - k_{p+1} \left(X_{t-p-1} - \sum_{k=1}^p \phi_{k,p} X_{t-p-1+k} \right) \\ &= \epsilon_{t,p}^+ - k_{p+1} \epsilon_{t-p-1,p}^- \end{aligned} \quad (5.27)$$

De façon similaire, nous avons :

$$\begin{aligned}
\epsilon_{t-p-1,p+1}^- &= X_{t-p-1} - \sum_{k=1}^{p+1} \phi_{k,p+1} X_{t-p-1+k} \\
&= \left(X_{t-p-1} - \sum_{k=1}^p \phi_{k,p} X_{t-p-1+k} \right) - k_{p+1} \left(X_t - \sum_{k=1}^p \phi_{k,p} X_{t-k} \right) \\
&= \epsilon_{t-p-1,p}^- - k_{p+1} \epsilon_{t,p}^+
\end{aligned} \tag{5.28}$$

Partant de la suite des autocorrélations, l'algorithme de Schur calcule récursivement les coefficients de corrélation partielle, sans avoir à déterminer les valeurs des coefficients de prédiction. Historiquement, l'algorithme de Schur a été introduit pour tester le caractère défini positif d'une suite (ou de façon équivalente, la positivité des matrices de Toeplitz construites à partir de cette suite). En effet, comme nous l'avons montré ci-dessus, une suite de coefficients de covariance est définie positive si et seulement si les coefficients de corrélation partielle sont de module strictement inférieur à 1. Déterminons à présent cet algorithme. En faisant $t = 0$ dans l'équation (5.27), en multipliant à gauche par X_m et en utilisant la stationnarité, il vient :

$$\langle X_m, \epsilon_{0,p+1}^+ \rangle = \langle X_m, \epsilon_{0,p}^+ \rangle - k_{p+1} \langle X_m, \epsilon_{-p-1,p}^- \rangle = \langle X_m, \epsilon_{0,p}^+ \rangle - k_{p+1} \langle X_{m+p+1}, \epsilon_{0,p}^- \rangle. \tag{5.29}$$

En faisant $t = p + 1$ dans l'équation (5.28), en multipliant à gauche par X_{m+p+1} et en utilisant la stationnarité, il vient :

$$\begin{aligned}
\langle X_{m+p+1}, \epsilon_{0,p+1}^- \rangle &= \langle X_{m+p+1}, \epsilon_{0,p}^- \rangle - k_{p+1} \langle X_{m+p+1}, \epsilon_{p+1,p}^+ \rangle \\
&= \langle X_{m+p+1}, \epsilon_{0,p}^- \rangle - k_{p+1} \langle X_m, \epsilon_{0,p}^+ \rangle.
\end{aligned} \tag{5.30}$$

En faisant $m = 0$ dans (5.30), il vient :

$$\langle X_{p+1}, \epsilon_{0,p+1}^- \rangle = \langle X_{p+1}, \epsilon_{0,p}^- \rangle - k_{p+1} \langle X_{p+1}, \epsilon_{p+1,p}^+ \rangle = \langle X_{p+1}, \epsilon_{0,p}^- \rangle - k_{p+1} \langle X_0, \epsilon_{0,p}^+ \rangle. \tag{5.31}$$

Mais on a aussi :

$$\langle X_{p+1}, \epsilon_{0,p+1}^- \rangle = \langle X_{p+1}, X_0 - \text{proj}(X_0 | \text{Vect}(X_1, \dots, X_{p+1})) \rangle = 0.$$

Nous pouvons donc déduire de l'équation (5.31) :

$$k_{p+1} = \frac{\langle X_{p+1}, \epsilon_{0,p}^- \rangle}{\langle X_0, \epsilon_{0,p}^+ \rangle} \tag{5.32}$$

En couplant les équations (5.29), (5.30) et (5.32) et en partant des conditions initiales :

$$\langle X_m, \epsilon_{0,0}^+ \rangle = \gamma(m) \quad \text{et} \quad \langle X_{m+1}, \epsilon_{0,0}^- \rangle = \gamma(m+1).$$

on peut déterminer les coefficients de corrélation partielle directement, sans avoir à évaluer explicitement les coefficients de prédiction.

On note $u(m, p) = \langle X_m, \epsilon_{0,p}^+ \rangle$ et $v(m, p) = \langle X_{m+p+1}, \epsilon_{0,p}^- \rangle$. Partant des $(K + 1)$ coefficients de covariance $\{\gamma(0), \dots, \gamma(K)\}$, l'algorithme de Schur calcule les K premiers coefficients de corrélation partielle :

Initialisation Pour $m = \{0, \dots, K - 1\}$:

$$\begin{aligned} u(m, 0) &= \gamma(m) \\ v(m, 0) &= \gamma(m + 1) \end{aligned}$$

Récursion — Pour $p = \{1, \dots, K\}$, calculer

$$k_p = \frac{v(0, p - 1)}{u(0, p - 1)}$$

— Pour $m = \{0, \dots, K - p - 1\}$ calculer :

$$\begin{cases} u(m, p) = u(m, p - 1) - k_p v(m, p - 1) \\ v(m, p) = v(m + 1, p - 1) - k_p u(m + 1, p - 1) \end{cases}$$

La complexité de l'algorithme de Schur est équivalente à l'algorithme de Levinson.

Filtres en treillis

En notant $e(t, p) = [\epsilon_{t,p}^+ \quad \epsilon_{t-p,p}^-]^T$ et en utilisant l'opérateur de retard B , les expressions (5.26) peuvent se mettre sous la forme matricielle :

$$e(t, p + 1) = \begin{bmatrix} 1 & -k_{p+1}B \\ -k_{p+1}B & 1 \end{bmatrix} e(t, p)$$

Les erreurs initiales ($p = 0$) sont $e(t, 0) = [X_t \quad X_t]^T$. Ces équations débouchent sur une structure de filtrage dite en treillis qui calcule, au moyen des coefficients de corrélation partielle, les erreurs de prédiction directe et rétrograde à partir du processus $\{X_t, t \in \mathbb{Z}\}$. Ce filtre d'analyse est représenté figure 5.1. Les équations (5.26) peuvent encore s'écrire :

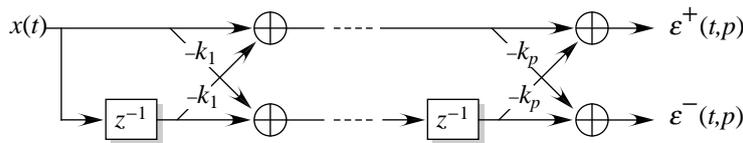


FIGURE 5.1 — Filtre d'analyse en treillis. Ce filtre permet de construire les erreurs de prédiction directe et rétrograde à partir du processus et de la donnée des coefficients de corrélation partielle.

$$\begin{cases} \epsilon_{t,p}^+ = \epsilon_{t,p+1}^+ + k_{p+1} \epsilon_{(t-1)-p,p}^- \\ \epsilon_{t-(p+1),p+1}^- = \epsilon_{(t-1)-p,p}^- k_{p+1} \epsilon_{t,p}^+ \end{cases}$$

qui donne le schéma de filtrage de la figure 5.2.

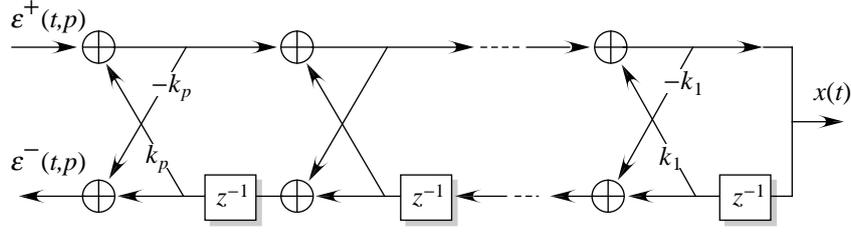


FIGURE 5.2 – Filtre de synthèse en treillis. Ce filtre permet de reconstruire le processus à partir de la suite des erreurs de prédiction directe et de la donnée des coefficients de corrélation partielle.

5.4 Algorithme des innovations

L'algorithme des innovations est une application directe de la méthode de gram-Schmidt et est, à cet égard, plus élémentaire que l'algorithme de Levinson-Durbin. Il ne suppose de plus pas que le processus $\{X_t, t \in \mathbb{Z}\}$ est stationnaire. Supposons, sans perte de généralité que $\mathbb{E}[X_t] = 0$ et notons

$$\kappa(i, j) = \langle X_i, X_j \rangle = \mathbb{E}[X_i X_j],$$

la fonction d'autocovariance de ce processus. Nous supposons dans ce paragraphe, que pour tout $n \geq 1$, la matrice $[\kappa(i, j)]_{i,j=1}^n$ est non singulière. Définissons, pour $n \geq 1$, $\mathcal{H}_n = \text{Vect}(X_1, \dots, X_n)$, et $\sigma_n^2 = \|X_{n+1} - \text{proj}(X_{n+1} | \mathcal{H}_n)\|^2$. Il est clair que, pour tout $n \geq 1$,

$$\mathcal{H}_n = \text{Vect}(X_1, X_2 - \text{proj}(X_2 | X_1), \dots, X_n - \text{proj}(X_n | \mathcal{H}_{n-1})),$$

ce qui implique que

$$\text{proj}(X_{n+1} | \mathcal{H}_n) = \sum_{j=1}^n \theta_{n,j} (X_{n+1-j} - \text{proj}(X_{n+1-j} | \mathcal{H}_{n-j})). \quad (5.33)$$

Nous allons maintenant montrer qu'il est possible de déterminer de façon récursive les coefficients $\{\theta_{n,j}, 1 \leq j \leq n\}$. Remarquons en effet que les vecteurs $\{X_i - \text{proj}(X_i | \mathcal{H}_{i-1}), i \geq 1\}$ forment une famille orthogonale. En effet, pour $i < j$, $X_i - \text{proj}(X_i | \mathcal{H}_{i-1}) \in \mathcal{H}_{j-1}$ et $X_j - \text{proj}(X_j | \mathcal{H}_{j-1}) \perp \mathcal{H}_{j-1}$. Par conséquent, pour $0 \leq k < n$,

$$\langle \text{proj}(X_{n+1} | \mathcal{H}_n), X_{k+1} - \text{proj}(X_{k+1} | \mathcal{H}_k) \rangle = \theta_{n,n-k} \sigma_{k+1}^2.$$

Since $\langle X_{n+1} - \text{proj}(X_{n+1} | \mathcal{H}_n), X_{k+1} - \text{proj}(X_{k+1} | \mathcal{H}_k) \rangle = 0$, les coefficients $\theta_{n,n-k}$, $k = 0, \dots, n-1$ sont donnés par

$$\theta_{n,n-k} = \sigma_{k+1}^{-2} \langle X_{n+1}, X_{k+1} - \text{proj}(X_{k+1} | \mathcal{H}_k) \rangle.$$

En utilisant la représentation (5.33), nous avons donc

$$\theta_{n,n-k} = \sigma_{k+1}^{-2} \left(\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \langle X_{n+1}, X_{j+1} - \text{proj}(X_{j+1} | \mathcal{H}_j) \rangle \right).$$

Comme $\langle X_{n+1}, X_{j+1} - \text{proj}(X_{j+1} | \mathcal{H}_j) \rangle = \sigma_{j+1}^2 \theta_{n,n-j}$ pour $0 \leq j < n$, nous avons donc pour $k \in \{1, \dots, n\}$,

$$\theta_{n,n-k} = \sigma_{k+1}^{-2} \left(\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} \sigma_{j+1}^2 \right). \quad (5.34)$$

Le théorème de projection implique que

$$\begin{aligned} \sigma_{n+1}^2 &= \|X_{n+1} - \text{proj}(X_{n+1} | \mathcal{H}_n)\|^2 = \|X_{n+1}\|^2 - \|\text{proj}(X_{n+1} | \mathcal{H}_n)\|^2 \\ &= \kappa(n+1, n+1) - \sum_{k=0}^{n-1} \theta_{n,n-k}^2 \sigma_{k+1}^2. \end{aligned} \quad (5.35)$$

Remarquons qu'alors que l'algorithme de Durbin Levinson permet de déterminer les coefficients du développement de $\text{proj}(X_{n+1} | \mathcal{H}_n)$ sur X_1, \dots, X_n , $\text{proj}(X_{n+1} | \mathcal{H}_n) = \sum_{j=1}^n \phi_{n,j} X_{n+1-j}$, l'algorithme précédent calcule les coefficients du développement de $\text{proj}(X_{n+1} | \mathcal{H}_n)$ sur la suite des *innovations*, $X_1, X_2 - \text{proj}(X_2 | X_1), \dots, X_n - \text{proj}(X_n | \mathcal{H}_{n-1})$.

Exemple 5.4.1 (Prédiction d'un processus MA(1)). *Considérons le processus $X_t = Z_t + \theta Z_{t-1}$ où $\{Z_t\} \sim \text{BB}(0, \sigma^2)$. Nous avons donc $\kappa(i, j) = 0$ pour $|i - j| > 1$, $\kappa(i, i) = \sigma^2(1 + \theta^2)$ et $\kappa(i, i+1) = \theta\sigma^2$. Dans ce cas, nous avons*

$$\begin{aligned} \theta_{n,j} &= 0 & 2 \leq j \leq n, \\ \theta_{n,1} &= \sigma_{n-1}^{-2} \theta \sigma^2, \end{aligned}$$

où les variances des innovations sont données par les récursions

$$\begin{aligned} \sigma_1^2 &= (1 + \theta^2)\sigma^2, \\ \sigma_{n+1}^2 &= [1 + \theta^2 - \sigma_n^{-2} \theta^2 \sigma^2] \sigma^2. \end{aligned}$$

Si nous posons $r_n = \sigma_n^2 / \sigma^2$, nous avons

$$\text{proj}(X_{n+1} | \mathcal{H}_n) = \theta (X_n - \text{proj}(X_n | \mathcal{H}_{n-1})) / r_n,$$

avec $r_1 = 1 + \theta^2$, et pour $n \geq 1$, $r_{n+1} = 1 + \theta^2 - \theta^2 / r_n$.

5.5 Décomposition de Wold

Un des résultats fondamentaux de la théorie des processus stationnaires au second-ordre est la décomposition de Wold. Cette décomposition permet de décomposer n'importe quel processus stationnaire au second-ordre comme la somme d'un processus résultant du filtrage linéaire d'un bruit blanc et d'un processus déterministe (définition 5.1.4). La preuve de ce résultat est de nature géométrique. L'idée de base est la suivante. Soit $\mathcal{H}_t^X = \overline{\text{Vect}(X_s, s \leq t)}$. \mathcal{H}_t^X est appelé le *passé linéaire* du processus à la date t . Par construction, $\mathcal{H}_t^X \subset \mathcal{H}_{t+1}^X$, et nous disposons ainsi d'une famille de sous-espace emboîtés de $\mathcal{H}_\infty^X = \overline{\cup_{t \in \mathbb{Z}} \mathcal{H}_t^X}$. \mathcal{H}_∞^X est l'enveloppe linéaire du processus. L'espace $\bigcap_{t \in \mathbb{Z}} \mathcal{H}_t^X$, appelé le *passé infini du processus* $\{X_t, t \in \mathbb{Z}\}$ jouera aussi un rôle particulier. Par définition X_t appartient

à \mathcal{H}_t^X , mais il n'appartient généralement pas à \mathcal{H}_{t-1}^X . Le théorème de projection dit qu'il existe un unique élément noté $\text{proj}(X_t | \mathcal{H}_{t-1}^X)$ et appartenant à \mathcal{H}_{t-1}^X tel que :

$$\epsilon_t = X_t - \text{proj}(X_t | \mathcal{H}_{t-1}^X) \perp \mathcal{H}_{t-1}^X$$

Dans ce contexte ϵ_t s'appelle l'*innovation* (linéaire) du processus. Il découle de cette construction géométrique que le processus d'innovation est un *processus orthogonal* dans le sens où :

$$\forall s \neq t, \quad \epsilon_s \perp \epsilon_t \tag{5.36}$$

En effet, pour $s < t$, nous pouvons écrire $\epsilon_s \in \mathcal{H}_s^X \subset \mathcal{H}_{t-1}^X$ et $\epsilon_t \perp \mathcal{H}_{t-1}^X$. Et donc $\epsilon_s \perp \epsilon_t$.

La proposition qui suit montre que le processus d'innovation est la limite des processus d'innovations partielles à l'ordre p .

Proposition 5.5.1. *Pour tout $Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ et tout $t \in \mathbb{Z}$ nous avons :*

$$\lim_{p \rightarrow \infty} \text{proj}(Y | \mathcal{H}_{t,p}^X) = \text{proj}(Y | \mathcal{H}_t^X)$$

où $\mathcal{H}_{t,p}^X = \text{Vect}(X_t, X_{t-1}, \dots, X_{t-p+1})$.

Exemple 5.5.2 (Bruit blanc). *Supposons que $\{X_t\}$ soit un bruit blanc. Nous avons $\text{proj}(X_t | \mathcal{H}_{t-1,p}^X) = 0$ pour tout p et donc $\text{proj}(X_t | \mathcal{H}_{t-1}^X) = 0$. Nous avons donc $\epsilon_t = X_t - \text{proj}(X_t | \mathcal{H}_{t-1}^X) = X_t$: le processus X_t coïncide avec son innovation. Ceci signifie qu'un bruit blanc ne peut être prédit de façon linéaire à partir de son passé.*

Exemple 5.5.3 (Prédiction d'un processus AR(p) causal). *On considère le processus AR(p) causal défini par l'équation récurrente $X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t$ où $Z_t \sim \text{BB}(0, \sigma^2)$. Dans le cas causal, on a $\mathcal{H}_t^X = \mathcal{H}_t^Z$ et, pour tout $k \geq 1$, on a $\mathbb{E}[X_{t-k} Z_t] = 0$. Par conséquent $Z_t \perp \mathcal{H}_{t-1}^X$ et $\mathcal{H}_t^X = \mathcal{H}_{t-1}^X \oplus \text{Vect}(Z_t)$. On en déduit que :*

$$\text{proj}(X_t | \mathcal{H}_{t-1}^X) = \sum_{k=1}^p \phi_k \text{proj}(X_{t-k} | \mathcal{H}_{t-1}^X) + \text{proj}(Z_t | \mathcal{H}_{t-1}^X) = \sum_{k=1}^p \phi_k X_{t-k}$$

et donc $X_t - \text{proj}(X_t | \mathcal{H}_{t-1}^X) = X_t - \sum_{k=1}^p \phi_k X_{t-k} = Z_t$. Par conséquent le bruit blanc Z_t , qui intervient dans l'équation récurrente d'un AR causal, est précisément l'innovation du processus AR. Ce résultat montre que $\sum_{k=1}^p \phi_k X_{t-k}$ est la projection de $X(t)$ sur tout le passé \mathcal{H}_{t-1} et qu'elle coïncide avec la projection orthogonale sur le passé $\mathcal{H}_{t-1,p}$ de durée p . Par conséquent, pour tout $m \geq p$, la suite des coefficients de prédiction est $\{\phi_1, \dots, \phi_p, \underbrace{0, \dots, 0}_{m-p}\}$. Ce résultat est faux pour un AR non causal.

Exemple 5.5.4 (Processus harmonique). *Soit le processus harmonique $X_t = A \cos(\lambda_0 t + \Phi)$ où A est une variable aléatoire, centrée, de variance σ_A^2 et Φ une variable aléatoire, indépendante de A et distribuée suivant une loi uniforme sur $[-\pi, \pi]$. Le processus X_t est stationnaire au second-ordre, centré, de fonction d'autocovariance $\gamma(\tau) = (\sigma_A^2/2) \cos(\lambda_0 \tau)$. Les coefficients du prédicteur linéaire optimal à l'ordre 2 sont donnés par :*

$$\begin{bmatrix} \phi_{1,2} \\ \phi_{2,2} \end{bmatrix} = \begin{bmatrix} 1 & \cos(\lambda_0) \\ \cos(\lambda_0) & 1 \end{bmatrix}^{-1} \begin{bmatrix} \cos(\lambda_0) \\ \cos(2\lambda_0) \end{bmatrix} = \begin{bmatrix} 2 \cos(\lambda_0) \\ -1 \end{bmatrix}$$

On vérifie facilement que $\sigma_2^2 = \|X_t - \text{proj}(X_t | \mathcal{H}_{t-1,2}^X)\|^2 = 0$. Par conséquent, on a :

$$X_t = \text{proj}(X_t | \mathcal{H}_{t-1,2}^X) = 2 \cos(\lambda_0) X_{t-1} - X_{t-2} \in \mathcal{H}_{t-1}^X$$

et donc la projection $\text{proj}(X_t | \mathcal{H}_{t-1}^X) = X_t$, ce qui implique que $\epsilon_t = 0$. A l'inverse du bruit blanc, le processus est entièrement prédictible à partir de son passé.

En appliquant la proposition 5.5.1 à X_t , nous pouvons écrire :

$$\lim_{p \rightarrow \infty} \text{proj}(X_t | \mathcal{H}_{t-1,p}^X) = \text{proj}(X_t | \mathcal{H}_{t-1}^X) \quad \text{et} \quad \lim_{p \rightarrow \infty} \epsilon_{t,p}^+ = \epsilon_t \quad (5.37)$$

Le processus d'innovation ϵ_t est donc la limite en moyenne quadratique de la suite des innovations partielles $\epsilon_{t,p}^+ = X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p}^X)$. Une conséquence immédiate est que le processus d'innovation est un processus stationnaire au second ordre. En utilisant, en effet, la continuité du produit scalaire et la stationnarité au second ordre de l'innovation partielle d'ordre p , on peut écrire :

$$\langle \epsilon_{t+\tau}, \epsilon_t \rangle = \lim_{p \rightarrow \infty} \langle \epsilon_{t+\tau,p}^+, \epsilon_{t,p}^+ \rangle = \lim_{p \rightarrow \infty} \langle \epsilon_{\tau,p}^+, \epsilon_{0,p}^+ \rangle \quad (5.38)$$

qui ne dépend que de τ . En particulier nous avons :

$$\sigma^2 = \|\epsilon_t\|^2 = \lim_{p \rightarrow \infty} \|X_t - \text{proj}(X_t | \mathcal{H}_{t,p}^X)\|^2 = \lim_{p \rightarrow \infty} \sigma_p^2$$

Dans le cas du bruit blanc on obtient $\sigma^2 = \mathbb{E}[X_t^2] \neq 0$ et donc, d'après la définition 5.1.4, le bruit blanc est un processus régulier. D'un autre côté, le processus harmonique, pour lequel $\sigma^2 = 0$, est déterministe. Nous remarquons aussi que la somme d'un bruit blanc et d'un processus harmonique est un processus régulier.

La structure géométrique emboîtée des espaces $\{\mathcal{H}_t^X\}$ et l'orthogonalité des innovations fournissent, pour tout $s < t$, la formule suivante de décomposition en somme directe :

$$\mathcal{H}_t^X = \mathcal{H}_s^X \oplus \text{Vect}(\epsilon_{s+1}, \dots, \epsilon_t) \quad (5.39)$$

Notons, tout d'abord, que $\epsilon_t = X_t - \text{proj}(X_t | \mathcal{H}_{t-1}^X) \in \mathcal{H}_t^X$ et que $\epsilon_t \perp \mathcal{H}_{t-1}^X$, ce qui implique que $\mathcal{H}_{t-1}^X \oplus \text{Vect}(\epsilon_t) \subseteq \mathcal{H}_t^X$. D'un autre côté, puisque $X_t = \epsilon_t + \text{proj}(X_t | \mathcal{H}_{t-1}^X)$,

$$\mathcal{H}_t^X = \overline{\text{Vect}(\epsilon_t + \text{proj}(X_t | \mathcal{H}_{t-1}^X), \{X_s, s \leq t-1\})} = \overline{\text{Vect}(\epsilon_t, \{X_s, s \leq t-1\})},$$

ce qui entraîne que $\mathcal{H}_t^X \subseteq \mathcal{H}_{t-1}^X \oplus \text{Vect}(\epsilon_t)$. En conclusion $\mathcal{H}_t^X = \mathcal{H}_{t-1}^X \oplus \text{Vect}(\epsilon_t)$. En réitérant ce raisonnement, on en déduit la décomposition (5.39). Cette décomposition orthogonale de l'espace \mathcal{H}_t^X n'est pas sans rappeler la décomposition de Gram-Schmidt. Notons qu'à l'inverse de la décomposition de Gram-Schmidt classique, nous procédons ici dans le sens rétrograde. Définissons pour tout $s \geq 0$:

$$\psi_s = \frac{\langle X_t, \epsilon_{t-s} \rangle}{\sigma^2} \quad (5.40)$$

Remarquons que ψ_s ne dépend pas de t . En effet, la continuité du produit scalaire et la stationnarité conjointe du processus X_t et de l'innovation partielle impliquent que :

$$\langle X_t, \epsilon_{t-s} \rangle = \lim_{p \rightarrow \infty} \langle X_t, \epsilon_{t-s,p}^+ \rangle = \lim_{p \rightarrow \infty} \langle X_0, \epsilon_{-s,p}^+ \rangle$$

Lemme 5.5.5. *La suite $\{\psi_s, s \in \mathbb{Z}\}$ est de carré sommable et $\psi_0 = 1$.*

Démonstration. Remarquons, tout d'abord, que la relation $\langle \text{proj}(X_t | \mathcal{H}_{t-1}^X), \epsilon_t \rangle = 0$ entraîne que :

$$\begin{aligned} \psi_0 &= \frac{\langle X_t, \epsilon_t \rangle}{\sigma^2} \\ &= \frac{\langle X_t - \text{proj}(X_t | \mathcal{H}_{t-1}^X), \epsilon_t \rangle}{\sigma^2} \\ &= 1. \end{aligned}$$

D'autre part, pour tout $s \geq 0$, la projection orthogonale de X_t sur $\mathcal{H}_{t,s}^\epsilon = \text{Vect}(\epsilon_t, \epsilon_{t-1}, \dots, \epsilon_{t-s+1})$ s'écrit, du fait de l'orthogonalité du processus d'innovation, $\text{proj}(X_t | \mathcal{H}_{t,s}^\epsilon) = \sum_{k=0}^{s-1} \psi_k \epsilon_{t-k}$. On en déduit que $\|\text{proj}(X_t | \mathcal{H}_{t,s}^\epsilon)\|^2 = \sigma^2 \sum_{k=0}^{s-1} \psi_k^2$. On a alors d'après l'égalité de Pythagore (proposition A.2.2) :

$$\|\text{proj}(X_t | \mathcal{H}_{t,s}^\epsilon)\|^2 = \sigma^2 \sum_{k=0}^{s-1} \psi_k^2 = \|X_t\|^2 - \|X_t - \text{proj}(X_t | \mathcal{H}_{t,s}^\epsilon)\|^2 \leq \|X_t\|^2$$

ce qui conclut la preuve. \square

La suite $(\psi_s)_{s \geq 0}$ étant de carré sommable, la suite $s \rightarrow X_{t,s} = \sum_{k=0}^s \psi_k \epsilon_{t-k}$ est, pour t fixé, une suite de Cauchy dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Elle admet donc, quand $s \rightarrow \infty$, une limite que nous notons :

$$U_t = \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}$$

et qui est un processus stationnaire au second-ordre. On a, en effet :

$$\mathbb{E}[U_t] = (U_t, 1) = \lim_{s \rightarrow \infty} \sum_{k=0}^s \psi_k (\epsilon_{t-k}, 1) = 0$$

et

$$\mathbb{E}[U_{t+\tau} U_t] = (U_{t+\tau}, U_t) = \lim_{s \rightarrow \infty} \left(\sum_{k=0}^s \psi_k \epsilon_{t+\tau-k}, \sum_{k=0}^s \psi_k \epsilon_{t-k} \right) = \lim_{s \rightarrow \infty} \left(\sum_{k=0}^s \psi_k \epsilon_{\tau-k}, \sum_{k=0}^s \psi_k \epsilon_{-k} \right)$$

qui est indépendant de t .

Le théorème suivant, connu sous le nom de *décomposition de Wold*, est vraisemblablement le résultat le plus important de la théorie des processus stationnaires au second-ordre.

Théorème 5.5.6 (Décomposition de Wold). *Soit X_t un processus stationnaire au second ordre et ϵ_t son processus d'innovation. On suppose que X_t est un processus régulier ($\sigma^2 = \|\epsilon_t\|^2 \neq 0$). On note $U_t = \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}$ où $\psi_k = \langle X_t, \epsilon_{t-k} \rangle / \sigma^2$. Alors il existe un processus V_t tel que :*

$$X_t = U_t + V_t, \tag{5.41}$$

et tel que :

- (i) pour tout (t, s) , $\langle V_t, \epsilon_s \rangle = 0$, qui implique que $\langle V_t, U_s \rangle = 0$,
- (ii) $V_t = \text{proj}(X_t | \mathcal{H}_{-\infty}^X)$ est la projection orthogonale de X_t sur $\mathcal{H}_{-\infty}^X = \bigcap_{t=-\infty}^{\infty} \mathcal{H}_t^X$,
- (iii) U_t est un processus régulier et $\epsilon_t = U_t - \text{proj}(U_t | \mathcal{H}_{t-1}^U)$ est l'innovation du processus $\{U_t, t \in \mathbb{Z}\}$.
De plus, $\mathcal{H}_t^\epsilon = \mathcal{H}_t^U$.
- (iv) V_t est un processus déterministe et $\mathcal{H}_t^V = \mathcal{H}_{-\infty}^X$.

Démonstration. Elle est donnée en fin de chapitre. □

Un processus $\{X_t\}$ tel que $\mathcal{H}_{-\infty}^X = \{0\}$ est dit *purement non déterministe*. Pour un tel processus la partie déterministe de la décomposition de Wold est identiquement nulle. Par exemple, le processus régulier U_t de la décomposition de Wold est purement non déterministe. En effet, en appliquant la décomposition de Wold au processus U_t on a, pour tout t , $U_t = U_t + V_t$ avec $V_t = 0$ et donc, d'après le point iv, $\mathcal{H}_{-\infty}^U = \{0\}$. Le théorème de Wold permet donc de décomposer tout processus stationnaire au second-ordre sous la forme d'une somme de deux processus orthogonaux, le premier étant *purement non déterministe* et le second étant *déterministe*. La partie purement non-déterministe s'exprime comme le filtrage d'un bruit blanc par un filtre linéaire invariant dans le temps de réponse impulsionnelle $\{\psi_k\}$ causale ($\psi_k = 0$ pour $k < 0$) et de carré sommable (pas nécessairement de module sommable).

Exemple 5.5.7 (Processus MA(1)). Soit $\{Z_t\}$ un bruit blanc et soit le processus $X_t = Z_t + \theta_1 Z_{t-1}$. Remarquons que, par construction, $\mathcal{H}_t^X \subseteq \mathcal{H}_t^Z$ mais que l'inclusion réciproque n'est pas nécessairement vérifiée. Montrons par contre que, pour $|\theta_1| < 1$, nous avons effectivement $\mathcal{H}_t^X = \mathcal{H}_t^Z$. En effet, en réitérant p fois l'équation $X_t = Z_t + \theta_1 Z_{t-1}$ et en résolvant par rapport à Z_t , nous obtenons :

$$Z_t = X_t - \theta_1 X_{t-1} + \theta_1^2 X_{t-2} + \dots + (-1)^p \theta_1^p X_{t-p} - (-1)^p \theta_1^{p+1} Z_{t-p}$$

En prenant la limite en p , nous en déduisons que, si $|\theta_1| < 1$, alors :

$$Z_t = \sum_{k=0}^{\infty} (-\theta_1)^k X_{t-k}$$

ce qui montre que $\mathcal{H}_t^Z \subset \mathcal{H}_t^X$ et donc que $\mathcal{H}_t^X = \mathcal{H}_t^Z$. Dans ce cas, nous pouvons écrire :

$$\begin{aligned} \text{proj}(X_t | \mathcal{H}_{t-1}^X) &= \text{proj}(Z_t | \mathcal{H}_{t-1}^X) + \theta_1 \text{proj}(Z_{t-1} | \mathcal{H}_{t-1}^X) = \text{proj}(Z_t | \mathcal{H}_{t-1}^Z) + \theta_1 \text{proj}(Z_{t-1} | \mathcal{H}_{t-1}^Z) \\ &= 0 + \theta_1 Z_{t-1}, \end{aligned}$$

en remarquant que $\text{proj}(Z_t | \mathcal{H}_{t-1}^Z) = 0$ car $\{Z_t, t \in \mathbb{Z}\}$ est un bruit blanc. On en déduit que $X_t - \text{proj}(X_t | \mathcal{H}_{t-1}^X) = X_t - \theta_1 Z_{t-1} = Z_t$. Par conséquent, lorsque $|\theta_1| < 1$, le processus $\{Z_t, t \in \mathbb{Z}\}$ est l'innovation du processus $\{X_t, t \in \mathbb{Z}\}$. Notons que le processus $\{X_t, t \in \mathbb{Z}\}$ est purement non déterministe et que les coefficients de la décomposition de Wold sont simplement donnés par $\psi_0 = 1$, $\psi_1 = \theta$, et $\psi_k = 0$ pour $k > 1$.

5.6 Preuves des théorèmes 5.1.5 et 5.5.6

Preuve du théorème 5.1.5. Nous allons tout d'abord montrer que le prédicteur optimal n'a pas de racines sur le cercle unité. Raisonnons par contradiction. Supposons que le polynôme $\phi_p(z)$ ait deux racines complexes conjuguées, de la forme $\exp(\pm i\pi\theta)$, sur le cercle unité. (on traite de façon similaire le cas de racines réelles, $\theta = 0$ ou π). Nous pouvons écrire :

$$\phi_p(z) = \phi_p^*(z)(1 - 2\cos(\theta)z + z^2)$$

On note $\bar{\nu}_X(d\lambda) = \nu_X(d\lambda)|\phi_p^*(e^{-i\lambda})|^2$. $\bar{\nu}_X$ est une mesure positive sur $[-\pi, \pi]$ de masse finie. On note $\bar{\gamma}(\tau)$ la suite des coefficients de Fourier associés à $\bar{\nu}_X$:

$$\bar{\gamma}(\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\tau\lambda} \bar{\nu}_X(d\lambda)$$

Nous avons donc :

$$\sigma_p^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} (1 - 2\cos(\theta)e^{-i\lambda} + e^{-2i\lambda}) \bar{\nu}_X(d\lambda) = \inf_{\psi \in \mathcal{P}_2} \frac{1}{2\pi} \int_{-\pi}^{\pi} |1 + \psi_1 e^{-i\lambda} + \psi_2 e^{-2i\lambda}|^2 \bar{\nu}_X(d\lambda).$$

Comme on l'a dit (page 48), la minimisation de σ_p^2 par rapport à ψ_1 et ψ_2 est équivalent à la résolution des équations de Yule-Walker à l'ordre $p = 2$ pour la suite des covariances $\bar{\gamma}(h)$. Par conséquent la suite des coefficients $\{1, -2\cos(\theta), 1\}$ doit vérifier l'équation :

$$\begin{bmatrix} \bar{\gamma}(0) & \bar{\gamma}(1) & \bar{\gamma}(2) \\ \bar{\gamma}(1) & \bar{\gamma}(0) & \bar{\gamma}(1) \\ \bar{\gamma}(2) & \bar{\gamma}(1) & \bar{\gamma}(0) \end{bmatrix} \begin{bmatrix} 1 \\ -2\cos(\theta) \\ 1 \end{bmatrix} = \begin{bmatrix} \sigma_p^2 \\ 0 \\ 0 \end{bmatrix}$$

De cette équation il s'en suit (les première et troisième lignes sont égales) que $\sigma_p^2 = 0$. Ce qui est contraire à l'hypothèse que le processus est régulier.

Démontrons maintenant que les racines des polynômes prédicteurs sont toutes *strictement à l'extérieur du cercle unité*. Raisonnons encore par l'absurde. Supposons que le polynôme prédicteur à l'ordre p ait m racines $\{a_k, |a_k| < 1, 1 \leq k \leq m\}$ à l'intérieur du cercle unité et $(p - m)$ racines $\{b_\ell, |b_\ell| > 1, 1 \leq \ell \leq p - m\}$ à l'extérieur du cercle unité. Le polynôme prédicteur à l'ordre p s'écrit donc :

$$\phi_p(z) = \prod_{k=1}^m (1 - a_k^{-1}z) \prod_{\ell=1}^{p-m} (1 - b_\ell^{-1}z)$$

Considérons alors le polynôme :

$$\bar{\phi}_p(z) = \prod_{k=1}^m (1 - a_k^* z) \prod_{\ell=1}^{p-m} (1 - b_\ell^{-1} z)$$

Il a d'une part toutes ses racines strictement à l'extérieur du cercle unité et d'autre part il vérifie $|\bar{\phi}_p(e^{-i\lambda})|^2 < |\phi_p(e^{-i\lambda})|^2$. On a en effet $|1 - a_k^* e^{-i\lambda}| = |1 - a_k e^{i\lambda}| = |a_k| |1 - a_k^{-1} e^{-i\lambda}|$ et donc $|\bar{\phi}_p(e^{-i\lambda})|^2 =$

$(\prod_{k=1}^m |a_k|^2) |\phi_p(e^{-i\lambda})|^2$, ce qui démontre le résultat annoncé compte tenu du fait que $|a_k| < 1$. On en déduit alors que :

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |\bar{\phi}_p(e^{-i\lambda})|^2 \nu_X(d\lambda) < \sigma_p^2$$

ce qui contredit que $\phi_p(z) = \inf_{\psi \in \mathcal{P}_p} (2\pi)^{-1} \int_{-\pi}^{\pi} |\psi(e^{-i\lambda})|^2 \nu_X(d\lambda)$. \square

Preuve du théorème 5.5.6. (i). Par définition, $V_t = X_t - \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k} \in \mathcal{H}_t^X$. Pour $s > t$, $\epsilon_s \perp \mathcal{H}_t^X$, et donc $\langle V_t, \epsilon_s \rangle = 0$. Pour $s \leq t$, $\langle V_t, \epsilon_s \rangle = \langle X_t, \epsilon_s \rangle - \psi_{t-s} \sigma^2$ qui est égal à 0 par définition de ψ_k .

(ii). Montrons tout d'abord que $V_t \in \mathcal{H}_{-\infty}^X$. La preuve se fait par récurrence. Nous avons $V_t \in \mathcal{H}_t^X$ et $V_t \perp \epsilon_t$ (d'après la propriété précédente). Comme $\mathcal{H}_t^X = \mathcal{H}_{t-1}^X \oplus \text{Vect}(\epsilon_t)$, on en déduit que $V_t \in \mathcal{H}_{t-1}^X$. Supposons à présent que $V_t \in \mathcal{H}_{t-s}^X$, pour $s \geq 0$. Comme $V_t \perp \epsilon_{t-s}$ et que $\mathcal{H}_{t-s}^X = \mathcal{H}_{t-s-1}^X \oplus \text{Vect}(\epsilon_{t-s})$, nous avons $V_t \in \mathcal{H}_{t-s-1}^X$. On a donc $V_t \in \mathcal{H}_{-\infty}^X = \bigcap_{s=-\infty}^{\infty} \mathcal{H}_s^X$. Il reste à montrer que $X_t - V_t = \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}$ est orthogonal à $\mathcal{H}_{-\infty}^X$. Pour cela considérons $Y \in \mathcal{H}_{-\infty}^X$. Nous avons :

$$\langle X_t - V_t, Y \rangle = \left\langle \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}, Y \right\rangle = \lim_{s \rightarrow +\infty} \sum_{k=0}^s \psi_k \langle \epsilon_{t-k}, Y \rangle$$

Mais, par définition, $Y \in \mathcal{H}_{-\infty}^X$ implique que, pour tout t , $Y \in \mathcal{H}_t^X$. Comme $\epsilon_{t-k} \perp \mathcal{H}_{t-s-1}^X$ pour $0 \leq k \leq s$, nous avons $\sum_{k=0}^s \psi_k \langle \epsilon_{t-k}, Y \rangle = 0$. Et donc, pour tout $Y \in \mathcal{H}_{-\infty}^X$, on a :

$$\langle X_t - V_t, Y \rangle = \langle U_t, Y \rangle = 0. \quad (5.42)$$

(iii). Notons que (5.42) implique que, pour tout t , $U_t \perp \mathcal{H}_{-\infty}^X$ et donc $\mathcal{H}_t^U = \overline{\text{Vect}(U_s, s \leq t)} \perp \mathcal{H}_{-\infty}^X$. On peut alors poser $\mathcal{L}_t = \mathcal{H}_t^U \oplus \mathcal{H}_{-\infty}^X$. La décomposition $X_t = U_t + V_t$ et la propriété précédente ($V_t = \text{proj}(X_t | \mathcal{H}_{-\infty}^X)$) impliquent que, pour tout t , $\mathcal{H}_t^X \subset \mathcal{L}_t$, et donc $\epsilon_t \in \mathcal{L}_t$. Comme, pour tout t , $\epsilon_t \perp \mathcal{H}_{t-u}$ pour tout $u \geq 0$, $\epsilon_t \perp Y$ pour tout $Y \in \mathcal{H}_{-\infty}^X$, puisque, en particulier, $Y \in \mathcal{H}_{t-u}$. Nous avons $\epsilon_t \perp \mathcal{H}_{-\infty}^X$. Et donc $\epsilon_t \in \mathcal{H}_t^U$. Cela entraîne que $\sum_{k=1}^{\infty} \psi_k \epsilon_{t-k} \in \mathcal{H}_{t-1}^U$. Notons que $\sum_{k=1}^{\infty} \psi_k \epsilon_{t-k} = U_t - \epsilon_t$ ($\psi_0 = 1$). Par conséquent, pour tout $Y \in \mathcal{H}_{t-1}^U$ on a :

$$\left\langle U_t - \sum_{k=1}^{\infty} \psi_k \epsilon_{t-k}, Y \right\rangle = \langle \epsilon_t, Y \rangle = 0.$$

Cela implique que $\sum_{k=1}^{\infty} \psi_k \epsilon_{t-k}$ est la projection orthogonale de U_t sur \mathcal{H}_{t-1}^U et donc que :

$$\epsilon_t = U_t - \text{proj}(U_t | \mathcal{H}_{t-1}^U).$$

Cela signifie que $\{\epsilon_t, t \in \mathbb{Z}\}$ est le processus d'innovation du processus $\{U_t, t \in \mathbb{Z}\}$. Comme, par hypothèse, $\sigma^2 = \|\epsilon_t\|^2 \neq 0$, le processus $\{U_t, t \in \mathbb{Z}\}$ est donc régulier. Remarquons que, comme $\epsilon_t \in \mathcal{H}_t^U$, nous avons $\mathcal{H}_t^\epsilon \subset \mathcal{H}_t^U$. Comme, par construction, $\mathcal{H}_t^U \subset \mathcal{H}_t^\epsilon$, nous avons $\mathcal{H}_t^U = \mathcal{H}_t^\epsilon$.

(iv). Montrons tout d'abord que, pour tout t , on a :

$$\mathcal{H}_t^V = \overline{\text{Vect}(V_s, s \leq t)} = \mathcal{H}_{-\infty}^X \quad (5.43)$$

Pour tout t , $V_t \in \mathcal{H}_{-\infty}^X$ et donc $\mathcal{H}_t^V \subseteq \mathcal{H}_{-\infty}^X$. D'un autre côté, puisque $X_t = \sum_{k=0}^{+\infty} \psi_k \epsilon_{t-k} + V_t$, $\mathcal{H}_t^X = \mathcal{H}_t^\epsilon \oplus \mathcal{H}_t^V$. Et donc, quel que soit $Y \in \mathcal{H}_{-\infty}^X$, alors $Y \in \mathcal{H}_{s-1}^X$ pour tout s , de telle sorte que $(Y, \epsilon_s) = 0$ et donc $Y \in \mathcal{H}_t^V$, ce qui implique que $\mathcal{H}_{-\infty}^X \subseteq \mathcal{H}_t^V$. Ce qui démontre (5.43). Partant de (5.43), on déduit que $\text{proj}(V_t | \mathcal{H}_{t-1}^V) = \text{proj}(V_t | \mathcal{H}_{-\infty}^X) = \text{proj}(V_t | \mathcal{H}_t^V) = V_t$ et que $\|V_t - \text{proj}(V_t | \mathcal{H}_{t-1}^V)\|^2 = 0$: le processus $\{V_t, t \in \mathbb{Z}\}$ est donc déterministe. □

Deuxième partie

Estimation spectrale

Chapitre 6

Estimation de la moyenne et de la fonction d'autocovariance

6.1 Estimation de la moyenne

Soit $\{X_t\}$ un processus aléatoire à temps discret stationnaire au second ordre, de moyenne $\mathbb{E}[X_0] = \mu$, et de fonction d'autocovariance γ . On suppose avoir observé n échantillons consécutifs X_1, \dots, X_n du processus. L'estimateur de μ que nous considérons est la *moyenne empirique* définie par :

$$\hat{\mu}_n = \frac{1}{n} \sum_{t=1}^n X_t \quad (6.1)$$

On constate tout d'abord que $\hat{\mu}_n$ est un estimateur *sans biais* de la moyenne μ car

$$\mathbb{E}[\hat{\mu}_n] = \frac{1}{n} \sum_{t=1}^n \mathbb{E}[X_t] = \mu \quad (6.2)$$

du fait de la stationnarité. Le *risque quadratique* de l'estimateur, qui mesure sa dispersion autour de la valeur inconnue μ de la moyenne, a pour expression

$$\begin{aligned} R(\hat{\mu}_n, \mu) &= \mathbb{E}[(\hat{\mu}_n - \mu)^2] \\ &= \mathbb{E}\left[\frac{1}{n^2} \sum_{s=1}^n \sum_{t=1}^n (X_t - \mu)(X_s - \mu)\right] = \frac{1}{n^2} \sum_{s=1}^n \sum_{t=1}^n \gamma(t-s) = \frac{1}{n} \sum_{h=-n+1}^{n-1} \left(1 - \frac{|h|}{n}\right) \gamma(h) \end{aligned} \quad (6.3)$$

D'où la proposition suivante :

Proposition 6.1.1. *Soit $\{X_t\}$ un processus stationnaire au second ordre de moyenne μ et de fonction d'autocovariance $\gamma(h)$ avec $\sum |\gamma(h)| < \infty$. Alors, le risque quadratique de l'estimateur de la moyenne empirique $\hat{\mu}_n = n^{-1} \sum_{t=1}^n X_t$ vérifie*

$$\lim_{n \rightarrow \infty} n \mathbb{E}[(\hat{\mu}_n - \mu)^2] = 2\pi f(0) \quad \text{où} \quad f(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma(\tau) e^{-i\tau\lambda}. \quad (6.4)$$

c'est à dire que $\hat{\mu}_n$ converge en moyenne quadratique vers μ , à la vitesse \sqrt{n} . De plus $\lim_{n \rightarrow \infty} \hat{\mu}_n = \mu$ \mathbb{P} -p.s.

Démonstration. Lorsque $\gamma(h)$ est absolument sommable, le théorème de la convergence dominée appliquée à (6.3) montre que

$$\lim_{n \rightarrow \infty} nR(\hat{\mu}_n, \mu) = \sum_{h=-\infty}^{\infty} \lim_{n \rightarrow \infty} \left(1 - \frac{|h|}{n}\right) \gamma(h) = \sum_{h=-\infty}^{\infty} \gamma(h) = 2\pi f(0)$$

où $f(\lambda) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-ih\lambda}$ est la densité spectrale du processus $\{X_t\}$. La preuve de la convergence presque sûre de $\hat{\mu}_n$ est laissée à titre d'exercice. \square

Cette proposition montre que la loi des grands nombres, établie classiquement pour des variables aléatoires indépendantes, est également valable pour un processus stationnaire au second ordre, du moment que la fonction d'autocovariance décroît suffisamment rapidement à l'infini. Sous cette condition, il est possible d'estimer la moyenne à partir d'une seule réalisation de celui-ci. La proposition 6.1.1 nous donne accès à la valeur limite de $\mathbb{E}[(\sqrt{n}(\hat{\mu}_n - \mu))^2]$. Cependant pour construire des intervalles de confiance pour les paramètres estimés ou pour tester des hypothèses concernant la valeur des paramètres, il est nécessaire d'obtenir un résultat plus précis portant sur la distribution limite de $\sqrt{n}(\hat{\mu}_n - \mu)$. L'obtention de théorèmes de type limite centrale pour des suites de variables aléatoires dépendantes est un sujet délicat, qui a donné lieu à une vaste littérature. Il n'est bien entendu pas question ici de présenter une théorie générale et nous nous contentons donc d'énoncer un résultat valable dans le cas de processus linéaire fort.

Le fait de devoir émettre une hypothèse aussi contraignante sur la loi du processus dans un contexte où, en fait, seules les propriétés au second ordre nous intéressent est bien sûr frustrant, mais il traduit la difficulté technique d'un tel résultat (la preuve de ce théorème est donnée dans la Section 6.3).

Théorème 6.1.2. Soit $\{X_t\}$ un processus linéaire fort de moyenne μ . On a $X_t = \mu + \sum_{k=-\infty}^{\infty} \psi_k Z_{t-k}$ avec $\sum_k |\psi_k| < \infty$ et $Z_t \sim \text{IID}(0, \sigma^2)$. On pose $\hat{\mu}_n = n^{-1} \sum_{t=1}^n X_t$. Alors :

$$\sqrt{n}(\hat{\mu}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 2\pi f(0)) \quad (6.5)$$

où $f(0) = \sigma^2 |\hat{\psi}(0)|^2 / (2\pi)$, $\hat{\psi}(\lambda) = \sum_{j=-\infty}^{\infty} \psi_j e^{ij\lambda}$, est la densité spectrale de $\{X_t, t \in \mathbb{Z}\}$ à la fréquence nulle 0.

Exemple 6.1.3 (Moyenne empirique pour un processus AR(1) (fort)). Soit X_t un processus autorégressif d'ordre 1 fort, de moyenne μ , solution stationnaire au second ordre défini par l'équation de récurrence

$$X_t - \mu = \phi(X_{t-1} - \mu) + Z_t$$

où $\{Z_t\} \sim \text{IID}(0, \sigma^2)$ et $|\phi| < 1$. Nous rappelons que la fonction d'autocovariance d'un processus AR(1) pour $|\phi| < 1$ est donnée par

$$\gamma_X(k) = \frac{\sigma^2}{(1 - \phi^2)} \phi^{|k|}$$

et que la densité spectrale de ce processus a pour expression

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

Dans ce cas, la variance limite qui intervient dans l'équation (6.5), est donnée par $2\pi f(0) = \sigma^2/(1 - \phi)^2$. Cette valeur est à comparer avec la variance de X_t donnée par $\gamma(0) = \sigma^2/(1 - \phi^2)$. On constate que le rapport $2\pi f(0)/\gamma(0) = (1 + \phi)/(1 - \phi)$ tend vers 0 lorsque $\phi \rightarrow -1$ et vers $+\infty$ lorsque $\phi \rightarrow 1$. Ce qui implique par exemple lorsque l'on considère l'intervalle de confiance asymptotique de niveau 95% pour la moyenne μ donné par $[\hat{\mu}_n - 1.96\sigma n^{-1/2}/(1 - \phi), \hat{\mu}_n + 1.96\sigma n^{-1/2}/(1 - \phi)]$ que la longueur de l'intervalle de confiance est minimale lorsque $\phi = -1$ (corrélation négative). et maximale lorsque $\phi \rightarrow 1$ (corrélation positive).

6.2 Estimation des coefficients d'autocovariance et d'auto-corrélation

Considérons à nouveau un processus $\{X_t\}$ stationnaire au second ordre, de moyenne μ et de fonction d'autocovariance $\gamma(h)$ supposée de module sommable. Pour estimer la suite $\gamma(h)$, nous considérons les estimateurs, dits de covariances empiriques, définis par :

$$\hat{\gamma}_n(h) = \begin{cases} n^{-1} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \hat{\mu}_n)(X_t - \hat{\mu}_n) & \text{si } |h| \leq n-1 \\ 0 & \text{sinon} \end{cases} \quad (6.6)$$

où $\hat{\mu}_n = n^{-1} \sum_{t=1}^n X_t$. Remarquons que le nombre d'observations, dont nous disposons, étant précisément égal à n , il n'existe pas de paires d'observations séparées de plus de $n-1$ intervalles de temps et donc l'expression (6.6) ne permet pas d'estimer les valeurs de $\gamma(h)$ pour $|h| \geq n$. De plus, lorsque $|h|$ est proche de n , il est clair que l'estimateur (6.6) de la covariance n'est pas fiable, dans la mesure où on ne dispose que de peu de paires d'observations $(X_t, X_{t+|h|})$, ce qui implique que l'effet de moyennage statistique ne peut pas jouer. La partie la plus utile de la fonction d'autocovariance empirique est celle qui correspond aux valeurs du décalage h significativement plus faibles que le nombre d'observations n .

Certains auteurs remplacent n^{-1} par $(n - |h|)^{-1}$ dans l'expression (6.6). L'inconvénient de cette normalisation (qui peut sembler plus naturelle car on divise la somme par le nombre de termes intervenant dans celle-ci) est que la suite d'estimateurs des coefficients d'autocovariance ainsi définie n'est pas de type positif. Ajouté au fait que ces deux estimateurs sont asymptotiquement équivalents, l'utilisation de la normalisation $(n - |h|)^{-1}$ ne présente que peu d'intérêt pour les séries temporelles et n'est que très rarement utilisé en pratique.

Une propriété importante de cet estimateur est que la suite $\hat{\gamma}_n(h)$ est de type positif. En effet, si on définit le *périodogramme* par¹

$$I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^n (X_t - \hat{\mu}_n) e^{-it\lambda} \right|^2 \quad (6.7)$$

Par construction, $\lambda \mapsto I_n(\lambda)$ est une fonction positive pour $\lambda \in [-\pi, \pi]$. Par ailleurs,

$$\int_{-\pi}^{\pi} e^{i\lambda h} I_n(\lambda) d\lambda = \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n (X_t - \hat{\mu}_n)(X_s - \hat{\mu}_n) \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\lambda(h-t+s)} = \hat{\gamma}_n(h)$$

Par conséquent, d'après le théorème d'Herglotz 3.3.1, la suite $\hat{\gamma}_n(h)$ est de type positif.

1. Le périodogramme joue un rôle fondamental pour l'estimation de la densité spectrale étudiée dans le chapitre 7.

Proposition 6.2.1. Si $\hat{\gamma}_n(0) > 0$ alors, pour tout $p \leq n$, la matrice $\hat{\Gamma}_{n,p}$ définie par

$$\hat{\Gamma}_n = \begin{bmatrix} \hat{\gamma}_n(0) & \hat{\gamma}_n(1) & \cdots & \hat{\gamma}_n(p-1) \\ \hat{\gamma}_n(1) & \hat{\gamma}_n(0) & \cdots & \hat{\gamma}_n(p-2) \\ \vdots & & & \\ \hat{\gamma}_n(p-1) & \hat{\gamma}_n(p-2) & \cdots & \hat{\gamma}_n(0) \end{bmatrix} \quad (6.8)$$

est de rang plein et est donc inversible.

Démonstration. La suite $\hat{\gamma}_n(h)$ est de type positif, $\hat{\gamma}_n(0) > 0$ et $\hat{\gamma}_n(h)$ tend vers 0 quand n tend vers l'infini. On en déduit, d'après la propriété 3.3.9, que, pour tout p , la matrice est inversible. \square

L'estimateur d'autocovariance empirique est asymptotiquement sans biais et consistant, si nous supposons que le processus $\{X_t\}$ est linéaire au sens fort.

Théorème 6.2.2. Soit $\{X_t, t \in \mathbb{Z}\}$ un processus linéaire défini par $X_t = \mu + \sum_{s=-\infty}^{\infty} \psi_s Z_{t-s}$ avec $\sum_s |\psi_s| < \infty$. On suppose que $Z_t \sim \text{IID}(0, \sigma^2)$ vérifie $\mathbb{E}[Z_t^4] = \eta\sigma^4$. Alors, pour tout $(p, q) \in \mathbb{N} \times \mathbb{N}$,

$$\mathbb{E}[\hat{\gamma}_n(p)] = \gamma(p) + O(n^{-1}), \quad (6.9)$$

$$\lim_{n \rightarrow \infty} n \text{Cov}(\hat{\gamma}_n(p), \hat{\gamma}_n(q)) = V_{p,q}, \quad (6.10)$$

où

$$V_{p,q} \stackrel{\text{def}}{=} (\eta - 3)\gamma(p)\gamma(q) + \sum_{-\infty}^{\infty} [\gamma(u)\gamma(u-p+q) + \gamma(u+q)\gamma(u-p)]. \quad (6.11)$$

Les coefficients d'autocovariance empiriques interviennent quasiment dans tous les problèmes d'inférence statistique portant sur les processus stationnaires. A l'instar de la moyenne empirique, il est donc indispensable de disposer de résultats concernant leur distribution. Cependant, même pour les modèles de processus les plus simples, il est en général impossible de déterminer la distribution exacte de la suite des coefficients d'autocovariance empiriques $\hat{\gamma}_n(0), \dots, \hat{\gamma}_n(K)$ pour un nombre d'échantillons n donné. Nous ne considérons ici que des résultats asymptotiques concernant la distribution limite jointe des coefficients d'autocovariance empirique $[\hat{\gamma}_n(0), \dots, \hat{\gamma}_n(K)]$, pour un nombre de retard K fixé et lorsque le nombre d'échantillons n tend vers l'infini.

Théorème 6.2.3. Soit $\{X_t, t \in \mathbb{Z}\}$ un processus linéaire défini par $X_t = \mu + \sum_{s=-\infty}^{\infty} \psi_s Z_{t-s}$ avec $\sum_s |\psi_s| < \infty$. On suppose que $Z_t \sim \text{IID}(0, \sigma^2)$ vérifie $\mathbb{E}[Z_t^4] = \eta\sigma^4$. Pour $K \geq 1$, notons $\hat{\gamma}_n \stackrel{\text{def}}{=} [\hat{\gamma}_n(1), \dots, \hat{\gamma}_n(K)]^T$, $\gamma \stackrel{\text{def}}{=} [\gamma(1), \dots, \gamma(K)]^T$ et V la matrice de dimension $K \times K$ dont l'élément $V_{p,q}$ est donné par (6.11). Alors,

$$\sqrt{n}(\hat{\gamma}_n - \gamma) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V).$$

La preuve de ce résultat est donnée dans le paragraphe 6.4.

Il est aussi intéressant, et souvent plus pratique, de considérer la distribution limite des coefficients d'autocorrélations $[\hat{\rho}_n(0), \dots, \hat{\rho}_n(K)]$. On rappelle que les coefficients d'autocorrélation sont définis par

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

et qu'ils vérifient $|\rho(h)| \leq \rho(0) = 1$. On définit les coefficients d'autocorrélation empiriques par

$$\hat{\rho}_n(h) = \frac{\hat{\gamma}_n(h)}{\hat{\gamma}_n(0)} \quad (6.12)$$

où $\hat{\gamma}_n(h)$ est donné par (6.6).

Théorème 6.2.4. *Soit $\{X_t\}$ un processus linéaire défini par $X_t - \mu = \sum_{s=-\infty}^{\infty} \psi_s Z_{t-s}$ avec $\sum_s |\psi_s| < \infty$. On suppose que $Z_t \sim \text{IID}(0, \sigma^2)$ vérifie $\mathbb{E}[Z_t^4] < \infty$. Pour $K \geq 1$, on note $\hat{\boldsymbol{\rho}}_n = (\hat{\rho}_n(1), \dots, \hat{\rho}_n(K))^T$, $\boldsymbol{\rho} = (\rho(1), \dots, \rho(K))^T$ et $W = [w_{p,q}, p, q = 1, \dots, K]$ la matrice de dimension $K \times K$ définie, pour $1 \leq p, q \leq K$, par l'élément :*

$$w_{p,q} \stackrel{\text{def}}{=} \sum_{u=-\infty}^{\infty} \{\rho(u+p) + \rho(u-p) - 2\rho(u)\rho(p)\} \{\rho(u+q) + \rho(u-q) - 2\rho(u)\rho(q)\}. \quad (6.13)$$

Alors :

$$\sqrt{n}(\hat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}) \rightarrow_d \mathcal{N}(0, W) \quad (6.14)$$

Il est remarquable de noter que la distribution des coefficients d'autocorrélation ne dépend pas des moments du processus Z_t (on a uniquement supposé que $Z_t \sim \text{IID}(0, \sigma^2)$ avec un moment du 4ème ordre fini). Comme dans le cas du théorème 6.1.2, on constate qu'il est nécessaire d'admettre des hypothèses relativement fortes pour garantir ce résultat. La preuve est donnée dans le paragraphe 6.4.

Exemple 6.2.5 (Bruit blanc fort). *Soit $\{X_t\} \sim \text{IID}(0, \sigma^2)$. Dans ce cas $\rho(h) = 0$ pour tout $h \neq 0$ et la matrice de covariance asymptotique W est égale à la matrice identité. L'expression (6.14) montre que, lorsque la taille de l'échantillon n tend vers l'infini, le vecteur des coefficients d'autocorrélation empiriques multiplié par la racine carrée du nombre d'échantillons $\sqrt{n}[\hat{\rho}_n(1), \dots, \hat{\rho}_n(K)]$ converge vers un vecteur gaussien, centré et de covariance identité. On en déduit que, si $\{X_t\}$ est un bruit blanc fort, pour tout $h \neq 0$:*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[-1.96n^{-1/2} \leq \hat{\rho}_n(h) \leq 1.96n^{-1/2} \right] = 0.95 \quad (6.15)$$

Ce résultat peut être utilisé pour définir des tests asymptotiques de l'hypothèse nulle H_0 : " $\{X_t\}$ est un bruit blanc fort". Considérons en effet la procédure de test consistant à accepter l'hypothèse nulle si $\hat{\rho}_n(h)$ appartient à l'intervalle $[-1.96n^{-1/2}, 1.96n^{-1/2}]$ et à la rejeter sinon est un test de l'hypothèse nulle dont l'erreur de première espèce tend, lorsque $n \rightarrow \infty$, vers 0,05. Nous avons représenté figure 6.1 les 60 premiers coefficients d'autocorrélation empiriques d'un échantillon de taille $n = 500$, d'un bruit blanc fort, gaussien, centré, de variance $\sigma^2 = 1$. En utilisant la formule (6.15), nous avons représenté l'intervalle asymptotique $[-1.96n^{-1/2}, 1.96n^{-1/2}]$ autour de la vraie valeur $\rho(h) = 0$. Pour les valeurs des retards h pour lesquelles l'autocorrélation empirique appartient à l'intervalle $[-1.96n^{-1/2}, 1.96n^{-1/2}]$ le test d'hypothèse décrit ci-dessus est accepté.

Ce type de visualisation où l'on représente les coefficients d'autocorrélation empiriques ainsi que les extrémités des intervalles $[-1.96n^{-1/2}, 1.96n^{-1/2}]$ pour les estimateurs correspondants dans le cas du bruit blanc (fort) est classique dans le domaine des séries temporelles où il est désigné sous le nom de corrélogramme. Il permet de détecter visuellement les retards pour lesquels les coefficients de

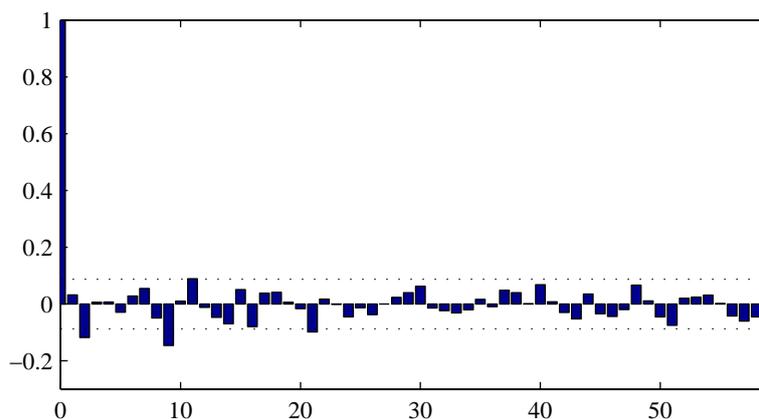


FIGURE 6.1 – Fonction d'autocorrélation empirique pour un échantillon de bruit blanc fort, gaussien, centré, de variance $\sigma^2 = 1$. Le nombre d'échantillons est égal à $n = 500$. Les droites en pointillé représentent les extrémités des intervalles $[-1.96n^{-1/2}, 1.96n^{-1/2}]$.

corrélation sont ou ne sont pas compatibles avec l'hypothèse de bruit blanc fort (comme dans le cas de la figure 6.2 par exemple).

Il faut toutefois faire attention quand on considère simultanément plusieurs valeurs de retards, car si le risque de première espèce de chaque test individuel est asymptotique égal à 0,05, il est difficile d'évaluer le risque d'un test consistant à prendre en considération simultanément un ensemble de valeurs de retards, car nous ne prenons pas en compte la dépendance entre les différents tests.

Nous pouvons toutefois déduire du théorème précédent un test de l'hypothèse nulle que le processus soit un bruit blanc en prenant en compte simultanément plusieurs valeurs de retard. En effet, le théorème 6.2.4 montre que, sous l'hypothèse que $\{X_t, t \in \mathbb{Z}\}$ est un bruit blanc fort, la statistique $T_n = \sum_{l=1}^K \hat{\rho}_n(l)^2$ est distribuée suivant une loi du χ^2 centré à K degrés de liberté. Le test consistant à accepter l'hypothèse nulle si la valeur prise par la statistique T_n est inférieure au quantile à 95% de la loi du χ^2 centré à K degrés de liberté, a une erreur de première espèce asymptotique égale à 0,05.

Exemple 6.2.6 (Processus MA(1)). On considère le processus MA(1) défini par $X_t = Z_t + \theta_1 Z_{t-1}$ où Z_t est un bruit blanc fort, centré, de variance σ^2 . Ici, la suite des coefficients d'autocorrélation est donnée par :

$$\rho(h) = \begin{cases} 1 & \text{pour } h = 0 \\ \frac{\theta_1}{1 + \theta_1^2} & \text{pour } |h| = 1 \\ 0 & \text{pour } |h| \geq 2 \end{cases}$$

On en déduit, d'après (6.13), que les éléments diagonaux de la matrice de covariance de la distribution limite des coefficients d'autocovariance empiriques ont pour expression :

$$W_{h,h} = \begin{cases} 1 - 3\rho^2(1) + 4\rho^4(1) & \text{pour } |h| = 1 \\ 1 + 2\rho(1)^2 & \text{pour } |h| \geq 2 \end{cases}$$

Par conséquent les intervalles de confiance de taux de couverture 95% pour les coefficients d'autocorrélation empiriques sont donnés, pour $h = 1$, par :

$$\hat{\rho}_n(1) \in \left[\rho(1) - 1.96W_{1,1}^{1/2} n^{-1/2} \quad \rho(1) + 1.96W_{1,1}^{1/2} n^{-1/2} \right]$$

et, pour $h \geq 2$, par :

$$\hat{\rho}_n(h) \in \left[-1.96W_{2,2}^{1/2} n^{-1/2} \quad + 1.96W_{2,2}^{1/2} n^{-1/2} \right]$$

Notons ici que ces régions dépendent, par l'intermédiaire de $\rho(1)$, de la quantité a priori inconnue θ_1 . Nous avons représenté figure 6.2 les 60 premiers coefficients d'autocorrélation empiriques d'un échantillon de longueur $n = 500$ d'un processus $MA(1)$ défini par $\theta_1 = -0.8$ et $\sigma = 1$. Les traits en pointillé représentent les bornes asymptotiques autour des vraies valeurs au niveau 95%.

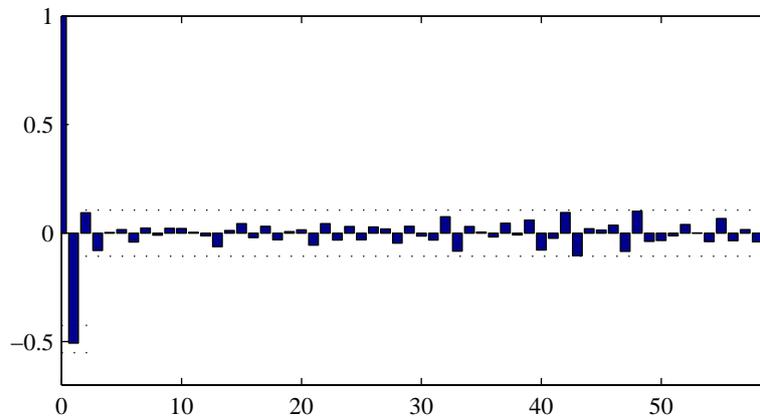


FIGURE 6.2 – Fonction d'autocorrélation empirique d'un échantillon de longueur $n = 500$ d'un processus $MA(1)$ pour $\theta_1 = -0.8$ et donc $\rho(1) = -0.4878$.

Exemple 6.2.7 (Processus autorégressif fort d'ordre 1). On considère le processus aléatoire X_t défini par :

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

où $\{Z_t\} \sim \text{IID}(0, \sigma^2)$ et où $|\phi| < 1$. La fonction d'autocorrélation d'un tel processus est donnée par $\rho(h) = \phi^{|h|}$ et les éléments diagonaux de la matrice de covariance W sont donnés par

$$\begin{aligned} W_{h,h} &= \sum_{m=1}^h \phi^{2h} (\phi^{-m} - \phi^m)^2 + \sum_{m=h+1}^{\infty} \phi^{2m} (\phi^{-i} - \phi^i)^2 \\ &= (1 - \phi^{2h})(1 + \phi^2)(1 - \phi^2)^{-1} - 2h\phi^{2h} \end{aligned}$$

Considérons la séquence, de longueur $n = 1800$, des battements cardiaques représentés figure 1.1 (chapitre 1). La figure 3.2 qui représente les couples (X_t, X_{t-1}) suggère fortement la présence d'une relation linéaire entre les variables X_t et X_{t-1} et invite donc à tester la validité d'un modèle autorégressif d'ordre 1. Pour estimer le paramètre ϕ du modèle autorégressif, une méthode naturelle, compte tenu de

l'allure de la fonction d'autocorrélation de l'AR(1), consiste à utiliser comme estimateur $\hat{\phi}_n = \hat{\rho}_n(1)$ qui donne $\hat{\phi}_n = 0.966$. Pour tester la validité du modèle, deux solutions s'offrent à nous : (i) tester que les résidus de prédiction donnés par $\hat{Z}_t = X_t - \hat{\mu}_n - \hat{\phi}_n(X_{t-1} - \hat{\mu}_n)$ sont compatibles avec un modèle de bruit blanc, (ii) vérifier directement que les coefficients d'autocorrélation empiriques sont compatibles avec ceux d'un modèle AR(1). Les résidus de prédiction sont reportés figure 6.3 et la fonction d'autocorrélation de ces résidus figure 6.4, où nous avons aussi indiqué les bornes de la zone crédible à 95% pour le bruit blanc avec un nombre d'observations $n = 1800$. Les corrélations empiriques, en particulier pour $h = 2$, sont significativement à l'extérieur des intervalles de confiance du bruit blanc, ce qui conduit à rejeter le modèle de bruit blanc pour les résidus et donc le modèle autorégressif d'ordre 1 pour les observations. Les résultats de l'analyse de la suite des coefficients d'autocorrélation empiriques du processus et des zones crédibles à 95% sous l'hypothèse d'un modèle AR(1) avec $\phi = 0.966$ sont reportés figure 6.5. On observe que les premières valeurs des coefficients de corrélation sont nettement à l'extérieur de cette zone, ce qui contribue ici encore à rejeter le modèle AR(1).

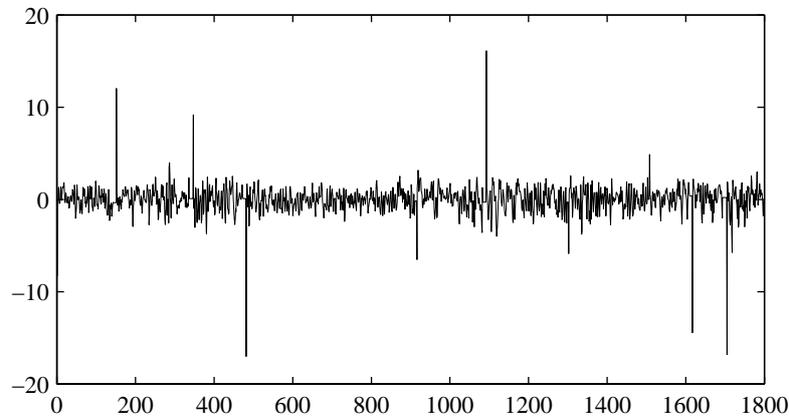


FIGURE 6.3 – Série des battements cardiaques : Résidu de prédiction $\hat{Z}_t = (X_t - \hat{\mu}_n) - \hat{\phi}_n(X_{t-1} - \hat{\mu}_n)$.

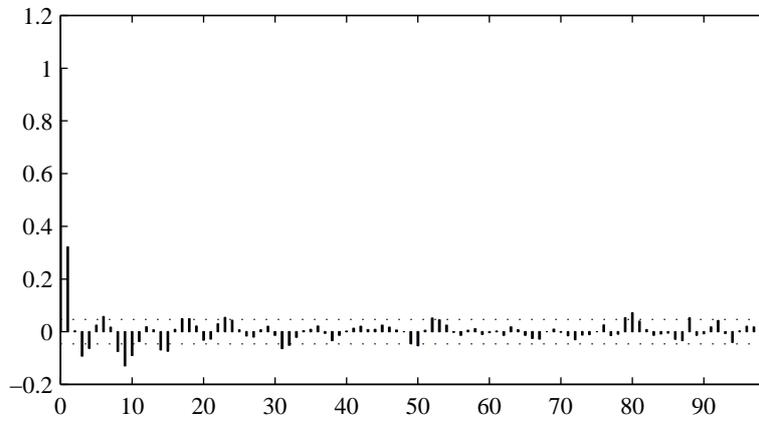


FIGURE 6.4 – Série des battements cardiaques : coefficients d'autocorrélation empiriques des résidus de prédiction $\hat{Z}_t = (X_t - \hat{\mu}_n) - \hat{\phi}_n(X_{t-1} - \hat{\mu}_n)$ et zones crédibles à 95% pour le bruit blanc ($n = 1800$).

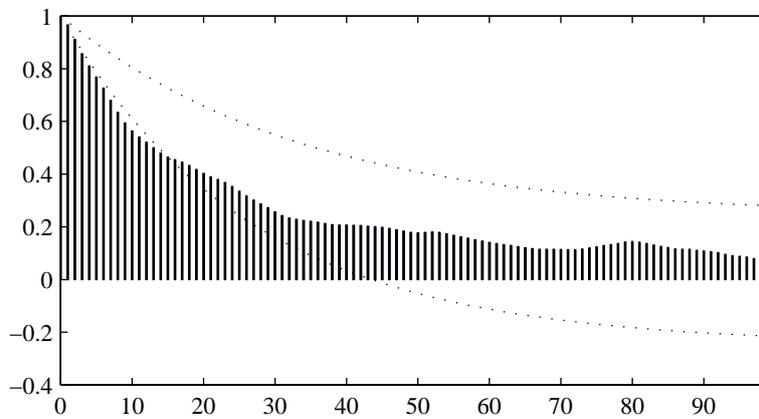


FIGURE 6.5 – Série des battements cardiaques : coefficients d'autocorrélation empiriques de la série et bornes des zones crédibles à 95% pour un modèle $AR(1)$ de paramètre $\phi = 0.966$.

6.3 Preuve du Théorème 6.1.2

Les preuves sont basées sur le théorème d'approximation suivant

Théorème 6.3.1. Soient $(A_n, n \geq 0)$, $(B_m, m \geq 0)$, $(B_{m,n}, m \geq 0, n \geq 0)$ et B des vecteurs aléatoires vérifiant

(i) Pour tout m , $B_{m,n} \xrightarrow{\mathcal{L}} B_m$ quand $n \rightarrow \infty$,

(ii) $B_m \xrightarrow{\mathcal{L}} B$ pour $m \rightarrow \infty$,

(iii) Pour tout $\epsilon > 0$, $\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(|A_n - B_{m,n}| \geq \epsilon) = 0$.

Alors $A_n \xrightarrow{\mathcal{L}} B$ quand $n \rightarrow \infty$.

Remarque 6.3.2. Pour établir la condition (iii) il est souvent pratique d'utiliser l'inégalité de Markov.

Démonstration. En appliquant le théorème de Levy (voir théorème B.3.3), il suffit de montrer que, pour tout λ , $\phi_{A_n}(\lambda) \rightarrow \phi_B(\lambda)$, où $\phi_Z(\lambda)$ est la fonction caractéristique du vecteur aléatoire Z . L'inégalité triangulaire montre que

$$|\phi_{A_n}(\lambda) - \phi_B(\lambda)| \leq |\phi_{A_n}(\lambda) - \phi_{B_{m,n}}(\lambda)| + |\phi_{B_{m,n}}(\lambda) - \phi_{B_m}(\lambda)| + |\phi_{B_m}(\lambda) - \phi_B(\lambda)|.$$

Considérons tout d'abord le premier terme.

$$\begin{aligned} |\phi_{A_n}(\lambda) - \phi_{B_{m,n}}(\lambda)| &= \left| \mathbb{E} \left[e^{i\lambda^t A_n} \right] - \mathbb{E} \left[e^{i\lambda^t B_{m,n}} \right] \right| \\ &\leq \mathbb{E} \left[\left| e^{i\lambda^t A_n} \left(1 - e^{i(\lambda^t B_{m,n} - \lambda^t A_n)} \right) \right| \right] \\ &\leq \mathbb{E} \left[\left| 1 - e^{i(\lambda^t B_{m,n} - \lambda^t A_n)} \right| \right]. \end{aligned}$$

Pour $\delta > 0$, posons $A_{m,n}(\delta) \stackrel{\text{def}}{=} \{|\lambda^t B_{m,n} - \lambda^t A_n| \geq \delta\}$. Nous avons donc :

$$|\phi_{A_n}(\lambda) - \phi_{B_{m,n}}(\lambda)| \leq \mathbb{E} \left[\left| 1 - e^{i(\lambda^t B_{m,n} - \lambda^t A_n)} \right| \mathbb{1}_{A_{m,n}(\delta)} \right] + \mathbb{E} \left[\left| 1 - e^{i(\lambda^t B_{m,n} - \lambda^t A_n)} \right| \mathbb{1}_{A_{m,n}^c(\delta)} \right].$$

Pour λ et $\epsilon > 0$, nous choisissons $\delta(\epsilon)$ tel que $\left| 1 - e^{i(\lambda^t B_{m,n} - \lambda^t A_n)} \right| \mathbb{1}_{A_{m,n}^c(\delta(\epsilon))} < \epsilon$. Comme $\left| 1 - e^{i(\lambda^t B_{m,n} - \lambda^t A_n)} \right| \leq 2$, l'inégalité précédente implique

$$|\phi_{A_n}(\lambda) - \phi_{B_{m,n}}(\lambda)| \leq 2\mathbb{P}[A_{m,n}(\delta(\epsilon))] + \epsilon.$$

Comme $\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}[A_{m,n}(\delta(\epsilon))] = 0$, nous pouvons choisir un entier $m(\epsilon)$ tel que, pour tout $m \geq m(\epsilon)$ il existe un entier $n(m, \epsilon)$ tel que

$$\mathbb{P}[A_{m,n}(\delta(\epsilon))] \leq \epsilon, \quad \text{pour tout } n \geq n(m, \epsilon).$$

La condition (ii) montre que l'on peut choisir $m \geq m(\epsilon)$ assez grand pour que $|\phi_{B_m}(\lambda) - \phi_B(\lambda)| \geq \epsilon$. La condition (i) montre que l'on peut choisir $n \geq n(m, \epsilon)$ suffisamment grand pour que $|\phi_{B_{m,n}}(\lambda) - \phi_{B_m}(\lambda)| < \epsilon$, ce qui conclut la preuve du théorème. \square

Nous allons maintenant étendre le théorème de la Limite Centrale dans un cadre dépendant. Nous nous allons tout d'abord établir ce théorème pour des suites m -dépendantes, puis nous étendrons ce résultat aux processus linéaires au sens fort.

Définition 6.3.3 (Processus M -dépendant). *Nous dirons que le processus $\{X_k, k \in \mathbb{Z}\}$ est M -dépendant, si pour tout $n \in \mathbb{Z}$, les tribus $\mathcal{P}_n \stackrel{\text{def}}{=} \sigma(X_k, k \leq n)$ et $\mathcal{F}_{n+m+1} \stackrel{\text{def}}{=} \sigma(X_k, k \geq n+m+1)$ sont indépendantes.*

Théorème 6.3.4. *Soit $\{X_k, k \in \mathbb{Z}\}$ un processus stationnaire au sens strict M -dépendant. Nous supposons que $\mathbb{E}[X_0^2] < \infty$ et nous notons par $\gamma(\cdot)$ la fonction d'autocovariance du processus. Alors,*

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V_M),$$

où $\mu = \mathbb{E}[X_0]$ et $V_M = \sum_{h=-M}^M \gamma(h)$.

Démonstration. Sans perte de généralité, nous supposons que $\mu = 0$. Nous allons utiliser le théorème d'approximation 6.3.1 en construisant un tableau de variables aléatoires $\{B_{m,n}, (m,n) \in \mathbb{N} \times \mathbb{N}\}$ approchant $A_n \stackrel{\text{def}}{=} n^{1/2} \bar{X}_n$. Pour $m \geq 2M$, considérons

$$B_{m,n} = n^{-1/2} [(X_1 + \cdots + X_{m-M}) + (X_{m+1} + \cdots + X_{2m-M}) + (X_{(r_n-1)m+1} + \cdots + X_{r_n m-M})]$$

où $r_n = \lfloor n/m \rfloor$. Cette approximation contient une partie des termes de \bar{X}_n , mais les variables aléatoires $Z_{m,k} = (X_{(k-1)m+1} + \cdots + X_{km-M})$, $k \in \{1, \dots, r\}$ sont indépendantes. Comme le processus est stationnaire au sens strict, les vecteurs aléatoires $\{Z_{m,1}, \dots, Z_{m,r}\}$ sont indépendants et identiquement distribués de moyenne nulle et de variance

$$S_{m-M} \stackrel{\text{def}}{=} \sum_{|h| \leq M} (m - M - |h|) \gamma(h), m \geq 2M. \quad (6.16)$$

Nous allons vérifier que cette approximation satisfait les conditions du théorème 6.3.1.

- (i) En appliquant le théorème de Limite Centrale aux vecteurs aléatoires $\{Z_{m,k}, k \geq 0\}$, nous obtenons

$$B_{m,n} = n^{-1/2} \sum_{k=1}^{r_n} Z_{m,k} = (n/r_n)^{-1/2} r_n^{-1/2} \sum_{k=1}^{r_n} Z_{m,k},$$

et en utilisant $\lim_{n \rightarrow \infty} (n/r_n)^{-1/2} = m^{1/2}$, nous obtenons $B_{m,n} \xrightarrow{\mathcal{L}} B_m$ lorsque $n \rightarrow \infty$, où B_m est une variable aléatoire Gaussienne de moyenne nulle et de variance S_{m-M}/m .

- (ii) Comme $\lim_{m \rightarrow \infty} S_{m-M}/m = V_M$, nous avons $B_m \xrightarrow{\mathcal{L}} B$, où B est une variable aléatoire gaussienne de moyenne nulle et de variance V_M . En effet, la fonction caractéristique est donnée par $\mathbb{E}[e^{i\lambda B_m}] = e^{-\lambda^2 S_{m-M}/(2m)}$ et, donc, pour tout $\lambda \in \mathbb{R}$, $\lim_{m \rightarrow \infty} \mathbb{E}[e^{i\lambda B_m}] = e^{-\lambda^2 V_M/2}$. Notons que $\lambda \mapsto e^{-\lambda^2 V_M/2}$ est la fonction caractéristique d'une variable aléatoire gaussienne centrée de variance V_M . le résultat est donc une conséquence de la caractérisation du théorème Levy (voir Théorème B.3.3).

(iii) Pour vérifier la dernière condition, considérons la différence

$$n^{1/2}\bar{X}_n - B_{m,n} = n^{-1/2}(W_{m,1} + \cdots + W_{m,r_n}),$$

où $W_{m,k} = X_{km-M+1} + \cdots + X_{km}$ pour $1 \leq k \leq r_n - 1$ et $W_{m,r_n} = X_{r_nm-M+1} + \cdots + X_n$. Les variables aléatoires $W_{m,k}$, $1 \leq k \leq r_n$ sont indépendantes. La variance des $r_n - 1$ premières variables est égale à S_M où S_M est défini par (6.16). La variance de W_{m,r_n} est donnée par

$$\text{Var}(W_{m,r_n}) = \sum_{|u| \leq m-M} (n - \lfloor n/m \rfloor m + M - |u|) \gamma(u) \leq \sum_{|u| \leq m-M} (m + M - |u|) \gamma(u).$$

En utilisant

$$\text{Var}\left(n^{1/2}\bar{X}_n - B_{m,n}\right) = n^{-1}[(r_n - 1)S_M + \text{Var}(W_{m,r_n})],$$

nous avons

$$\limsup_{n \rightarrow \infty} \text{Var}\left(n^{1/2}\bar{X}_n - B_{m,n}\right) = m^{-1}S_M,$$

et donc

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \text{Var}\left(n^{-1}\bar{X}_n - B_{m,n}\right) = 0.$$

□

En utilisant le résultat précédent, nous allons établir le théorème Central Limite pour la moyenne empirique d'un processus linéaire au sens fort

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad (6.17)$$

où $\{Z_t, t \in \mathbb{Z}\}$ sont des variables aléatoires indépendantes et identiquement distribuées (bruit blanc fort), de moyenne nulle et de variance σ^2 et

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty. \quad (6.18)$$

Remarquons tout d'abord que si le processus $\{Z_t, t \in \mathbb{Z}\}$ est un bruit blanc fort gaussien, la variable \bar{X}_n est elle aussi gaussienne, de moyenne nulle et de variance

$$\text{Var}(\bar{X}_n) = n^{-1} \sum_{|u| < n} \left(1 - \frac{|u|}{n}\right) \gamma(u),$$

et comme

$$\lim_{n \rightarrow \infty} \sum_{|u| < n} \left(1 - \frac{|u|}{n}\right) \gamma(u) = \sum_{u=-\infty}^{\infty} \gamma(u) = \sigma^2 \left(\sum_{j=-\infty}^{\infty} \psi_j \right)^2 \stackrel{\text{def}}{=} V, \quad (6.19)$$

nous avons donc $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V)$. Nous allons montrer que ce résultat reste vrai de façon générale.

Théorème 6.3.5. Soit $\{X_t, t \in \mathbb{Z}\}$ un processus linéaire fort (6.21) tel que

1. $\{Z_t, t \in \mathbb{Z}\}$ est un bruit blanc fort de moyenne nulle et de variance σ^2
2. $\sum_{j \in \mathbb{Z}} |\psi_j| < \infty$.

Alors

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V), \quad V = \sigma^2 \left(\sum_{j=-\infty}^{\infty} \psi_j \right)^2.$$

Démonstration. Sans perte de généralité, nous supposons que $\mu = 0$. Pour établir ce résultat, nous allons utiliser encore le théorème d'approximation 6.3.1. L'idée est d'approcher le processus $\{X_t, t \in \mathbb{Z}\}$, par une suite de processus $(2m+1)$ -dépendant, $X_t^m = \sum_{i=-m}^m \psi_i Z_{t-i}$ et nous construisons les variables :

$$Y_{m,n} = n^{1/2} \sum_{t=1}^n X_t^m.$$

1. En appliquant le théorème 6.3.4, nous avons, pour tout m , $Y_{m,n} \xrightarrow{\mathcal{L}} Y_m$ quand $n \rightarrow \infty$, où Y_m est une variable aléatoire gaussienne de moyenne nulle et de variance V_m , avec

$$V_m = \sum_{h=-2m}^{2m} \gamma^m(h) = \sigma^2 \left(\sum_{j=-m}^m \psi_j \right)^2. \quad (6.20)$$

2. Comme $V_m \rightarrow V$ quand $m \rightarrow \infty$, nous avons, $Y_m \xrightarrow{\mathcal{L}} Y$, où Y est une variable aléatoire gaussienne de moyenne nulle et de variance $V = \sigma^2 \left(\sum_{j=-\infty}^{\infty} \psi_j \right)^2$.
3. Finalement,

$$\begin{aligned} \text{Var} \left(n^{1/2} \bar{X}_n - Y_{m,n} \right) &= n \text{Var} \left(n^{-1} \sum_{t=1}^n \sum_{|j|>m} \psi_j Z_{t-j} \right) \\ &= \sigma^2 \left(\sum_{|j|>m} \psi_j \right)^2 \end{aligned}$$

qui converge vers 0 quand $m \rightarrow \infty$.

□

6.4 Preuves des Théorèmes 6.2.4 et 6.2.3

Nous allons maintenant étudier les distributions asymptotique de la fonction d'autocovariance et d'autocorrélation. Dans toute cette partie, nous supposons que $\{X_t, t \in \mathbb{Z}\}$ est un processus linéaire défini par

$$X_t = \mu + \sum_{s=-\infty}^{\infty} \psi_s Z_{t-s}, \quad \sum_s |\psi_s| < \infty, \quad (6.21)$$

où

$$Z_t \sim \text{IID}(0, \sigma^2) \text{ vérifie } \mathbb{E}[Z_t^4] = \eta\sigma^4. \quad (6.22)$$

Sans perte de généralité, nous posons $\mu = 0$. Nous notons $\{\gamma(h), h \in \mathbb{Z}\}$ la fonction d'autocovariance de ce processus,

$$\gamma(h) = \sigma^2 \sum_{i=-\infty}^{\infty} \psi_i \psi_{i+h}, \quad h \in \mathbb{Z}. \quad (6.23)$$

Considérons, pour $h \geq 0$,

$$\tilde{\gamma}_n(h) = n^{-1} \sum_{t=1}^n (X_{t+h} - \mu)(X_t - \mu). \quad (6.24)$$

Cette quantité est plus facile à étudier que l'autocovariance empirique $\hat{\gamma}_n(h)$. De plus, $\tilde{\gamma}_n(h)$ est une *bonne* approximation de l'autocovariance empirique comme le démontre la Proposition 6.4.2. Nous aurons besoin dans la suite du résultat suivant :

Lemme 6.4.1. *Supposons que $\{X_t\}$ satisfait (6.21) et (6.22). Alors, pour tout $(k, l, p, q) \in \mathbb{Z}^4$*

$$\begin{aligned} \mathbb{E}[X_k X_\ell X_p X_q] &= (\eta - 3)\sigma^4 \sum_{i=-\infty}^{\infty} \psi_{k+i} \psi_{\ell+i} \psi_{p+i} \psi_{q+i} \\ &\quad + \sigma^4 \{ \gamma(k - \ell)\gamma(p - q) + \gamma(k - p)\gamma(\ell - q) + \gamma(k - q)\gamma(\ell - p) \}. \end{aligned} \quad (6.25)$$

De plus, il existe une constante C telle que, pour tout $m \in \mathbb{N}$,

$$\mathbb{E} \left[\left(\sum_{t=1}^m X_t \right)^4 \right] \leq Cm^2. \quad (6.26)$$

Démonstration. Ce calcul est élémentaire mais un peu compliqué ; nous n'en donnons que les éléments essentiels en laissant les détails au lecteur. Nous évaluons ensuite $\mathbb{E}[Z_i Z_j Z_k Z_\ell]$. Un calcul élémentaire montre que

$$\mathbb{E}[Z_i Z_j Z_k Z_\ell] = \begin{cases} \eta\sigma^4 & \text{si } i = j = k = \ell \\ \sigma^4 & \text{si } i = j \neq k = \ell \text{ ou } i = k \neq j = \ell \text{ ou } i = \ell \neq j = k \\ 0 & \text{sinon} \end{cases} \quad (6.27)$$

Le résultat (6.25) en découle.

$$\mathbb{E} \left[\left(\sum_{t=1}^m X_t \right)^4 \right] = \sum_{k, \ell, p, q=1}^m \mathbb{E}[X_k X_\ell X_p X_q] = A + B$$

avec

$$\begin{aligned} A &= (\eta - 3)\sigma^4 \sum_{k, \ell, p, q=1}^m \sum_{i=-\infty}^{\infty} \psi_{k+i} \psi_{\ell+i} \psi_{p+i} \psi_{q+i} \\ B &= \sigma^4 \sum_{k, \ell, p, q=1}^m \{ \gamma(k - \ell)\gamma(p - q) + \gamma(k - p)\gamma(\ell - q) + \gamma(k - q)\gamma(\ell - p) \}. \end{aligned}$$

Notons tout d'abord que

$$\sum_{k,\ell,p,q=1}^m \left| \sum_{i=-\infty}^{\infty} \psi_{k+i} \psi_{\ell+i} \psi_{p+i} \psi_{q+i} \right| \leq \sum_{k=1}^n \sum_{\ell,p,q=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} |\psi_{k+i} \psi_{\ell+i} \psi_{p+i} \psi_{q+i}|$$

$$\leq \sum_{k=1}^m \sum_{i=-\infty}^{\infty} |\psi_{k+i}| \left(\sum_{j=-\infty}^{\infty} |\psi_j| \right)^3,$$

ce qui montre que

$$A \leq m \left(\sum_{j=-\infty}^{\infty} |\psi_j| \right)^4.$$

Remarquons que d'autre part que

$$\sum_{k,\ell,p,q=1}^m |\gamma(k-\ell)\gamma(p-q) + \gamma(k-p)\gamma(\ell-q) + \gamma(k-q)\gamma(\ell-p)| \leq 3 \left(\sum_{h=-m+1}^{m-1} (m-|h|)|\gamma(h)| \right)^2.$$

Par conséquent

$$B \leq 3m^2 \left(\sum_{h=-\infty}^{\infty} |\gamma(h)| \right)^2.$$

□

Proposition 6.4.2. *Supposons que $\{X_t\}$ satisfait (6.21) et (6.22). Pour tout $h \geq 0$,*

$$\mathbb{E} \left[(\hat{\gamma}_n(h) - \tilde{\gamma}_n(h))^2 \right] = O(n^{-2}). \quad (6.28)$$

De plus, pour tout $h \in \mathbb{Z}$

$$\mathbb{E} [\hat{\gamma}_n(h)] = \mathbb{E} [\tilde{\gamma}_n(h)] + O(n^{-1}), \quad (6.29)$$

et pour tout $(p, q) \in \mathbb{Z} \times \mathbb{Z}$,

$$\text{Cov}(\hat{\gamma}_n(p), \hat{\gamma}_n(q)) = \text{Cov}(\tilde{\gamma}_n(p), \tilde{\gamma}_n(q)) + O(n^{-1}). \quad (6.30)$$

Démonstration. Sans perte de généralité, nous posons $\mu = 0$. Un calcul élémentaire montre que

$$\hat{\gamma}_n(h) = \frac{1}{n} \sum_{t=1}^{n-h} X_{t+h} X_t + \frac{n-h}{n} \bar{X}_n^2 - \frac{1}{n} \bar{X}_n \sum_{t=1}^{n-h} \{X_t + X_{t+h}\}.$$

Nous avons d'autre part

$$n^{-1} \sum_{t=1}^{n-h} \{X_t + X_{t+h}\} = 2\bar{X}_n - n^{-1} \sum_{t=1}^h X_t - n^{-1} \sum_{t=n-h+1}^n X_t.$$

Par conséquent

$$\hat{\gamma}_n(h) = \tilde{\gamma}_n(h) - \left(1 + \frac{h}{n}\right) \bar{X}_n^2 - \bar{X}_n \left[n^{-1} \sum_{t=1}^h X_t + n^{-1} \sum_{t=n-h+1}^n X_t \right].$$

En utilisant le Lemme 6.4.1, nous avons $\mathbb{E}[\bar{X}_n^4] \leq Cn^{-2}$ et

$$\mathbb{E} \left[\left(\bar{X}_n n^{-1} \sum_{t=1}^h X_t \right)^2 \right] \leq n^{-2} (\mathbb{E}[\bar{X}_n^4])^{1/2} \left(\mathbb{E} \left[\left(\sum_{t=1}^h X_t \right)^4 \right] \right)^{1/2} \leq Cn^{-3}h.$$

De même,

$$\mathbb{E} \left[\left(\bar{X}_n n^{-1} \sum_{t=n-h+1}^n X_t \right)^2 \right] \leq Cn^{-3}h,$$

ce qui conclut la preuve de (6.28). L'équation (6.29) découle de

$$|\mathbb{E}[\hat{\gamma}_n(h)] - \mathbb{E}[\tilde{\gamma}_n(h)]| \leq \mathbb{E}[|\hat{\gamma}_n(h) - \tilde{\gamma}_n(h)|] \leq (\mathbb{E}[|\hat{\gamma}_n(h) - \tilde{\gamma}_n(h)|^2])^{1/2}.$$

L'équation (6.30) s'obtient de façon similaire. \square

Nous allons tout d'abord calculer la variance et la variance asymptotique de $\tilde{\gamma}_n(h)$ lorsque $\{X_t, t \in \mathbb{Z}\}$ est un processus linéaire fort vérifiant (6.21) et (6.22).

Proposition 6.4.3. *Soit $\{X_t, t \in \mathbb{Z}\}$ un processus linéaire vérifiant (6.21) et (6.22). Alors :*

$$\lim_{n \rightarrow \infty} n \text{Cov}(\tilde{\gamma}_n(p), \tilde{\gamma}_n(q)) = (\eta - 3)\gamma(p)\gamma(q) + \sum_{u=-\infty}^{\infty} [\gamma(u)\gamma(u+p-q) + \gamma(u+p)\gamma(u-q)]. \quad (6.31)$$

Démonstration. Sans perte de généralité, nous posons ici $\mu = 0$. Notons que $\mathbb{E}[\tilde{\gamma}_n(h)] = \gamma(h)$. Nous allons vérifier tout d'abord que pour tout $p, q \geq 0$,

$$\text{Cov}(\tilde{\gamma}_n(p), \tilde{\gamma}_n(q)) = n^{-1} \sum_{u=-(n-1)}^{(n-1)} \left(1 - \frac{|u|}{n}\right) V_u, \quad (6.32)$$

où

$$V_u \stackrel{\text{def}}{=} \gamma(u)\gamma(u+p-q) + \gamma(u+p)\gamma(u-q) + (\eta - 3)\sigma^4 \sum_i \psi_{i+u+q}\psi_{i+u}\psi_{i+p}\psi_i. \quad (6.33)$$

L'absolue sommabilité de $\{\psi_t, t \in \mathbb{Z}\}$ implique que $\sum_{u \in \mathbb{Z}} |V_u| < \infty$ et le théorème de convergence dominé implique (6.31). Notons tout d'abord que :

$$\mathbb{E}[\tilde{\gamma}_n(p)\tilde{\gamma}_n(q)] = n^{-2} \sum_{s,t} \sum_{i,j,k,\ell} \psi_{s+p-i}\psi_{s-j}\psi_{t+q-k}\psi_{t-\ell} \mathbb{E}[Z_i Z_j Z_k Z_\ell].$$

La formule s'obtient ensuite directement en utilisant (6.27). \square

Nous disposons maintenant de l'ensemble des résultats permettant d'établir le Théorème

Preuve du Théorème 6.4. Le théorème découle de (6.29), (6.30) et de la Proposition 6.4.3. \square

La proposition 6.4.2 montre que

$$n^{1/2} (\hat{\gamma}_n(h) - \tilde{\gamma}_n(h)) = o_P(1).$$

Par conséquent, les distributions limites de $\tilde{\gamma}_n(h)$ et de $\hat{\gamma}_n(h)$ coïncident. En utilisant la proposition précédente, nous pouvons obtenir un premier résultat sur la distribution asymptotique des coefficients d'autocovariance empirique :

Théorème 6.4.4. *Soit $\{X_t, t \in \mathbb{Z}\}$ un processus linéaire vérifiant (6.21) et (6.22). Pour $K \geq 1$, notons $\hat{\gamma}_n \stackrel{\text{def}}{=} [\hat{\gamma}_n(1), \dots, \hat{\gamma}_n(K)]^T$, $\gamma \stackrel{\text{def}}{=} [\gamma(1), \dots, \gamma(K)]^T$ et V la matrice de dimension $K \times K$ dont l'élément $V_{p,q}$ est donné par*

$$V_{p,q} \stackrel{\text{def}}{=} (\eta - 3)\gamma(p)\gamma(q) + \sum_{u=-\infty}^{\infty} [\gamma(u)\gamma(u-p+q) + \gamma(u+q)\gamma(u-p)]. \quad (6.34)$$

Alors,

$$\sqrt{n}(\hat{\gamma}_n - \gamma) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V).$$

Démonstration. Nous allons établir ce résultat en remplaçant $\hat{\gamma}_n$ par $\tilde{\gamma}_n = [\tilde{\gamma}_n(1), \dots, \tilde{\gamma}_n(K)]^T$, car nous savons que ces deux quantités sont asymptotiquement équivalentes. Nous considérons tout d'abord le processus stationnaire au sens strict $(2m + K)$ -dépendant défini par

$$Y_t^m = \begin{bmatrix} (X_t^m - \mu)^2 \\ (X_{t+1}^m - \mu)(X_t^m - \mu) \\ \vdots \\ (X_{t+K}^m - \mu)(X_t^m - \mu) \end{bmatrix},$$

où $X_t^m \stackrel{\text{def}}{=} \mu + \sum_{|j| \leq m} \psi_j Z_{t-j}$. La moyenne empirique de ce processus est donnée par

$$\bar{Y}_n^m = n^{-1} \sum_{t=1}^n Y_t^m = \begin{bmatrix} \tilde{\gamma}_n^m(0) \\ \tilde{\gamma}_n^m(1) \\ \vdots \\ \tilde{\gamma}_n^m(K) \end{bmatrix},$$

où, pour $h \geq 0$,

$$\tilde{\gamma}_n^m(h) = n^{-1} \sum_{t=1}^n (X_{t+h}^m - \mu)(X_t^m - \mu).$$

Nous remarquons que

$$\mathbb{E} [\bar{Y}_{m,n}^m] = \begin{bmatrix} \gamma_n^m(0) \\ \gamma_n^m(1) \\ \vdots \\ \gamma_n^m(K) \end{bmatrix}$$

où $\gamma^m(h) = \text{Cov}(X_0^m, X_h^m)$. Nous allons appliquer le théorème d'approximation 6.3.1. Considérons le vecteur

$$B_{m,n} = n^{1/2} (\bar{Y}_{m,n} - \mathbb{E} [\bar{Y}_{m,n}]) ,$$

qui approche le vecteur

$$A_n = n^{1/2} (\hat{\gamma}_n - \gamma) .$$

1. Soit \mathbf{c} un $(K+1) \times 1$ vecteur déterministe, et appliquons le théorème Central Limite au processus $(2m+K)$ -dépendant $\mathbf{c}^T Y_t$. Nous obtenons

$$B_{m,n} = n^{1/2} (\bar{Y}_t^m - \mathbb{E} [\bar{Y}_0^m]) \xrightarrow{\mathcal{L}} B_m$$

où B_m est une variable aléatoire gaussienne de moyenne nulle et de variance $\mathbf{c}^T V^m \mathbf{c}$, les éléments $V_{p,q}^m$ de la matrice V^m étant donnés par,

$$V_{p,q}^m \stackrel{\text{def}}{=} (\eta - 3)\gamma^m(p)\gamma^m(q) + \sum_{u=-\infty}^{\infty} [\gamma^m(u)\gamma^m(u-p+q) + \gamma^m(u+q)\gamma^m(u-p)] .$$

2. En remarquant que comme $\lim_{m \rightarrow \infty} V_m = V$, nous avons $B_m \xrightarrow{\mathcal{L}} B$ quand $m \rightarrow \infty$, où B est une variable aléatoire gaussienne de moyenne nulle et de variance $\mathbf{c}^T V \mathbf{c}$.
3. Il reste à établir l'uniformité de la convergence : pour tout $\epsilon > 0$,

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(|B_{m,n} - A_n| \geq \epsilon) = 0 .$$

Comme

$$\mathbb{P}(|B_{m,n} - A_n| \geq \epsilon) \leq \sum_{h=0}^K \mathbb{P}(n^{1/2} |\tilde{\gamma}_n^m(h) - \hat{\gamma}_n(h)| \geq \epsilon_K)$$

il suffit d'établir (en utilisant l'inégalité de Bienaymé-Tchebyshev) que

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} n \text{Var}(\tilde{\gamma}_n^m(h) - \hat{\gamma}_n(h)) = 0 . \quad (6.35)$$

En utilisant l'identité

$$n \text{Var}(\tilde{\gamma}_n^m(h) - \hat{\gamma}_n(h)) = n [\text{Var}(\tilde{\gamma}_n^m(h)) + n \text{Var}(\hat{\gamma}_n(h)) + 2 \text{Cov}(\tilde{\gamma}_n^m(h), \hat{\gamma}_n(h))] ,$$

et des calculs similaires à ceux que nous avons effectué pour établir (6.32) and (6.31), nous obtenons (6.35). □

Nous allons étendre le résultat précédent aux coefficients d'autocorrélation $\hat{\rho}_n(h) = \hat{\gamma}_n(h)/\hat{\gamma}_n(0)$.

Preuve du Théorème 6.2.4. La preuve est une application directe du théorème 6.2.3 par application de la delta-méthode (voir paragraphe B.8). On considère la fonction $g : \mathbb{R}^{K+1} \rightarrow \mathbb{R}^K$ définie pour $x_0 \neq 0$ par

$$g(x_0, x_1, \dots, x_K) = [x_1/x_0, \dots, x_K/x_0]^T .$$

Nous avons,

$$g(\hat{\gamma}_n(0), \dots, \hat{\gamma}_n(K)) = (\hat{\rho}_n(1), \dots, \hat{\rho}_n(K))^T.$$

et une application directe de la Proposition B.8.1 montre que

$$\sqrt{n} (g(\hat{\gamma}_n(0), \dots, \hat{\gamma}_n(K)) - g(\gamma(0), \dots, \gamma(K))) \xrightarrow{\mathcal{L}} \mathcal{N}(0, DVD^T)$$

où V est la matrice de covariance des coefficients d'autocovariance donnée par (6.34) et D est la matrice jacobienne de la fonction g au point $[\gamma(0), \dots, \gamma(K)]$,

$$D = \frac{1}{\gamma(0)^2} \begin{bmatrix} -\gamma(1) & \gamma(0) & 0 & \dots & 0 \\ -\gamma(2) & 0 & \gamma(0) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\gamma(K) & 0 & 0 & \dots & \gamma(0) \end{bmatrix}.$$

Cette matrice peut se réécrire de façon plus compacte

$$D = \frac{1}{\gamma(0)} [-\boldsymbol{\rho} I_K]$$

où I_K est la matrice identité $K \times K$. En écrivant la matrice V sous la forme

$$V = \begin{bmatrix} v_{0,0} & \mathbf{v}_1^T \\ \mathbf{v}_1 & V_{2,2} \end{bmatrix}$$

la matrice W peut s'écrire

$$W = \gamma^{-2}(0) [v_0 \boldsymbol{\rho} \boldsymbol{\rho}^T - \boldsymbol{\rho} \mathbf{v}_1^T - \mathbf{v}_1 \boldsymbol{\rho}^T + V_{2,2}],$$

où $\mathbf{v}_1 = [v_{1,0}, v_{2,0}, \dots, v_{K,0}]^T$ et $V_{2,2} = [v_{p,q}, p, q = 1, \dots, K]$. Un calcul élémentaire montre que

$$\begin{aligned} w_{p,q} &= \gamma^{-2}(0) [v_{p,q} - \rho(p)v_{0,q} - \rho(q)v_{p,0} + \rho(p)\rho(q)v_{0,0}] \\ &= \sum_{u=-\infty}^{\infty} [\rho(u)\rho(u-p+q) + \rho(u-p)\rho(u+q) + 2\rho(p)\rho(q)\rho^2(u) - 2\rho(p)\rho(u)\rho(u+q) - 2\rho(q)\rho(u)\rho(u-p)]. \end{aligned}$$

□

Chapitre 7

Estimation de la densité spectrale

Dans le chapitre précédent, nous nous sommes intéressés à l'estimation de la fonction d'autocovariance. Dans certaines applications, il est plus pertinent d'estimer la densité spectrale qui est souvent plus facile à interpréter que la trajectoire ou fonction d'autocovariance, révélant des structures (par exemple, cycles ou pseudo-cycles) qui ne sont pas toujours visibles. Pour nous en convaincre considérons l'exemple de la forme d'onde représentée figure 7.1. Il s'agit d'un segment d'environ 40 millisecondes extrait d'un enregistrement d'un son produit par un harmonica. La trajectoire est complexe, reflétant les deux caractéristiques essentielles du signal produit par cet instrument : des composantes cycliques liées aux vibrations des lames métalliques modulant de façon quasi-périodique le flux d'air et un bruit de friction. La fonction d'autocorrélation, que nous avons représentée à gauche figure 7.2, révèle en effet des structures temporelles complexes mais cette représentation n'est pas apte à réellement mettre en évidence la présence de (pseudo)-cycles. Ceux-ci apparaissent, par contre, clairement quand on observe le module de la transformée de Fourier du signal (à droite figure 7.2). Cette représentation fréquentielle n'est toutefois pas tout à fait satisfaisante, car elle est très "bruitée", ce qui rend difficile son interprétation. Cette variabilité est simplement la traduction, dans le domaine de Fourier, de la variabilité que nous observons dans la forme d'onde.

L'objet de ce chapitre est de trouver une méthode d'estimation spectrale qui, tout en préservant les structures cycliques, soit capable de lisser les fluctuations.

7.1 Le périodogramme

Nous supposons dans cette partie que $\{X_t\}$ est un processus stationnaire au second-ordre de moyenne μ et de fonction de covariance $\gamma(h) \triangleq \mathbb{E}[(X_{t+h} - \mu)(X_t - \mu)]$ absolument sommable : $\sum |\gamma(h)| < \infty$. Sous ces hypothèses, le processus $\{X_t\}$ admet une densité spectrale donnée par :

$$f_X(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-ih\lambda}$$

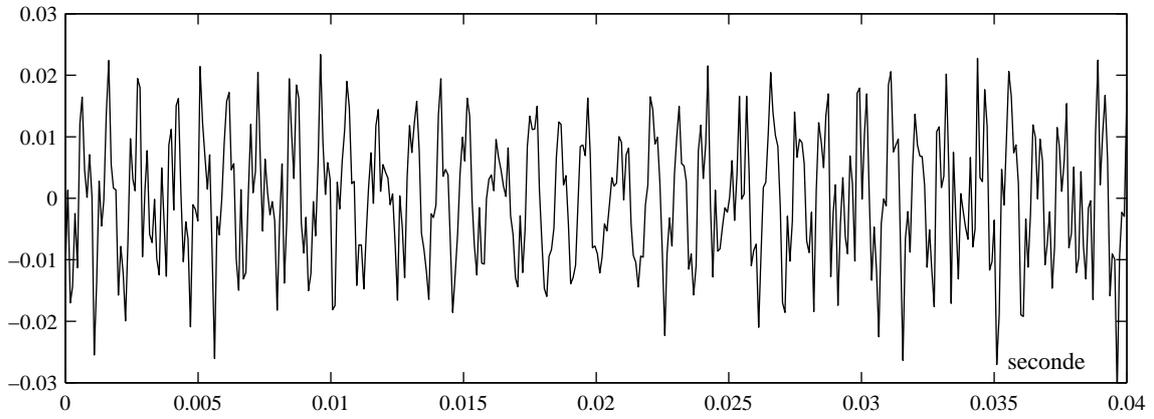


FIGURE 7.1 – *Signal d’harmonica échantillonné à 11.025 kHz (temps en seconde).*

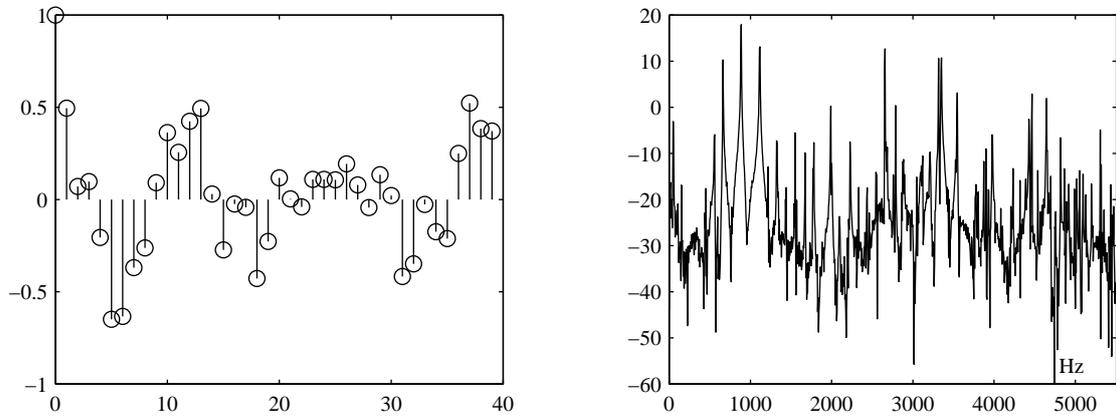


FIGURE 7.2 – *A gauche, suite des 40 premiers coefficients de corrélation du signal représenté figure 7.1. A droite, transformée de Fourier (en dB) de ce signal (fréquence en Hz).*

Pour estimer la densité spectrale de $\{X_t\}$, il est naturel de s'intéresser au périodogramme, défini comme le module au carré de la transformée de Fourier discrète des observations $\{X_1, X_2, \dots, X_n\}$:

$$I_n^X(\lambda_k) = |d_n^X(\lambda_k)|^2 \quad \text{où} \quad d_n^X(\lambda_k) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n X_t e^{-it\lambda_k} \quad (7.1)$$

où $\lambda_k = 2\pi k/n$ sont les fréquences de Fourier. Remarquons ici que la relation :

$$\sum_{t=0}^{n-1} e^{-it\lambda_k} = 0 \quad \text{pour} \quad \lambda_k = 2\pi k/n \text{ et } k \in \{1, \dots, (n-1)\}$$

montre que le périodogramme aux fréquences de Fourier λ_k , non nulles modulo 2π , est invariant par ajout d'une constante. Le périodogramme a été introduit par Sir Arthur Schuster (1898) pour étudier les "périodes cachées" apparaissant dans la série de tâches solaires. L'analyse spectrale des séries temporelles s'est ensuite considérablement développée avec le développement de moyens de calculs performants, et la découverte d'algorithmes de transformée de Fourier rapides.

Malheureusement nous allons voir dans la suite que le périodogramme n'est pas un "bon" estimateur de la densité spectrale, dans le sens où cet estimateur n'est pas consistant (il ne converge pas vers la vraie densité quand n tend vers l'infini). Néanmoins, il est à la base de la construction de la plupart des estimateurs de densité spectrale.

Rappelons tout d'abord que, comme nous l'avons déjà noté dans le chapitre 6 (voir expression (6.7)), le périodogramme est aussi égal à la transformée de Fourier discrète de la suite des coefficients d'autocovariance empiriques. En effet partant de :

$$\hat{\gamma}_n(h) = n^{-1} \sum_{t=1}^{n-|h|} (X_t - \hat{\mu}_n)(X_{t+|h|} - \hat{\mu}_n) \quad \text{où} \quad \hat{\mu}_n = n^{-1} \sum_{t=1}^n X_t$$

on vérifie aisément que

$$I_n^X(0) = \frac{1}{2\pi} n |\hat{\mu}_n|^2 \quad (7.2)$$

$$I_n^X(\lambda_k) = \frac{1}{2\pi} \sum_{h=-(n-1)}^{n-1} \hat{\gamma}_n(h) \exp(-ih\lambda_k) \quad \text{pour } \lambda_k \neq 0 \quad (7.3)$$

Pour estimer la densité spectrale $f_X(\lambda)$ à toutes les fréquences, il est pratique d'étendre le périodogramme pour les valeurs de fréquences normalisées ne coïncidant pas avec les fréquences de Fourier. Ceci peut être fait de différentes manières ; nous suivons l'extension adoptée par Fuller (1976) qui consiste à définir le périodogramme comme la fonction constante par morceaux donnée par :

$$I_n^X(\lambda) = \begin{cases} I_n^X(\lambda_k) & \text{si } \lambda_k - \pi/n < \lambda \leq \lambda_k + \pi/n \text{ et } 0 \leq \lambda \leq \pi \\ I_n^X(-\lambda) & \text{si } -\pi \leq \lambda < 0 \end{cases} \quad (7.4)$$

Par construction, cette définition garantit que le périodogramme est une fonction paire, qui coïncide avec l'équation (7.1) aux fréquences $\lambda_k = 2\pi k/n$. De façon plus concise on peut alors écrire que :

$$I_n^X(\lambda) = I_n^X(g(n, \lambda))$$

où $g(n, \lambda)$ désigne, pour $\lambda \in [0, \pi]$, le multiple de $2\pi/n$ le plus proche de λ et, pour $\lambda \in [-\pi, 0)$, $g(n, \lambda) = g(n, -\lambda)$. La proposition suivante établit que le périodogramme est asymptotiquement sans biais.

Théorème 7.1.1. *Soit $\{X_t\}$ un processus stationnaire de moyenne μ et de fonction d'autocovariance $\gamma(h)$ absolument sommable. Alors quand $n \rightarrow +\infty$ on a :*

$$\mathbb{E} [I_n^X(0)] - \frac{1}{2\pi} n\mu^2 \longrightarrow f_X(0)$$

et $\mathbb{E} [I_n^X(\lambda)] \longrightarrow f_X(\lambda) \quad \text{pour } \lambda \neq 0$

Démonstration. Remarquons que, pour $\lambda \neq 0$, on a :

$$\mathbb{E} [I_n^X(g(n, \lambda))] = \frac{1}{2\pi} \sum_{h=-(n+1)}^{(n-1)} \left(1 - \frac{|h|}{n}\right) \gamma(\tau) e^{-ihg(n, \lambda)}$$

Posons $\gamma_n(h, \lambda) = (2\pi)^{-1} \mathbb{1}_{[-n, n]}(h) (1 - |h|/n) \gamma(h) e^{-ihg(n, \lambda)}$. Nous avons $|\gamma_n(h, \lambda)| \leq |\gamma(h)|$ et $\lim_{n \rightarrow \infty} \gamma_n(h, x) = \gamma(h) e^{-ih\lambda}$. On conclut en appliquant le théorème de convergence dominée. \square

Pour comprendre les propriétés statistiques du périodogramme, nous allons tout d'abord nous intéresser à la distribution statistique du périodogramme d'un bruit blanc fort, c'est-à-dire d'une suite de variables aléatoires indépendantes et identiquement distribuées, de moyenne nulle et de variance finie.

Théorème 7.1.2. *Soit $\{Z_t\}$ une suite de variables aléatoires i.i.d., de moyenne nulle et de variance $\sigma^2 < \infty$. Sa distribution spectrale a pour densité $f_Z(\lambda) = \sigma^2/2\pi$.*

1. Soient $0 < \omega_1 < \dots < \omega_m < \pi$, m fréquences fixes. Le vecteur aléatoire $[I_n^Z(\omega_1), \dots, I_n^Z(\omega_m)]$ converge en loi vers un vecteur de variables aléatoires indépendantes, distribuées suivant une loi exponentielle, de moyenne $\sigma^2/2\pi$.
2. Supposons que $\mathbb{E} [Z_t^4] < \infty$, alors :

$$\text{var}\{I_n^Z(\lambda_k)\} = \begin{cases} 2f_Z^2(\lambda_k) + \kappa_4/4\pi^2n & \lambda_k \in \{0, \pi\} \\ f_Z^2(\lambda_k) + \kappa_4/4\pi^2n & 0 < \lambda_k < \pi \end{cases} \quad (7.5)$$

$$\text{et } \text{cov}\{I_n^Z(\lambda_j), I_n^Z(\lambda_k)\} = \kappa_4/4\pi^2n \quad \text{pour } \lambda_j \neq \lambda_k \quad (7.6)$$

où $\lambda_k = 2\pi k/n$ sont les fréquences de Fourier et où κ_4 est le cumulants d'ordre 4 de la variable Z_1 défini par :

$$\kappa_4 = \mathbb{E} [Z_1^4] - 3(\mathbb{E} [Z_1^2])^2$$

3. Supposons que les variables aléatoires Z_t soient gaussiennes. Alors $\kappa_4 = 0$ et, pour tout n , les variables aléatoires $I_n^Z(\lambda_k)/f_Z(\lambda)$, $k \in \{1, \dots, (n-1)/2\}$ sont indépendantes et identiquement distribuées suivant une loi exponentielle¹ de moyenne 1.

1. Cette loi a pour densité $p(u) = e^{-u} \mathbb{1}(u \geq 0)$.

Démonstration. Elle est donnée en fin de chapitre. □

La relation (7.5) du théorème 7.1.2 montre que *la variance de l'estimateur du périodogramme ne tend pas vers 0 lorsque le nombre d'échantillons tend vers l'infini*. Le périodogramme est bien un estimateur asymptotiquement sans biais de la densité spectrale du bruit blanc, mais *n'est pas consistant*. On voit même que $\sqrt{\text{var}(I_n^Z(\lambda_k))}$ est de l'ordre de σ^2 et donc les fluctuations autour de la vraie valeur sont de l'ordre de grandeur de ce que l'on cherche à estimer. C'est ce que montre la figure 7.3 où nous avons représenté le périodogramme en dB d'un bruit blanc pour différentes valeurs de n . On observe sur ces réalisations qu'à certaines fréquences de Fourier les écarts avec la vraie valeur $\sigma^2/2\pi$ restent très importants même lorsque n augmente. Nous avons aussi reporté (droite en pointillé) le seuil de confiance à $\alpha = 90\%$ de la loi asymptotique de $I_n(\lambda_k)/f_Z(\lambda_k)$. Ce seuil a pour expression $s = -\log(1 - \alpha)$. Partant du théorème 7.1.2, valable pour les processus i.i.d., nous allons

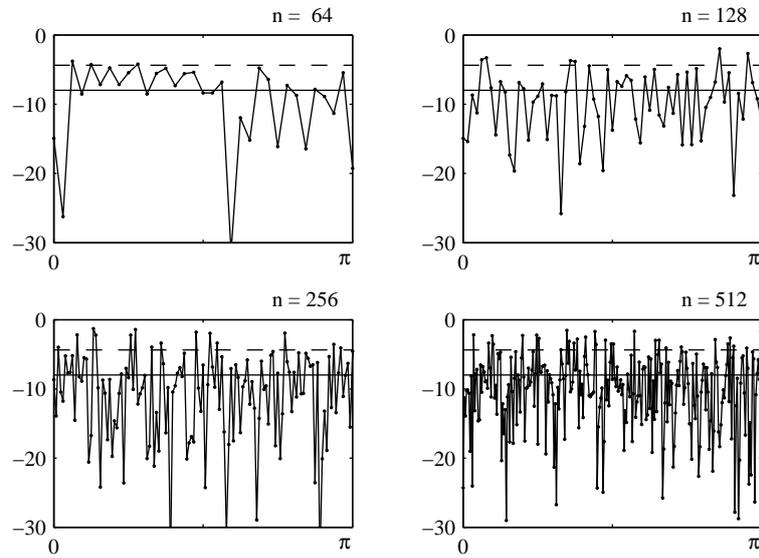


FIGURE 7.3 – Périodogramme en dB d'un bruit blanc de variance 1 en fonction de la fréquence $\lambda \in (0, \pi)$, pour différentes valeurs de n . La droite en trait plein représente la densité spectrale théorique $\sigma^2/2\pi$ et la droite en pointillé le seuil de confiance à 90%.

voir qu'il est encore possible d'étendre ce théorème à la classe plus large des processus linéaires forts centrés dont nous rappelons la définition.

Définition 7.1.3 (Processus linéaire fort). *Le processus $\{X_t\}$ est linéaire fort, s'il existe un bruit blanc fort $Z_t \sim IID(0, \sigma^2)$ et une suite de coefficients $\{\psi_k\}_{k \in \mathbb{Z}}$ absolument sommable telle que :*

$$X_t = \sum_{k=-\infty}^{\infty} \psi_k Z_{t-k} \quad (7.7)$$

On rappelle que X_t est stationnaire au second ordre, que $\mathbb{E}[X_t] = 0$ et que sa densité spectrale est

donnée par :

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} |\psi(e^{-i\lambda})|^2 \quad (7.8)$$

Le théorème 7.1.4 montre qu'il existe une relation analogue à (7.8) entre le périodogramme $I_n^X(\lambda)$ du processus $\{X_t\}$ et le périodogramme $I_n^Z(\lambda)$ du bruit blanc fort $\{Z_t\}$ qui définit X_t .

Théorème 7.1.4. *Soit $\{X_t\}$ un processus linéaire fort. Supposons que $\sum_{k=-\infty}^{\infty} |k|^{1/2} |\psi_k| < \infty$ et que $\mathbb{E}[Z_t^4] < \infty$. On a alors :*

$$I_n^X(\lambda_k) = |\psi(e^{-i\lambda_k})|^2 I_n^Z(\lambda_k) + R_n(\lambda_k)$$

où le terme $R_n(\lambda_k)$ vérifie :

$$\max_{k \in \{1, \dots, \lfloor (n-1)/2 \rfloor\}} \mathbb{E}[|R_n(\lambda_k)|^2] = O(n^{-1})$$

Démonstration. Elle est donnée en fin de chapitre. □

On comprend alors qu'en utilisant l'"approximation" donnée par le théorème 7.1.4 on puisse étendre le théorème 7.1.2 aux processus linéaires forts.

Théorème 7.1.5. *Soit $\{X_t\}$ un processus linéaire défini par :*

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

où $\{Z_t\}$ est un bruit blanc fort IID(0, σ^2) vérifiant $\mathbb{E}[Z_t^4] < \infty$. On suppose que $\sum_k |k|^{1/2} |\psi_k| < \infty$ et que $\psi(e^{-i\lambda}) = \sum_k \psi_k e^{-ik\lambda} \neq 0$. On note :

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

1. Soient $0 < \omega_1 < \dots < \omega_m < \pi$, m fréquences fixes. Le vecteur aléatoire $[I_n^X(\omega_1)/f_X(\omega_1), \dots, I_n^X(\omega_m)/f_X(\omega_m)]$ converge en loi vers un vecteur de variables aléatoires indépendantes, distribuées suivant une loi exponentielle, de moyenne 1.
2. De plus

$$\text{var}(I_n^X(\lambda_k)) = \begin{cases} 2f_X^2(\lambda_k) + O(n^{-1/2}) & \lambda_k \in \{0, \pi\} \\ f_X^2(\lambda_k) + O(n^{-1/2}) & 0 < \lambda_k < \pi \end{cases}$$

$$\text{cov}(I_n^X(\lambda_j), I_n^X(\lambda_k)) = O(n^{-1}) \quad \lambda_j \neq \lambda_k,$$

où les termes $O(\cdot)$ sont uniformes en k et en j .

Démonstration. La preuve est une conséquence directe des théorèmes 7.1.4 et 7.1.2. □

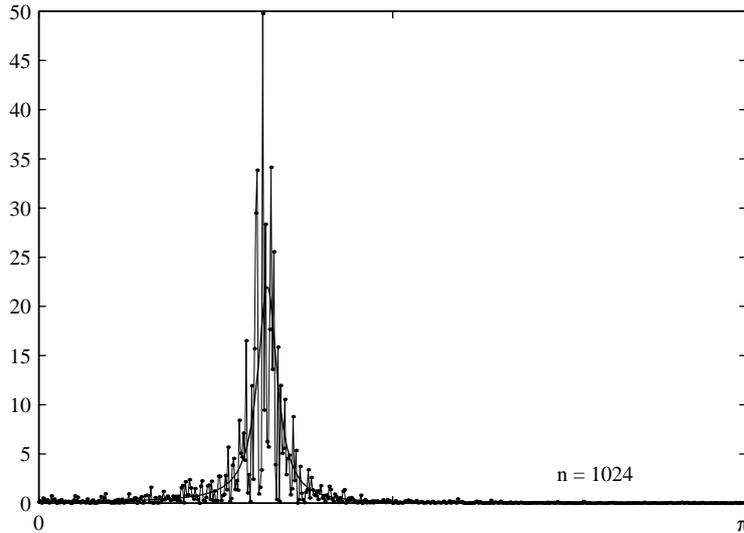


FIGURE 7.4 – Périodogramme pour un $AR(2)$ de paramètres $[1, -1, 0.9]$ et $\sigma^2 = 1$ calculé sur $n = 1024$ échantillons, en fonction de la fréquence $\lambda \in (0, \pi)$.

En conséquence, comme pour le bruit blanc fort, la variance du périodogramme d'un processus linéaire fort est, à une fréquence de Fourier, de l'ordre de grandeur du carré de la densité spectrale à cette fréquence. La figure 7.4 illustre ce résultat : elle montre le périodogramme, évalué sur 1024 échantillons, d'un processus $AR(2)$ gaussien. L'écart-type du périodogramme est proportionnelle à la densité spectrale, ce qui rend bien entendu l'interprétation du périodogramme difficile. Le théorème 7.1.5 implique qu'asymptotiquement les variables aléatoires $[I_n(\lambda_1), \dots, I_n(\lambda_{N/2})]$ se comportent comme un tableau de variables indépendantes distribuées marginalement comme $W f_X(\lambda_k)$ où W suit une loi exponentielle. Il s'agit donc d'une structure de bruit de type multiplicatif, où le paramètre d'intérêt, à savoir la densité spectrale, est *multipliée* par le "bruit" W . L'application d'une transformation logarithmique conduit naturellement à une structure de bruit additif : asymptotiquement le log-périodogramme est égal à la log-densité spectrale observée dans un bruit approximativement additif et de variance constante. Figure 7.4, nous avons représenté le spectre évalué en dB ainsi que l'intervalle de confiance à $\alpha = 90\%$ de la loi asymptotique de $I_n^X(\lambda_k)/f_X(\lambda_k)$ soit :

$$\lim_{n \rightarrow \infty} \mathbb{P} \{ I_n^X(\lambda_k)/f_X(\lambda_k) > c \} = 1 - e^{-c} = \alpha$$

qui donne $c = -\log(1 - \alpha)$.

7.2 Estimateur à noyau

Nous présentons ici une technique permettant de construire un estimateur non paramétrique de la densité spectrale, l'estimateur à noyau. Cette approche, qui effectue un lissage du périodogramme en fréquence, exploite les propriétés du périodogramme que nous avons mises en évidence dans le

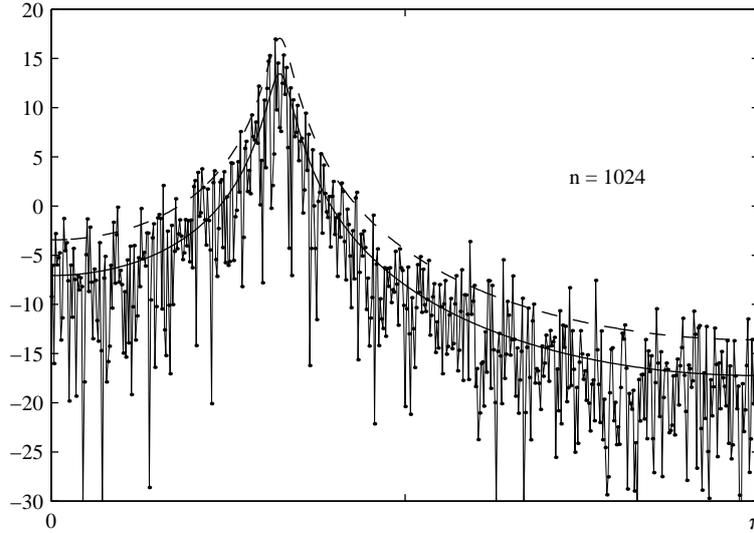


FIGURE 7.5 – Périodogramme en dB pour un $AR(2)$ de paramètres $[1, -1, 0.9]$ et $\sigma^2 = 1$ calculé sur $n = 1024$ échantillons, en fonction de la fréquence $\lambda \in (0, \pi)$. La courbe en pointillé donne le seuil de confiance à 90%.

paragraphe précédent. Nous supposons dans toute cette partie que $\{X_t\}$ est un processus linéaire fort, satisfaisant les conditions d’applications du théorème 7.1.5.

D’après le théorème 7.1.5, à la limite des grands échantillons, les coordonnées du périodogramme aux fréquences de Fourier $\lambda_k = 2\pi k/n$ sont des variables décorréées d’écart type $\sigma^2 |\psi(e^{-i\lambda_k})|^2 / (2\pi)$. La fonction $\lambda \rightarrow |\psi(e^{-i\lambda})|^2$ est continue, elle varie donc “peu” sur de “petits” intervalles de fréquence. Ceci suggère de construire un estimateur de la densité spectrale à la fréquence λ en moyennant les coordonnées du périodogramme aux fréquences de Fourier dans un “voisinage” de la fréquence λ .

Nous considérons l’estimateur à noyau de la densité

$$\hat{f}_n^X(\lambda) = \sum_{k=1}^n W_{m_n}(k) I_n^X(g(n, \lambda) + \lambda_k). \quad (7.9)$$

où $(m_n, n \geq 0)$ est une suite d’entiers et $(W_m, j \geq 0)$ est une suite de fonctions vérifiant

(W-i) $\lim_{n \rightarrow \infty} m_n = \infty$ et $\lim_{n \rightarrow \infty} m_n/n = 0$.

(W-ii) Pour tout $m \in \mathbb{N}$, et tout $k \in \{-m, \dots, m\}$, $W_m(k) = W_m(-k)$, $W_m(k) \geq 0$. De plus, pour $k \notin \{-m, \dots, m\}$, $W_m(k) = 0$.

(W-iii) $\sum_{|k| \leq m} W_m(k) = 1$.

(W-iv) $\lim_{m \rightarrow \infty} \sum_{|k| \leq m} W_m^2(k) = 0$.

Nous allons calculer le risque quadratique de cet estimateur,

$$\text{EQM}(\hat{f}_n^X(\lambda)) = b_{m_n}^2(\lambda) + \sigma_{m_n}^2(\lambda),$$

où $b_{m_n}(\lambda)$ est le biais de l'estimateur

$$b_{m_n}(\lambda) = \mathbb{E} \left[\hat{f}_n^X(\lambda) \right] - f_X(\lambda), \quad (7.10)$$

et $\sigma_{m_n}^2(\lambda)$ est la variance de l'estimateur

$$\sigma_{m_n}^2(\lambda) = \mathbb{E} \left[\left(\hat{f}_n^X(\lambda) - \mathbb{E} \left[\hat{f}_n^X(\lambda) \right] \right)^2 \right]. \quad (7.11)$$

Nous allons étudier ces deux quantités à une fréquence λ donnée. Nous supposons dans cette partie que le processus $\{X_t\}$ est linéaire fort,

$$X_t = \mu + \sum_{k=-\infty}^{\infty} \psi_k Z_{t-k}, \quad \sum_{k=-\infty}^{\infty} |k|^3 |\psi_k| < \infty, \quad (7.12)$$

et

$$\{Z_t\} \text{ est un IID}(0, \sigma^2) \text{ tel que } \mathbb{E} [Z_t^4] = \eta\sigma^4. \quad (7.13)$$

La condition $\sum_{k=-\infty}^{\infty} |k|^3 |\psi_k| < \infty$ implique que la densité spectrale est trois fois continûment différentiable.

Proposition 7.2.1. *Supposons que $\{X_t\}$ est un processus linéaire vérifiant (7.12) et (7.13). Supposons de plus que la suite $(m_n)_{n \geq 0}$ et le noyau W satisfasse les conditions (W-i-iv) pour un λ . Alors, il existe une constante C_1 telle que, pour tout n suffisamment grand ,*

$$\left| \sigma_{m_n}^2(\lambda) - \left(\sum_{k=-m_n}^{m_n} W_{m_n}^2(k) \right) f_X^2(\lambda) \right| \leq C_1 \left(n^{-1} + (n^{-1/2} + m_n/n) \sum_{k=-m_n}^{m_n} W_{m_n}^2(k) \right), \quad (7.14)$$

Démonstration. Supposons que $\lambda \in]0, \pi/2[$. Nous prenons n suffisamment grand pour que

$$0 < g(n, \lambda) - 2\pi m/n \leq g(n, \lambda) + 2\pi m/n < \pi/2.$$

Par construction de l'estimateur

$$\begin{aligned} \text{Var} \left(\hat{f}_n^X(\lambda) \right) &= \sum_k W_{m_n}^2(k) \text{Var} \left(I_n^X [g(n, \lambda) + \lambda_k] \right) \\ &\quad + \sum_{k \neq \ell} W_{m_n}(k) W_{m_n}(\ell) \text{Cov} \left(I_n^X [g(n, \lambda) + \lambda_k], I_n^X [g(n, \lambda) + \lambda_\ell] \right). \end{aligned}$$

En utilisant le Théorème 7.1.5, il existe une constante C telle que :

$$\begin{aligned} \left| \text{Var} \left(I_n^X (g(n, \lambda) + \lambda_k) \right) - f_X^2(g(n, \lambda) + \lambda_k) \right| &\leq C n^{-1/2} \\ \text{Cov} \left(I_n^X (g(n, \lambda) + \lambda_k), I_n^X (g(n, \lambda) + \lambda_\ell) \right) &\leq C n^{-1} \end{aligned}$$

pour tout $k, \ell \in \{-m_n, \dots, m_n\}^{\times 2}$.

□

Pour simplifier l'analyse, nous supposons dans la suite que la fonction $\lambda \rightarrow |\psi(e^{-i\lambda})|^2$ est trois fois différentiable sur $[-\pi, \pi]$ et que la dérivée troisième est bornée. En utilisant les résultats du théorème 7.1.4 nous avons :

$$\mathbb{E} \left[\hat{f}_n^X(\lambda) \right] = \sum_{|k| \leq m} W_{m,n}(k) f_X [g(n, \lambda) + 2\pi k/n] + O(n^{-1}) \quad (7.15)$$

où $f_X(\lambda) = (2\pi)^{-1} \sigma^2 |\psi(e^{-i\lambda})|^2$ est la densité spectrale du processus $\{X_t\}$. Comme la fonction f_X est deux fois continûment différentiables, nous avons, pour $|k| \leq m$,

$$f_X [g(n, \lambda) + 2\pi k/n] = f_X [g(n, \lambda)] + f_X' [g(n, \lambda)] (2\pi k/n) + (1/2) f_X'' [g(n, \lambda)] (2\pi k/n)^2 + R_{k,m,n}$$

où $R_{k,m,n} \leq c \max |f_X'''(\lambda)| (m/n)^3$ pour $|k| \leq m$. Comme la fenêtre de pondération est symétrique, nous avons $\sum_{|k| \leq m_n} W_{m_n}(k) k = 0$, ce qui implique en utilisant (iii) :

$$\sum_{|k| \leq m_n} W_{m_n}(k) f_X [g(n, \lambda) + 2\pi k/n] = f_X(\lambda) + (1/2) f_X''(\lambda) \bar{W}_{m,n} + R_{m,n},$$

où $\bar{W}_{m,n} = \frac{4\pi^2}{n^2} \sum_{|k| \leq m} k^2 W_{m,n}(k)$ et $|R_{m,n}| \leq c \max |f_X'''(\lambda)| (m/n)^3$. En prenant par exemple la fenêtre de pondération rectangulaire, nous avons $\bar{W}_{m,n} \propto m^2/n^2$ ce qui montre que le biais de l'estimateur varie comme le carré du nombre de points de fréquence pris en compte dans le calcul de la moyenne pondérée. Le calcul de la variance de cet estimateur s'écrit :

$$\mathbb{E} \left[\left(\hat{f}_n^X(\lambda) - \mathbb{E} \left[\hat{f}_n^X(\lambda) \right] \right)^2 \right] = \widetilde{W}_{m,n} f_X^2(g(n, \lambda)) + Q_{m,n},$$

où $\widetilde{W}_{m,n} = \frac{1}{4\pi^2} \sum_{|k| \leq m} W_{m,n}^2(k)$ et $|Q_{m,n}| \leq c \max [|f_X'(\lambda)|] \sum_{|k| \leq m} W_{m,n}^2(k) (m/n)$. La condition W-(iii) assure que la variance tend vers 0 quand n tend vers l'infini. En s'appuyant encore sur l'exemple de la fenêtre rectangulaire, nous avons $\widetilde{W}_{m,n} \propto 1/m$ ce qui montre que la variance de l'estimateur est inversement proportionnelle au nombre de points pris en compte dans le calcul de la moyenne locale. En conclusion dans le cas d'une fenêtre rectangulaire, le paramètre m (qui détermine le nombre de coordonnées de périodogramme moyennées) a un effet *néfaste* pour le biais et *bénéfique* pour la variance de l'estimateur. Le risque quadratique de l'estimateur (qui prend en compte ces deux effets) a pour expression :

$$\mathbb{E} \left[\left(\hat{f}_n^X(\lambda) - f_X(\lambda) \right)^2 \right] \approx (1/4) (f_X''(\lambda) \bar{W}_{m,n})^2 + \widetilde{W}_{m,n} f_{X,m}^2(\lambda)$$

Il est naturel de choisir le paramètre m de façon à minimiser l'erreur quadratique moyenne. Dans le cas où $W_{m,n}(k) = 1/(2m+1)$, cette optimisation peut être effectuée de façon explicite. Une autre fenêtre couramment utilisée est la fenêtre triangulaire définie par :

$$W_{m,n}(k) = \begin{cases} \frac{1}{m} \left(1 - \frac{|k|}{m} \right) & \text{pour } |k| \leq m \\ 0 & \text{sinon} \end{cases}$$

Elle vérifie les conditions W-(ii),(ii),(iv) et présente l'avantage d'assurer au spectre estimé d'être positif. Les résultats obtenus avec la fenêtre rectangulaire ont un caractère général : l'utilisation de fenêtre

de pondération permet d'obtenir un risque qui tend vers 0 quand n tend vers l'infini. Ce résultat s'accompagne en général d'un biais asymptotiquement non nul. En règle générale, la valeur de m , qui détermine la largeur de la fenêtre, doit tendre vers l'infini, quand $n \rightarrow +\infty$, mais suffisamment lentement pour que le rapport n/m tende aussi vers l'infini. Il faut donc ajouter aux conditions W-(ii),(iii),(iv) la condition suivante :

$$m(n) \rightarrow \infty \text{ et } m(n)/n \rightarrow 0 \text{ quand } n \rightarrow \infty$$

Typiquement on aura $m(n) = n^\alpha$ avec $0 < \alpha < 1$.

7.3 Preuves des théorèmes 7.1.2, 7.1.4

Preuve du théorème 7.1.2. (i). Notons :

$$\begin{cases} \alpha_n^Z(\lambda_k) = (1/2\pi n)^{-1/2} \sum_{t=1}^n Z_t \cos(\lambda_k t) \\ \beta_n^Z(\lambda_k) = (1/2\pi n)^{-1/2} \sum_{t=1}^n Z_t \sin(\lambda_k t) \end{cases} \quad (7.16)$$

les parties réelles et imaginaire de la transformée de Fourier discrète de $\{Z_t\}$ aux points de fréquences $\lambda_k = 2\pi k/n$. Pour une fréquence arbitraire λ , nous avons :

$$I_n^Z(\lambda) = \frac{1}{2} (\alpha_n^Z(g(n, \lambda))^2 + \beta_n^Z(g(n, \lambda))^2)$$

Rappelons que si une suite de vecteurs aléatoires Y_n converge en loi vers une variable aléatoire Y et que ϕ est une fonction continue, alors $\phi(Y_n)$ converge en loi vers $\phi(Y)$. Il suffit donc de montrer que le vecteur aléatoire :

$$(\alpha_n^Z(\lambda_1), \beta_n^Z(\lambda_1), \dots, \alpha_n^Z(\lambda_m), \beta_n^Z(\lambda_m)) \quad (7.17)$$

converge en loi vers une distribution normale de moyenne nulle et de matrice de covariance asymptotique $(\sigma^2/4\pi)I_{2m}$, où I_{2m} est la matrice identité $(2m \times 2m)$. Nous allons tout d'abord nous intéresser au cas $m = 1$. La preuve découle alors du théorème suivant :

Théorème 7.3.1 (Lindeberg). *Soit $U_{n,t}$, où $t = 1, \dots, n$ et $n = 1, 2, \dots$, une suite triangulaire de variables aléatoires centrées de variance finies. Pour tout n , les variables $\{U_{n,1}, \dots, U_{n,n}\}$ sont indépendantes. On pose $Y_n = \sum_{t=1}^n U_{n,t}$ et $w_n^2 = \sum_{t=1}^n \text{var}(U_{n,t})$. Alors si pour tout $\epsilon > 0$:*

$$\lim_{n \rightarrow \infty} \sum_{t=1}^n \frac{1}{w_n^2} \mathbb{E} [U_{n,t}^2 \mathbb{1}(|U_{n,t}| \geq \epsilon w_n)] = 0$$

on a :

$$Y_n/w_n \rightarrow_d \mathcal{N}(0, 1)$$

Soit u et v deux réels quelconques fixés et $\lambda \in (0, \pi)$. Considérons la variable $Y_n = u\alpha_n^Z(g(n, \lambda)) + v\beta_n^Z(g(n, \lambda))$ que nous pouvons encore écrire :

$$Y_n = \sum_{t=1}^n U_{n,t} \quad \text{où} \quad U_{n,t} = \frac{1}{\sqrt{2\pi n}} (u \cos(g(n, \lambda)t) + v \sin(g(n, \lambda)t)) Z_t$$

Notons que, pour n fixé les variables aléatoires $\{U_{n,t}\}$ sont indépendantes. D'autre part, pour tout $\lambda \neq 0$, on vérifie aisément que :

$$\sum_{t=1}^n \cos^2(g(n, \lambda)t) = \sum_{t=1}^n \sin^2(g(n, \lambda)t) = \frac{n}{2} \quad \text{et} \quad \sum_{t=1}^n \cos((g(n, \lambda)t) \sin(g(n, \lambda)t) = 0$$

Par suite, on peut écrire que :

$$\begin{aligned} w_n^2 &= \sum_{t=1}^n \text{var}(U_{n,t}) \\ &= \frac{1}{2\pi n} \sum_{t=1}^n (u^2 \cos^2(g(n, \lambda)t) + v^2 \sin^2(g(n, \lambda)t) + 2uv \cos((g(n, \lambda)t) \sin(g(n, \lambda)t))) \\ &= \frac{1}{4\pi} (u^2 + v^2) = w_1^2 \end{aligned}$$

Par suite, en posant $c_0 = (|u| + |v|)/2\pi w_1$ et $\epsilon' = \epsilon\sqrt{2\pi}w_1/(|u| + |v|)$, on a :

$$\sum_{t=1}^n \frac{1}{w_n^2} \mathbb{E} [U_{n,t}^2 \mathbb{1}(|U_{n,t}| \geq \epsilon w_n)] \leq \frac{c_0}{n} \sum_{t=1}^n \mathbb{E} [Z_t^2 \mathbb{1}(|Z_t| \geq \epsilon' \sqrt{n})] = c_0 \mathbb{E} [Z_1^2 \mathbb{1}(|Z_1| \geq \epsilon' \sqrt{n})]$$

Le dernier terme tend vers 0 puisque on a $\mathbb{E} [Z_1^2 \mathbb{1}(|Z_1| \geq \epsilon' \sqrt{n})] \leq \mathbb{E} [|Z_1|^3] / \epsilon' \sqrt{n}$ et que $\mathbb{E} [|Z_1|^3] < \infty$ puisque $\mathbb{E} [|Z_1|^4] < \infty$. La preuve s'étend aisément à un ensemble de fréquences $\lambda_1, \dots, \lambda_m$ en utilisant la méthode de Cramer-Wold (see Proposition B.3.4)

(ii). Par définition de $I_n^Z(\lambda_k)$, nous avons au premier ordre :

$$\mathbb{E} [I_n^Z(\lambda_k)] = (2\pi n)^{-1} \sum_{s,t=1}^n \mathbb{E} [Z_s Z_t] e^{i\lambda_k(t-s)} = (2\pi)^{-1} \sigma^2 \quad (7.18)$$

Au second ordre nous avons :

$$\mathbb{E} [I_n^Z(\lambda_j) I_n^Z(\lambda_k)] = (2\pi n)^{-2} \sum_{s,t,u,v=1}^n \mathbb{E} [Z_s Z_t Z_u Z_v] e^{i(\lambda_j(t-s) + \lambda_k(v-u))} \quad (7.19)$$

En utilisant que les variables aléatoires Z_t sont indépendantes, centrées, de même variance σ^2 et de moment d'ordre 4 fini et en posant $\mathbb{E} [Z_1^4] = \kappa_4 + 3\sigma^4$, on obtient :

$$\mathbb{E} [Z_s Z_t Z_u Z_v] = \kappa_4 \delta_{s,t,u,v} + \sigma^4 (\delta_{s,t} \delta_{u,v} + \delta_{s,u} \delta_{t,v} + \delta_{s,v} \delta_{t,u}) \quad (7.20)$$

En portant cette expression dans (7.19), nous avons :

$$\mathbb{E} [I_n^Z(\lambda_j) I_n^Z(\lambda_k)] = (2\pi)^{-2} n^{-1} \kappa_4 + (2\pi)^{-2} n^{-2} \sigma^4 \left(n^2 + \left| \sum_{t=1}^n e^{i(\lambda_j + \lambda_k)t} \right|^2 + \left| \sum_{t=1}^n e^{i(\lambda_k - \lambda_j)t} \right|^2 \right)$$

et donc :

$$\begin{aligned} \text{cov}(I_n^Z(\lambda_j), I_n^Z(\lambda_k)) &= \mathbb{E} [I_n^Z(\lambda_j) I_n^Z(\lambda_k)] - \mathbb{E} [I_n^Z(\lambda_j)] \mathbb{E} [I_n^Z(\lambda_k)] \\ &= (2\pi)^{-2} n^{-1} \kappa_4 + (2\pi)^{-2} n^{-2} \sigma^4 \left(\left| \sum_{t=1}^n e^{i(\lambda_j + \lambda_k)t} \right|^2 + \left| \sum_{t=1}^n e^{i(\lambda_k - \lambda_j)t} \right|^2 \right) \end{aligned}$$

ce qui permet de conclure.

(iii). Lorsque $\{Z_t\}$ est une variable gaussienne centrée, le vecteur :

$$Q_n = [\alpha_n^Z(\lambda_1) \quad \beta_n^Z(\lambda_1) \quad \cdots \quad \alpha_n^Z(\lambda_{\tilde{n}}) \quad \beta_n^Z(\lambda_{\tilde{n}})]$$

est gaussien comme transformée linéaire d'un vecteur gaussien. Il suffit donc de calculer le vecteur-moyenne et sa matrice de covariance. Il est facile de vérifier que le vecteur-moyenne est nul et que, pour $0 < \lambda_k \neq \lambda_j < \pi$, nous avons :

$$\begin{aligned} \mathbb{E} [(\alpha_n^Z(\lambda_k))^2] &= \mathbb{E} [(\beta_n^Z(\lambda_k))^2] = (4\pi)^{-1} \\ \mathbb{E} [\alpha_n^Z(\lambda_k) \beta_n^Z(\lambda_k)] &= 0 \\ \mathbb{E} [\alpha_n^Z(\lambda_k) \alpha_n^Z(\lambda_j)] &= \mathbb{E} [\beta_n^Z(\lambda_k) \beta_n^Z(\lambda_j)] = 0 \\ \mathbb{E} [\alpha_n^Z(\lambda_k) \beta_n^Z(\lambda_j)] &= 0 \end{aligned}$$

La matrice de covariance est donc $\sigma^2 I_{\tilde{n}}/4\pi$ où $I_{\tilde{n}}$ est la matrice identité de taille \tilde{n} . Par conséquent les composantes de Q_n sont indépendantes. Rappelons que :

$$I_n^Z(\lambda_k) = (\alpha_n^Z(\lambda_k))^2 + (\beta_n^Z(\lambda_k))^2$$

De l'indépendance des composantes de Q_n , on déduit que les variables aléatoires $I_n^Z(\lambda_k)$ sont elles-même indépendantes et que $4\pi I_n^Z(\lambda_k)/\sigma^2$ est la somme du carré de deux variables gaussiennes centrées, indépendantes, de même variance 1, dont la distribution de probabilité est la loi dite du χ^2 à deux degrés de liberté. Ce qui conclut la preuve. \square

Preuve du théorème 7.1.4. Notons respectivement $d_n^X(\lambda_k)$ et $d_n^Z(\lambda_k)$ les transformées de Fourier discrètes des suites $\{X_1, \dots, X_n\}$ et de $\{Z_1, \dots, Z_n\}$ au point de fréquence $2\pi k/n$ avec $k \in \{1, \dots, \lfloor (n-1)/2 \rfloor\}$. Nous pouvons écrire successivement :

$$\begin{aligned} d_n^X(\lambda_k) &= (2\pi n)^{-1/2} \sum_{t=1}^n X_t e^{-i\lambda_k t} \\ &= (2\pi n)^{-1/2} \sum_{j=-\infty}^{\infty} \psi_j e^{-i\lambda_k j} \left(\sum_{t=1}^n Z_{t-j} e^{-i\lambda_k (t-j)} \right) \\ &= (2\pi n)^{-1/2} \sum_{j=-\infty}^{\infty} \psi_j e^{-i\lambda_k j} \left(\sum_{t=1-j}^{n-j} Z_t e^{-i\lambda_k t} \right) \\ &= (2\pi n)^{-1/2} \sum_{j=-\infty}^{\infty} \psi_j e^{-i\lambda_k j} \left(\sum_{t=1}^n Z_t e^{-i\lambda_k t} + U_{n,j}(\lambda_k) \right) \\ &= \psi(e^{-i\lambda_k}) d_n^Z(\lambda_k) + Y_n(\lambda_k) \end{aligned}$$

où nous avons posé :

$$U_{n,j}(\lambda_k) = \sum_{t=1-j}^{n-j} Z_t e^{-i\lambda_k t} - \sum_{t=1}^n Z_t e^{-i\lambda_k t} \quad (7.21)$$

$$\text{et } Y_n(\lambda_k) = (2\pi n)^{-1/2} \sum_{j=-\infty}^{\infty} \psi_j e^{-i\lambda_k j} U_{n,j}(\lambda_k) \quad (7.22)$$

On remarque que, pour $|j| < n$, $U_{n,j}(\lambda_k)$ est une somme de $2|j|$ variables indépendantes centrées de variance σ^2 tandis que, pour $|j| \geq n$, $U_{n,j}(\lambda_k)$ est la somme de $2n$ variables centrées indépendantes de variance σ^2 . Par conséquent, partant de (7.21), on a :

$$\mathbb{E} [|U_{n,j}(\lambda_k)|^2] \leq 2\sigma^2 \min(|j|, n) \quad (7.23)$$

ainsi que :

$$\mathbb{E} [|U_{n,j}(\lambda_k)|^4] \leq C_R \sigma^4 (\min(|j|, n))^2 \quad (7.24)$$

où $C_R < \infty$ est une constante. Pour montrer (7.24), il suffit de poser $\mathbb{E} [Z_t^4] = \eta\sigma^4$ et d'utiliser l'inégalité (7.25) pour $p = 4$.

Proposition 7.3.2 (Inégalité de Rosenthal (Petrov, 1985)). *Soient (X_1, \dots, X_n) des variables indépendantes (mais pas nécessairement identiquement distribuées) et soit $p \geq 2$. Alors il existe une constante universelle $C(p) < \infty$ telle que :*

$$\mathbb{E} \left[\left| \sum_{k=1}^n X_k \right|^p \right] \leq C(p) \left(\left(\sum_{k=1}^n \mathbb{E} [X_k^2] \right)^{p/2} + \sum_{k=1}^n \mathbb{E} [|X_k|^p] \right) \quad (7.25)$$

Utilisons à présent (7.24) pour majorer $\mathbb{E} [|Y_n(\lambda_k)|^4]$. En adoptant la notation $\|X\|_p = (\mathbb{E} [|X|^p])^{1/p}$ (pour $p > 0$) on a, d'après l'inégalité triangulaire (inégalité de Minkovski) $\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$:

$$\sup_{k \in \{1, \dots, \lfloor (n-1)/2 \rfloor\}} \|Y_n(\lambda_k)\|_4 \leq \sup_{k \in \{1, \dots, \lfloor (n-1)/2 \rfloor\}} (2\pi n)^{-1/2} \sum_{j=-\infty}^{\infty} |\psi_j| \|U_{n,j}(\lambda_k)\|_4$$

D'après (7.24), $\|U_{n,j}(\lambda_k)\|_4 \leq c\sigma \min(|j|, n)^{1/2}$. Par conséquent :

$$\sup_{k \in \{1, \dots, \lfloor (n-1)/2 \rfloor\}} \|Y_n(\lambda_k)\|_4 \leq c\sigma (2\pi n)^{-1/2} \sum_{j=-\infty}^{\infty} |\psi_j| \min(|j|, n)^{1/2}$$

Maintenant on peut écrire :

$$\sum_{j=-\infty}^{\infty} |\psi_j| \min(|j|, n)^{1/2} \leq \sum_{j=-\infty}^{\infty} |\psi_j| |j|^{1/2}$$

Par conséquent $\|Y_n(\lambda_k)\|_4$ est d'un ordre égal à $O(n^{-1/2})$.

Nous pouvons à présent exprimer $R_n(\lambda_k) = I_n^X(\lambda_k) - |\psi(e^{-i\lambda_k})|^2 I_n^Z(\lambda_k)$ en fonction de $Y_n(\lambda_k) = d_n^X(\lambda_k) - \psi(e^{-i\lambda_k}) d_n^Z(\lambda_k)$. Il vient :

$$\begin{aligned} R_n(\lambda_k) &= |\psi(e^{-i\lambda_k}) d_n^Z(\lambda_k) + Y_n(\lambda_k)|^2 - |\psi(e^{-i\lambda_k})|^2 I_n^Z(\lambda_k) \\ &= \psi(e^{-i\lambda_k}) d_n^Z(\lambda_k) Y_n(-\lambda_k) + \psi(e^{i\lambda_k}) d_n^Z(-\lambda_k) Y_n(\lambda_k) + |Y_n(\lambda_k)|^2 \end{aligned}$$

D'après l'inégalité de Hölder, $\|XY\|_r \leq \|X\|_p \|Y\|_q$ si $p^{-1} + q^{-1} = r^{-1}$. En faisant $p = q = 4$ et $r = 2$, il vient :

$$(\mathbb{E} [|R_n(\lambda_k)|^2])^{1/2} = \|R_n(\lambda_k)\|_2 \leq 2 \sum_j |\psi_j| \|d_n^Z(\lambda_k)\|_4 \|Y_n(\lambda_k)\|_4 + \|Y_n(\lambda_k)\|_4$$

D'après le théorème 7.1.2, $\|d_n^Z(\lambda_k)\|_4$ est de l'ordre de $\sigma/\sqrt{2\pi}$. Par conséquent $\|R_n(\lambda_k)\|_2$ est de l'ordre de $n^{-1/2}$ et $\mathbb{E} [|R_n(\lambda_k)|^2] = \|R_n(\lambda_k)\|_2^2$ de l'ordre de $1/n$. Ce qui conclut la preuve. \square

Troisième partie

Annexes

Annexe A

Eléments d'Analyse Hilbertienne

A.1 Définitions

Définition A.1.1 (Espace pré-hilbertien). Soit \mathcal{H} un espace vectoriel sur l'ensemble des nombres complexes \mathbb{C} . L'espace \mathcal{H} est appelé pré-hilbertien si \mathcal{H} est muni d'un produit scalaire :

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

qui vérifie les propriétés suivantes :

- (i) pour tout $(x, y) \in \mathcal{H} \times \mathcal{H}$, $\langle x, y \rangle = \langle y, x \rangle^*$
- (ii) pour tout $(x, y) \in \mathcal{H} \times \mathcal{H}$ et tout $(\alpha, \beta) \in \mathbb{C} \times \mathbb{C}$, $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$
- (iii) pour tout $x \in \mathcal{H}$, $\langle x, x \rangle \geq 0$, et $\langle x, x \rangle = 0$ si et seulement si $x = 0$.

L'application :

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

définit une norme pour tout vecteur x .

Exemple A.1.2 (Espace \mathbb{R}^n). L'ensemble des vecteurs colonnes $x = [x_1 \ \dots \ x_n]^T$, où $x_k \in \mathbb{R}$, est un espace vectoriel dans lequel la relation :

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

définit par un produit scalaire.

Exemple A.1.3 (Espace ℓ^2). L'ensemble des suites numériques complexes $\{x_k\}_{k \in \mathbb{N}}$ vérifiant $\sum_{k=0}^{\infty} |x_k|^2 < \infty$ est un espace vectoriel sur \mathbb{C} . On munit cet espace du produit intérieur :

$$\langle x, y \rangle = \sum_{k=0}^{\infty} x_k y_k^* \leq (1/2) \sum_{k=0}^{\infty} (|x_k|^2 + |y_k|^2) < \infty$$

On vérifie aisément les propriétés (i-iii) de la définition A.1.1. L'espace ainsi défini est donc un espace pré-Hilbertien, que l'on note ℓ^2 .

Exemple A.1.4 (Fonctions de carré intégrable). *L'ensemble $\mathcal{L}^2(T)$ des fonctions boréliennes définies sur un intervalle T de \mathbb{R} , à valeurs complexes et de module de carré intégrable par rapport à la mesure de Lebesgue ($\int_T |f(t)|^2 dt < \infty$) est un espace vectoriel. Considérons alors le produit intérieur :*

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

On montre aisément que $\langle f, g \rangle < \infty$ ainsi que les propriétés (i) et (ii) de la définition A.1.1. Par contre la propriété (iii) n'est pas vérifiée puisque :

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

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

Exemple A.1.5 (Variables aléatoires de variance finie). *De façon similaire à l'exemple A.1.4, pour tout espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, on définit $\mathcal{H} = L^2(\Omega, \mathcal{F}, \mathbb{P})$ (noté $\mathcal{L}^2(\Omega)$ s'il n'y a pas de confusion possible) comme l'ensemble des v.a. X définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs complexes telles que*

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

(On dit que X a une "variance finie".) Sur cet ensemble, on définit

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

Pour les mêmes raisons que dans l'exemple A.1.4, on définit l'espace pré-Hilbertien $L^2(\Omega, \mathcal{F}, \mathbb{P})$ (ou $L^2(\Omega)$) comme l'ensemble des classes d'équivalences de $\mathcal{L}^2(\Omega)$ pour la relation d'équivalence définie par l'égalité presque sûre entre deux v.a. Cet exemple se généralise en fait à tout espace mesuré $(\Omega, \mathcal{F}, \mu)$ en posant

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

On montre aisément les propriétés suivantes :

Théorème A.1.6. *Pour tout $x, y \in \mathcal{H} \times \mathcal{H}$, nous avons :*

- (i) Inégalité de Cauchy-Schwarz : $|\langle x, y \rangle| \leq \|x\| \|y\|$,*
- (ii) Inégalité triangulaire : $|\|x\| - \|y\|| \leq \|x - y\| \leq \|x\| + \|y\|$,*
- (iii) Identité du parallélogramme :*

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

Définition A.1.7 (Convergence dans \mathcal{H}). *Soit x_n une suite de vecteurs et x un vecteur d'un espace \mathcal{H} muni d'un produit scalaire. On dit que x_n tend vers x si et seulement si $\|x_n - x\| \rightarrow 0$ quand $n \rightarrow +\infty$. On note $x_n \rightarrow x$.*

Proposition A.1.8. *Si dans un espace de Hilbert la suite $x_n \rightarrow x$, alors x_n est bornée.*

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

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

□

Proposition A.1.9 (Continuité du produit scalaire). *Soit $x_n \rightarrow x$ et $y_n \rightarrow y$ deux suites convergentes de vecteurs d'un espace pré-hilbertien \mathcal{H} . Alors quand $n \rightarrow +\infty$: $\langle x_n, y_n \rangle \rightarrow \langle x, y \rangle$. En particulier, si $x_n \rightarrow x$, $\|x_n\| \rightarrow \|x\|$.*

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

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

Il suffit ensuite d'évoquer la convergence et la bornitude des suites x_n et y_n . □

Définition A.1.10 (Suite de Cauchy). *Soit x_n une suite de vecteurs d'un espace pré-hilbertien \mathcal{H} . On dit que x_n est une suite de Cauchy si et seulement si :*

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

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

Notons qu'en vertu de l'inégalité triangulaire toute suite convergente est une suite de Cauchy. La réciproque est fautive : une suite de Cauchy peut ne pas être convergente. En voici un contre-exemple :

Exemple A.1.11 (Suite de Cauchy non convergente). *Soit $\mathcal{C}([-\pi, \pi])$ l'espace des fonctions continues sur $[-\pi, \pi]$. L'espace $\mathcal{C}([-\pi, \pi])$, muni du produit $\int_{-\pi}^{\pi} f(x)g^*(x)dx$, est un espace pré-hilbertien. Considérons la suite de fonctions :*

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

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

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

D'autre part on montre aisément que la limite de cette suite $f_\infty(x) = \sum_{k=1}^{\infty} k^{-1} \cos(kx) = \log |\sin(x/2)|$ n'est pas continue et n'appartient donc pas à $\mathcal{C}([-\pi, \pi])$.

Définition A.1.12 (Espace de Hilbert). *On dit qu'un espace vectoriel est complet si toute suite de Cauchy de \mathcal{H} converge dans \mathcal{H} . On dit \mathcal{H} est un espace de Hilbert si \mathcal{H} est pré-hilbertien et complet.*

Exemple A.1.13 (Espace de suite). *L'espace ℓ^2 est un espace de Hilbert. Soit (a_n) une suite de Cauchy dans ℓ^2 . Si nous notons*

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

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

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

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

En utilisant (A.1), nous avons pour tout $p \in \mathbb{N}$, et tout $m, n \geq N$,

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

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

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

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

Proposition A.1.14 (Espaces L^2). *Pour tout espace mesurable $(\Omega, \mathcal{F}, \mu)$, L'espace $L^2(\Omega, \mathcal{F}, \mu)$ (voir l'exemple A.1.5) des fonctions de carré intégrable pour la mesure μ est un espace de Hilbert.*

Définition A.1.15 (Sous espace vectoriel). *Un sous-espace \mathcal{E} d'un espace vectoriel \mathcal{H} est un sous-ensemble de \mathcal{H} tel que, pour tout $x, y \in \mathcal{E}$ et tout scalaire α, β , $\alpha x + \beta y \in \mathcal{E}$. Un sous-espace vectoriel est dit propre si $\mathcal{E} \neq \mathcal{H}$.*

Définition A.1.16 (Sous-espace fermé). *Soit \mathcal{E} un sous-espace d'un espace de Hilbert \mathcal{H} . On dit que \mathcal{E} est fermé, si toute suite $\{x_n\}$ de \mathcal{E} , qui converge, converge dans \mathcal{E} .*

Exemple A.1.17 (Contre-exemple). *Soit $L^2([-\pi, \pi])$ l'espace de Hilbert des fonctions de carré intégrable pour la mesure de Lebesgue sur $[-\pi, \pi]$. Comme le montre l'exemple A.1.11, l'ensemble des fonctions continues sur $[-\pi, \pi]$ est un sous-espace vectoriel de $L^2([-\pi, \pi])$ mais n'est pas fermé.*

Définition A.1.18 (Sous espace engendré par un sous-ensemble). *Soit \mathcal{X} un sous-ensemble de \mathcal{H} . Nous notons $\text{Vect}(\mathcal{X})$ le sous-espace vectoriel des combinaisons linéaires finies d'éléments de \mathcal{X} et $\overline{\text{Vect}(\mathcal{X})}$ la fermeture de $\text{Vect}(\mathcal{X})$ dans \mathcal{H} .*

Définition A.1.19 (Orthogonalité). *Deux vecteurs $x, y \in \mathcal{H}$ sont dit orthogonaux, si $\langle x, y \rangle = 0$, ce que nous notons $x \perp y$. Si \mathcal{S} est un sous-ensemble de \mathcal{H} , la notation $x \perp \mathcal{S}$, signifie que $x \perp s$ pour tout $s \in \mathcal{S}$. Nous notons $\mathcal{S} \perp \mathcal{T}$ si tout élément de \mathcal{S} est orthogonal à tout élément de \mathcal{T} .*

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

Définition A.1.20 (Complément orthogonal). *Soit \mathcal{E} un sous-ensemble d'un espace de Hilbert \mathcal{H} . On appelle ensemble orthogonal de \mathcal{E} , l'ensemble défini par :*

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

A.2 Projection et principe d'orthogonalité

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

Théorème A.2.1. *Soit \mathcal{E} est un sous-espace fermé d'un espace de Hilbert \mathcal{H} et soit x un élément quelconque de \mathcal{H} , alors :*

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

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

(ii) *$\text{proj}(x|\mathcal{E}) \in \mathcal{E}$ et $\|x - \text{proj}(x|\mathcal{E})\| = \inf_{w \in \mathcal{E}} \|x - w\|$ si et seulement si $\text{proj}(x|\mathcal{E}) \in \mathcal{E}$ et $x - \text{proj}(x|\mathcal{E}) \perp \mathcal{E}$.*

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

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

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

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

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

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

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

$$0 \leq \|y - z\|^2 = -4\|(y + z)/2 - x\|^2 + 2\|x - y\|^2 + 2\|x - z\|^2 \leq -4h^2 + 2h^2 + 2h^2 = 0$$

où nous avons utilisé que $(y + z)/2 \in \mathcal{E}$ et que $\|(y + z)/2 - x\|^2 \geq h^2$. Il s'en suit que $y = z$. \hat{x} est appelé la projection orthogonale de x sur \mathcal{E} .

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

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

et donc $u = \hat{x}$. Réciproquement supposons que $u \in \mathcal{E}$ et $x - u \not\perp \mathcal{E}$. Alors choisissons $y \in \mathcal{E}$ tel que $\|y\| = 1$ et tel que $c = \langle x - u, y \rangle \neq 0$ et notons $\tilde{x} = u + cy \in \mathcal{E}$. On a :

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

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

Proposition A.2.2. Soit \mathcal{H} un espace de Hilbert et $\text{proj}(\cdot | \mathcal{E})$ la projection orthogonale sur le sous-espace fermé \mathcal{E} . On a :

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

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

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

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

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

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

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

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

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

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

Exemple A.2.3 (Projection sur un vecteur). Soit \mathcal{H} un espace de Hilbert, $\mathcal{C} = \text{Vect}(v)$ le sous-espace engendré par un vecteur $v \in \mathcal{H}$ et x un vecteur quelconque de \mathcal{H} . On a alors $\text{proj}(x | \mathcal{C}) = \alpha v$ avec $\alpha = \langle x, v \rangle / \|v\|^2$. Si on note $\epsilon = x - \text{proj}(x | \mathcal{C})$, on a :

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

Appliquons ce résultat à $\mathcal{H} = \mathbb{C}^n$ et au vecteur $v(\lambda_0)$ de composantes $v_t = n^{-1/2} e^{i\lambda_0 t}$ où $t \in \{1, \dots, n\}$ et où la pulsation de Fourier $\lambda_0 \in (-\pi, \pi)$. On vérifie que $\|v(\lambda_0)\| = 1$. Soit $x = (x_1, \dots, x_n)^T$ un vecteur quelconque de \mathbb{C}^n . La projection orthogonale de x sur $\text{Vect}(v(\lambda_0))$ s'écrit $\alpha v(\lambda_0)$ avec :

$$\alpha = \sum_{t=1}^n x_t v_t^* = \frac{1}{\sqrt{n}} \sum_{t=1}^n x_t e^{-i\lambda_0 t}$$

qui est la transformée de Fourier à temps discret de la suite x_t calculée précisément à la pulsation λ_0 .

Théorème A.2.4. Si \mathcal{E} est un sous-ensemble d'un espace de Hilbert \mathcal{H} , alors \mathcal{E}^\perp est un sous-espace fermé.

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

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

et donc $x \in \mathcal{E}^\perp$. □

Lemme A.2.5. 1. Soit (\mathcal{M}_n) une suite croissante de sous-espaces vectoriels (s.e.v.) fermés d'un espace de Hilbert \mathcal{H} et notons $\mathcal{M}_{-\infty} = \bigcap_n \mathcal{M}_n$. Alors, pour tout $h \in \mathcal{H}$, nous avons

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

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

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

3. Soit $\{e_k, k \in \mathbb{N}\}$ une famille orthonormale de \mathcal{H} , $e_j \perp e_k$, for $j \neq k$, $\|e_j\| = 1$. Soit $\mathcal{E}_n = \text{Vect}(e_l, 0 \leq l \leq n)$ et $\mathcal{E}_\infty = \overline{\bigcup_{n \geq 0} \mathcal{E}_n}$. Alors

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

Démonstration. (a) Comme \mathcal{M}_n est un s.e.v. fermé de \mathcal{H} et donc $\mathcal{M}_{-\infty}$ est un s.e.v. fermé de \mathcal{H} . Le théorème de projection A.2.1 prouve que $\text{proj}(h| \mathcal{M}_{-\infty})$ existe. Pour $m < n$, définissons $\mathcal{M}_n \ominus \mathcal{M}_m$ le complément orthogonal de \mathcal{M}_m dans \mathcal{M}_n , c'est à dire l'ensemble des vecteurs $x \in \mathcal{M}_n$ tel que $x \perp \mathcal{M}_m$. $\mathcal{M}_n \ominus \mathcal{M}_m$ est un s.e.v. fermé de \mathcal{H} . Notons que

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

On a, pour tout $m \geq 0$,

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

et donc la suite $\{\text{proj}(h| \mathcal{M}_n), n = 0, -1, -2, \dots\}$ est une suite de Cauchy. Comme \mathcal{H} est complet, $\text{proj}(h| \mathcal{M}_n)$ converge dans \mathcal{H} . Notons $z \stackrel{\text{def}}{=} \lim_{m \rightarrow -\infty} \text{proj}(h| \mathcal{M}_n)$. Il reste à prouver que $z = \text{proj}(h| \mathcal{M}_{-\infty})$. En appliquant le théorème de projection A.2.1, nous devons donc démontrer que (i) $z \in \mathcal{M}_{-\infty}$ et (ii) $h - z \perp \mathcal{M}_{-\infty}$. Comme $\text{proj}(h| \mathcal{M}_n) \in \mathcal{M}_p$ pour tout $n \leq p$, nous avons donc $\lim_{n \rightarrow -\infty} \text{proj}(h| \mathcal{M}_n) \in \mathcal{M}_p$ pour tout p et donc $z \in \mathcal{M}_{-\infty}$, ce qui établit (i). Pour prouver (ii), prenons $p \in \mathcal{M}_{-\infty}$. Nous avons $p \in \mathcal{M}_n$ pour tout $n \in \mathbb{Z}$, et donc, pour tout $n \in \mathbb{Z}$, $\langle h - \text{proj}(h| \mathcal{M}_n), p \rangle = 0$ et (ii) découle de la continuité du produit scalaire. La preuve du point [(b)]

est similaire et est laissée au lecteur à titre d'exercice. Nous prouvons finalement le point [(c)]. En appliquant [(b)], nous avons

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

On vérifie aisément que

$$\text{proj}(h|\mathcal{E}_n) = \sum_{k=1}^n \langle h, e_k \rangle e_k.$$

Notons en effet que $\text{proj}(h|\mathcal{E}_n) \in \mathcal{E}_n$ et, pour tout $k \in \{1, \dots, n\}$,

$$\langle h - \text{proj}(h|\mathcal{E}_n), e_k \rangle = \langle h, e_k \rangle - \langle h, e_k \rangle = 0.$$

On conclut la preuve en combinant les deux résultats précédents. □

A.3 Bases Hilbertiennes

Définition A.3.1 (Famille orthonormale). *Soit $E = \{e_j; j \in T\}$ un sous ensemble de \mathcal{H} . On dit que E est une famille orthonormale ssi $\langle e_i, e_j \rangle = 1$ si $i = j$ et 0 sinon.*

Théorème A.3.2 (Inégalité de Bessel). *Si x est un vecteur d'un espace de Hilbert \mathcal{H} et si $E = \{e_1, \dots, e_k\}$ est une famille orthonormale finie, alors :*

$$\left\| x - \sum_{i=1}^k \langle x, e_i \rangle e_i \right\|^2 = \|x\|^2 - \sum_{i=1}^k |\langle x, e_i \rangle|^2$$

Démonstration. Notons $\mathcal{E} = \overline{\text{Vect}(E)}$ le sous-espace engendré par les vecteurs $\{e_1, \dots, e_k\}$. Nous avons $\|\text{proj}(x|\mathcal{E})\| \leq \|x\|$. On vérifie aisément que $\text{proj}(x|\mathcal{E}) = \sum_{i=1}^k \langle x, e_i \rangle e_i$ et que $\|\text{proj}(x|\mathcal{E})\|^2 = \sum_{i=1}^k |\langle x, e_i \rangle|^2$. Remarquons en effet, pour tout $j \in \{1, \dots, k\}$,

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

□

Si (e_i) est une suite de vecteurs orthonormaux, l'inégalité de Bessel implique $\sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2 \leq \|x\|^2$ et par conséquent, pour tout $x \in \mathcal{H}$, $\lim_{n \rightarrow \infty} \langle x, e_n \rangle = 0$ (la suite de vecteurs (e_n) converge *faiblement* vers 0). L'inégalité $\sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2 \leq \|x\|^2$ montre que la série $\sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$ est convergente pour tout $x \in \mathcal{H}$. Par conséquent, la suite $(\langle x, e_i \rangle)$ est un élément de l'espace ℓ^2 des suites de carrés sommables. La suite de vecteurs orthonormaux (e_i) de \mathcal{H} induit donc une application de \mathcal{H} dans ℓ^2 . Le développement

$$\sum_{i=1}^{\infty} \langle x, e_i \rangle e_i \tag{A.3}$$

est appelé le *développement de Fourier généralisé* de x ; les coefficients $\langle x, e_i \rangle$ sont appelés *coefficients de Fourier* par rapport à la suite orthonormale (e_i) . Comme le montre le théorème A.3.2, ces coefficients définissent la projection du vecteur x sur les sous-espaces vectoriels de \mathcal{H} engendré par un sous-ensemble fini de vecteurs (e_i) . Le théorème suivant montre que le développement (A.3) est convergent.

Théorème A.3.3. Soit (e_i) une suite orthonormale d'un espace de Hilbert \mathcal{H} et soit (α_i) une suite de nombre complexes. La série $\sum_{i=1}^{\infty} \alpha_i e_i$ converge si et seulement si $\sum_{i=1}^{\infty} |\alpha_i|^2 < \infty$ auquel cas

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

Démonstration. Pour tout $m > k > 0$, nous avons,

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

Comme $\sum_{i=1}^{\infty} |\alpha_i|^2 < \infty$, la suite $s_m = \sum_{i=1}^m \alpha_i e_i$ est une suite de Cauchy dans \mathcal{H} . Comme \mathcal{H} est complet, cette suite de Cauchy converge.

Réciproquement, si la série $\sum_{i=1}^{\infty} \alpha_i e_i$ converge, alors la suite $\sigma_m = \sum_{i=1}^m |\alpha_i|^2$ est une suite de Cauchy dans \mathbb{R} ; cette série est convergente. \square

Il faut prendre garde toutefois au fait que si la série $\sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$ converge, sa limite n'est pas nécessairement égale à x .

Exemple A.3.4. Considérons $\mathcal{H} = L^2(\mathbb{T})$ et soit $e_n(t) = \pi^{-1/2} \sin(nt)$ pour $n = 1, 2, \dots$. La suite (e_n) est orthonormale dans \mathcal{H} , mais pour $x(t) = \cos(t)$, nous avons

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

Définition A.3.5 (Suite orthonormale complète). Soit (e_i) une suite orthonormale de \mathcal{H} . On dit que (e_i) est une suite orthonormale complète si et seulement si pour tout $x \in \mathcal{H}$, $x = \sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$.

Il est important de rappeler que l'identité $x = \sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$ signifie que

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

Par exemple, si $\mathcal{H} = L^2(\mathbb{T})$ et (e_n) est une famille orthonormale de \mathcal{H} , alors l'égalité $x = \sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$ signifie que

$$\lim_{n \rightarrow \infty} \int_{\mathbb{T}} \left| x(t) - \sum_{i=1}^n \alpha_i e_i(t) \right|^2 dt = 0, \quad \text{avec} \quad \alpha_i = \int_{\mathbb{T}} x(t) e_i(t) dt.$$

Cette identité n'implique pas en général la convergence ponctuelle, nous n'avons pas en général $x(t) = \sum_{i=1}^{\infty} \alpha_i e_i(t)$.

Définition A.3.6 (Bases orthonormales). Une famille orthonormale B d'un espace de Hilbert \mathcal{H} est appelée une base orthonormale si tout élément $x \in \mathcal{H}$ a une représentation unique

$$x = \sum_{i=1}^{\infty} \alpha_i e_i,$$

où $\alpha_i \in \mathbb{C}$ et (e_i) est une suite d'éléments de B distincts.

remarquons qu'une suite orthonormale complète (e_i) est une base orthonormale. Il suffit en effet de montrer l'unicité de la décomposition : En effet, soit

$$x = \sum_{i=1}^{\infty} \alpha_i e_i \quad \text{et} \quad x = \sum_{i=1}^{\infty} \beta_i e_i.$$

En appliquant le théorème A.3.3, nous avons :

$$0 = \|x - x\|^2 = \left\| \sum_{i=1}^{\infty} \alpha_i e_i - \sum_{i=1}^{\infty} \beta_i e_i \right\|^2 = \sum_{i=1}^{\infty} |\alpha_i - \beta_i|^2.$$

Théorème A.3.7. Une suite orthonormale (e_i) de l'espace de Hilbert \mathcal{H} est complète si et seulement si

$$(\langle x, e_i \rangle = 0 \forall i \in \mathbb{N}) \Leftrightarrow x = 0.$$

Démonstration. Supposons que (e_i) est une suite orthonormale complète. Pour tout $x \in \mathcal{H}$, nous avons

$$x = \sum_{i=0}^{\infty} \langle x, e_i \rangle e_i,$$

et la condition $\langle x, e_i \rangle = 0$ pour tout $i \in \mathbb{N}$ implique que $x = 0$. Réciproquement, soit $x \in \mathcal{H}$ et considérons

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

Comme $\sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2 < \infty$ par l'inégalité de Bessel, la suite $y_n = \sum_{i=1}^n \langle x, e_i \rangle e_i$ est une suite de Cauchy dans \mathcal{H} et y est la limite de la suite (y_n) dans \mathcal{H} . Nous avons

$$\begin{aligned} \langle x - y, e_i \rangle &= \langle x, e_i \rangle - \left\langle \sum_{k=1}^{\infty} \langle x, e_k \rangle e_k, e_i \right\rangle = \\ &= \langle x, e_i \rangle - \sum_{k=1}^n \langle x, e_k \rangle \langle e_k, e_i \rangle = \langle x, e_i \rangle - \langle x, e_i \rangle = 0, \end{aligned}$$

et donc, pour tout $i \in \mathbb{N}$, $\langle x - y, e_i \rangle = 0$. Par hypothèse, cette condition implique que $x - y = 0$ et donc (e_i) est une suite orthonormale complète. \square

Théorème A.3.8 (Formule de Parseval). *Une suite orthonormale (e_i) dans un espace hilbertien \mathcal{H} est complète si et seulement si*

$$\|x\|^2 = \sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2, \quad (\text{A.4})$$

pour tout $x \in \mathcal{H}$.

Démonstration. Soit $x \in \mathcal{H}$. En utilisant le théorème A.3.2, nous avons pour tout $n \in \mathbb{N}$,

$$\left\| x - \sum_{i=1}^n \langle x, e_i \rangle e_i \right\|^2 = \|x\|^2 - \sum_{i=1}^n |\langle x, e_i \rangle|^2. \quad (\text{A.5})$$

Si (e_i) est une suite orthonormale complète, $\lim_{n \rightarrow \infty} \|x - \sum_{i=1}^n \langle x, e_i \rangle e_i\|^2 = 0$, ce qui implique

$$\lim_{n \rightarrow \infty} \left[\|x\|^2 - \sum_{i=1}^n |\langle x, e_i \rangle|^2 \right] = 0,$$

ce qui prouve (A.4). Réciproquement, si $\lim_{n \rightarrow \infty} [\|x\|^2 - \sum_{i=1}^n |\langle x, e_i \rangle|^2] = 0$, Eq. A.5 montre que

$$\lim_{n \rightarrow \infty} \left\| x - \sum_{i=1}^n \langle x, e_i \rangle e_i \right\|^2 = 0,$$

ce qui montre que suite (e_i) est complète. □

Exemple A.3.9 (base de Fourier). *Le système de fonctions*

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

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

Définition A.3.10 (Espace de Hilbert séparable). *On dit qu'un espace de Hilbert est séparable si et seulement si il existe une suite orthonormale complète.*

La plupart des espaces de Hilbert que nous rencontrerons seront séparables. En particulier le sous-espace fermé engendré par une famille dénombrable d'un espace de Hilbert, que celui-ci soit séparable ou non séparable, est séparable.

Théorème A.3.11. *Un espace de Hilbert \mathcal{H} séparable contient un sous-ensemble dénombrable dense.*

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

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

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

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

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

Théorème A.3.12. *Toute famille orthogonale d'un espace de Hilbert séparable est dénombrable.*

La preuve de ce résultat est laissée au lecteur.

A.3.1 Isométries et isomorphismes d'espaces de Hilbert

Définition A.3.13. *Isométrie Soient \mathcal{H} et \mathcal{I} deux espaces de Hilbert complexes et \mathcal{G} un sous-espace vectoriel de \mathcal{H} . Une isométrie S de \mathcal{G} dans \mathcal{I} est une application linéaire $S : \mathcal{G} \rightarrow \mathcal{I}$ telle que $\langle Sv, Sw \rangle_{\mathcal{I}} = \langle v, w \rangle_{\mathcal{H}}$ pour tout $(v, w) \in \mathcal{G}$.*

On vérifie aisément qu'une isométrie entre deux espaces de Hilbert est automatiquement linéaire.

Définition A.3.14 (isomorphisme d'espaces de Hilbert). *Un espace de Hilbert \mathcal{H} est isomorphe à un espace de Hilbert \mathcal{I} s'il existe une isométrie bijective T de \mathcal{H} dans \mathcal{I}*

Théorème A.3.15. *Soit \mathcal{H} un espace de Hilbert séparable.*

- (a) *Si \mathcal{H} est infini dimensionnel, alors il est isomorphe à ℓ^2 .*
- (b) *Si \mathcal{H} est de dimension fini, alors in est isomorphe à \mathbb{C}^n .*

Démonstration. Soit (e_i) une suite orthonormale complète de \mathcal{H} . Si \mathcal{H} est infini-dimensionnel, alors (e_i) est une suite infinie. Soit $x \in \mathcal{H}$. Pour $x \in \mathcal{H}$, définissons $Tx = (\alpha_i)$, où $\alpha_i = \langle x, e_i \rangle$. Le théorème A.3.3 montre que T est une isométrie de \mathcal{H} dans ℓ^2 . \square

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

Le résultat suivant permet de construire facilement des isométries.

Théorème A.3.16. *Soit $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ et $(\mathcal{I}, \langle \cdot, \cdot \rangle_{\mathcal{I}})$ deux espaces de Hilbert complexes. Soit \mathcal{G} un sous-espace vectoriel de \mathcal{H} .*

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

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

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

Par construction $\bar{S}(\bar{\mathcal{G}})$ est inclus dans l'adhérence de $S(\mathcal{G})$. Inversement, soit $w \in \overline{S(\mathcal{G})}$. Il existe une suite $(v_n) \in \mathcal{G}$ telle que $w = \lim_{n \rightarrow \infty} Sv_n$. La suite (Sv_n) est de Cauchy et comme S est une

isométrie, la suite (v_n) est aussi de Cauchy dans \mathcal{G} . Soit $v \in \bar{\mathcal{G}}$ sa limite. Nous avons $\bar{S}v = \lim_{n \rightarrow \infty} S v_n$ et donc $\bar{S}v = w$, ce qui montre $\overline{S(\mathcal{G})} \subseteq \bar{S}(\bar{\mathcal{G}})$. Ceci établit le point (i) de la proposition.

Notons par \mathcal{G} l'ensemble des combinaisons linéaires finies, à coefficients complexes, d'éléments de $\{v_t, t \in \mathbb{T}\}$:

$$\mathcal{G} = \left\{ v = \sum_{t \in I} a_t v_t, (a_t)_{t \in I} \in \mathbb{C}^I, I \in \mathcal{P}_F(\mathbb{T}) \right\},$$

où $\mathcal{P}_F(\mathbb{T})$ est l'ensemble des parties finies de \mathbb{T} .

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

$$\left\| \sum_{t \in I} a_t v_t \right\|_{\mathcal{H}} = \left\| \sum_{t \in I} b_t w_t \right\|_{\mathcal{I}}. \quad (\text{A.6})$$

Pour $f \in \mathcal{G}$, $f = \sum_{t \in I} a_t v_t$, $I \in \mathcal{P}_F(\mathbb{T})$, posons $Sf = \sum_{t \in I} a_t w_t$. Eq. A.6 montre que même si la représentation $\sum_{t \in I} a_t v_t$ n'est pas unique, Sf est unique (si nous avons $\sum_{t \in I} a_t v_t = \sum_{t \in I'} a'_t v_t$, (A.6) montre que $\sum_{t \in I} a_t w_t = \sum_{t \in I'} a'_t w_t$). Nous avons clairement, pour tout $f, g \in \mathcal{G}$,

$$\langle Sf, Sg \rangle_{\mathcal{I}} = \langle f, g \rangle_{\mathcal{H}},$$

ce qui montre que $S : \mathcal{G} \rightarrow \mathcal{I}$ est une isométrie. Cette isométrie, en vertu de i se prolonge de façon unique en une isométrie $\bar{S} : \bar{\mathcal{G}} \rightarrow \mathcal{I}$ telle que $\bar{S}(\bar{\mathcal{G}}) = \overline{S(\mathcal{G})}$. Par construction, $\bar{\mathcal{G}} = \overline{\text{Vect}(v_t, t \in \mathbb{T})}$ et $S(\mathcal{G}) = \text{Vect}(w_t, t \in \mathbb{T})$. \square

S

Annexe B

Statistique Asymptotique

B.1 Notions de convergence

Soit $\{X_n\}_{n \geq 0}$ une suite de vecteurs aléatoires à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ et définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Nous notons, pour $x, y \in \mathbb{R}^d$, $d(x, y) = \|x - y\|$ la distance euclidienne.

Nous donnons les trois notions de convergence pour une suite de v.a. $\{X_n\}_{n \geq 0}$ dans l'ordre *croissant*, c'est-à-dire de la notion la plus faible à la plus forte (l'implication d'une convergence à une autre est établie au théorème B.7).

Définition B.1.1. *Nous dirons que la suite $\{X_n\}_{n \geq 0}$ converge en loi vers X et nous noterons*

$$X_n \xrightarrow{\mathcal{L}} X,$$

si pour toute fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$ continue bornée,

$$\lim_n \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)].$$

Il serait plus approprié de définir la convergence en loi uniquement à partir des lois :

Définition B.1.2. *Nous dirons que la suite de probabilités $\{\mu_n\}$ définies sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ converge étroitement vers la probabilité μ définie sur $(X, \mathcal{B}(X))$ et nous noterons*

$$\mu_n \rightsquigarrow \mu,$$

si pour toute fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$ continue bornée,

$$\lim_{n \rightarrow \infty} \mu_n(f) = \mu(f),$$

où pour une fonction f intégrable et une mesure ν , $\nu(f) = \int f(x)\nu(dx)$.

Il est donc clair que $X_n \xrightarrow{\mathcal{L}} X$ est équivalent à $P_{X_n} \rightsquigarrow P_X$, où l'on a noté P_{X_n} et P_X les probabilités images de X_n et X . On voit à cette occasion qu'il n'est pas nécessaire de définir les v.a. $\{X_n\}_{n \geq 0}$ et X sur le même espace de probabilité pour définir la convergence en loi.

Il existe divers caractérisations de cette convergence très utiles en pratique qui seront abordées dans la section B.7 (voir le Lemme B.3.1 et le théorème B.3.3).

Définition B.1.3. Nous dirons que la suite $\{X_n\}_{n \geq 0}$ converge en probabilité vers X et nous noterons

$$X_n \xrightarrow{\mathbb{P}} X ,$$

si pour tout $\epsilon > 0$, nous avons :

$$\lim_{n \rightarrow \infty} \mathbb{P}(d(X_n, X) > \epsilon) = 0 .$$

Au contraire de la convergence en loi, cette définition utilise que X_n et X sont définies sur le même espace de probabilité ; autrement, on ne serait calculer la loi de $d(X_n, X)$. Nous verrons les nombreuses relations qui existent entre convergence en loi et convergence en probabilité dans la section B.7. Notons cependant qu'il est un cas où ces définitions coïncident. Si X est déterministe, i.e. s'il existe $c \in \mathbb{R}^d$ tel que $\mathbb{P}(X = c) = 1$, alors il est facile de voir que $X_n \xrightarrow{\mathbb{P}} X$ (on notera plutôt $X_n \xrightarrow{\mathbb{P}} c$ dans ce cas très particulier) est équivalent à $X_n \xrightarrow{\mathcal{L}} X$ (voir théorème B.7-(ii)). Il n'y a pas de contradiction avec la remarque précédente, les loi v.a. déterministes sont définies sans ambiguïté indépendamment de l'espace de probabilité sur lesquelles elles sont définies ; il s'en suit que pour X constante la loi de $d(X_n, X)$ est connue uniquement à partir de la loi de X_n .

La proposition B.5.1 peut bien entendu se formuler sous une forme plus générale, en choisissant f telle que $\mathbb{E}[|f(X_n)|] < \infty$, et en l'appliquant à la suite définie par $Y_n = f(X_n)$ pour $n > 0$. Une formulation plus élégante (et plus précise) peut être obtenue en introduisant la définition suivante.

Définition B.1.4. Nous dirons que la suite $\{X_n\}_{n \geq 0}$ converge presque-sûrement vers X et nous noterons

$$X_n \xrightarrow{\mathbb{P}\text{-p.s.}} X ,$$

si :

$$\mathbb{P}(\limsup_n d(X_n, X) = 0) = 1 .$$

Cette fois-ci, il est nécessaire de définir $\{X_n\}_{n \geq 0}$ et X sur le même espace de probabilité sinon la loi de $\limsup_{n \rightarrow \infty} d(X_n, X)$ n'a pas de sens. Une méthode essentielle permettant de prouver une telle convergence est le Lemme de Borel-Cantelli :

Lemme B.1.5 (Lemme de Borel-Cantelli). Soit $\{A_n, n \in \mathbb{N}\}$ une suite d'événements tels que $\sum_{n \geq 0} \mathbb{P}(A_n) < \infty$. Alors il existe une variable aléatoire N à valeurs dans \mathbb{N} telle que, $\mathbb{P}(A_n \subseteq \{n \leq N\}) = 1$. De plus,

$$\mathbb{P}\left(\sum_{n \geq 0} \mathbb{1}(A_n) < \infty\right) = 1. \tag{B.1}$$

Démonstration. Ce lemme est une application bien connue de la théorie des probabilités. Remarquons que $\left(\bigcup_{n \geq m} A_n\right)_m$ est une suite décroissante d'événements dont l'intersection est précisément l'événement $\left\{\sum_{n \geq 0} \mathbb{1}(A_n) = \infty\right\}$. Or $\mathbb{P}\left(\bigcup_{n \geq m} A_n\right) \leq \sum_{n \geq m} \mathbb{P}(A_n)$ qui tend vers zéro quand $m \rightarrow \infty$ par hypothèse. D'où (B.1). Observons alors qu'il s'en suit que pour tout ω dans un ensemble Ω^* de probabilité un, la plus grande valeur de n telle que $\omega \in A_n$ est un entier fini. Notons $N(\omega)$ ce nombre et complétons la définition de cette variable par zéro en dehors Ω^* . Le résultat est alors obtenu. \square

On en déduit facilement que s'il existe une suite de nombres positifs (ϵ_n) telle que $\lim \epsilon_n = 0$ et $\sum_{n \in \mathbb{N}} \mathbb{P}\{|X_n| \geq \epsilon_n\} < \infty$, alors $X_n \xrightarrow{\mathbb{P}\text{-p.s.}} 0$.

B.2 Suites tendues

Introduisons la notion de *tension* pour une variable aléatoire.

Définition B.2.1. Une v.a. X à valeur dans un espace topologique quelconque est dite *tendue* si pour tout $\epsilon > 0$, il existe un compact K tel que $\mathbb{P}(X \notin K) \leq \epsilon$. Une famille $\{X_\alpha, \alpha \in A\}$ est dite *uniformément tendue* (ou *bornée en probabilité*) si, pour tout $\epsilon > 0$, il existe un compact K tel que $\sup_{\alpha \in A} \mathbb{P}(X_\alpha \notin K) \leq \epsilon$.

Toute v.a. à valeurs dans \mathbb{R}^d équipé de la topologie associée à la distance euclidienne est *tendue* puisque $\bigcap_{n \geq 1} \{\|x\| \geq n\} = \emptyset$. De façon générale toute famille finie de variables aléatoires réelles est tendue. Il s'en suit qu'une suite $(X_n)_{n \in \mathbb{N}}$ de v.a. à valeurs dans \mathbb{R}^d est bornée en probabilité si, pour tout $\epsilon > 0$, il existe $M > 0$ tel que

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\|X_n\| > M) \leq \epsilon.$$

Un critère simple de tension uniforme pour une famille $\{X_\alpha, \alpha \in A\}$ de v.a. à valeurs dans \mathbb{R}^d est obtenu comme suit. Si, pour $p > 0$, nous avons $\sup_{\alpha} \mathbb{E} [|X_\alpha|^p] < \infty$, l'inégalité de Markov montre que :

$$\mathbb{P}(\|X_\alpha\| \geq M) \leq M^{-1} \mathbb{E} [|X_\alpha|^p],$$

et donc que $\{X_\alpha\}$ est uniformément tendue. Plus généralement la notion de tension est intimement liée à la notion de convergence en loi. Le résultat suivant montre en effet que toute suite convergente en loi est tendue et qu'il existe une réciproque (partielle) à ce résultat.

Théorème B.2.2 (Prohorov dans \mathbb{R}^d). Soit $\{X_n\}$ une suite de v.a. de \mathbb{R}^d .

- (i) Si $X_n \xrightarrow{\mathcal{L}} X$, alors la famille $\{X_n, n \in \mathbb{N}\}$ est uniformément tendue,
- (ii) Si la famille $\{X_n, n \in \mathbb{N}\}$ est uniformément tendue, alors il existe une sous-suite telle que $X_{n_j} \xrightarrow{\mathcal{L}} X$ pour une v.a. X .

Démonstration. (i) : Si $X_n \xrightarrow{\mathcal{L}} X$, on, pour tout M , $\limsup_n \mathbb{P}(\|X_n\| \geq M) \leq \limsup_n \mathbb{E} [\phi_M(\|X_n\|)] = \mathbb{E} [\phi_M(\|X\|)] \leq \mathbb{P}(\|X\| \geq M - 1)$, où ϕ_M est une fonction continue bornée sur \mathbb{R} telle que $\mathbb{1}_{[M, \infty)} \leq \phi_M \leq \mathbb{1}_{[M-1, \infty)}$. Donc pour tout $\epsilon > 0$, on peut trouver $M > 0$ et $N \in \mathbb{N}$ tels que $\mathbb{P}(\|X_n\| \geq M) \leq \epsilon$ pour tout $n \geq N$. En augmentant M suffisamment, on obtient que $\mathbb{P}(\|X_n\| \geq M) \leq \epsilon$ pour tout $n \in \{0, \dots, N\}$.

(ii) : Soit $\{g_k\}_{k \in \mathbb{N}}$ une suite de fonction mesurables bornées de \mathbb{R}^d dans \mathbb{R} . Pour tout k la suite $\{\mathbb{E} [g_k(X_n)]\}_{n \geq 0}$ est bornée par $\sup |g_k|$ et on peut donc extraire une sous-suite convergente de chacune de ses sous-suites. En procédant itérativement on construit une suite de sous-suites $\{n_{k,l}\}$ telle que $\{n_{k+1,l}\} \subset \{n_{k,l}\}$ et $\{\mathbb{E} [g_k(X_{n_{k,l}})]\}_{l \geq 0}$ converge pour tout k . On définit alors récursivement une suite d'entiers $\{n_\ell\}_{\ell \geq 0}$ par

$$n_0 = 1 \text{ et, pour tout } \ell \in \mathbb{N}, n_{\ell+1} = \inf\{n_{\ell+1,l} > n_\ell : l \in \mathbb{N}\}.$$

Par construction, la suite $\{n_\ell\}_{\ell \geq 0}$ est croissante et, pour tout k , à partir d'un certain rang, c'est une sous-suite de la suite $\{n_{k,l}\}_l$. On a donc trouvé pour toute suite $\{g_k\}_{k \in \mathbb{N}}$ de fonctions mesurables bornées de \mathbb{R}^d dans \mathbb{R} une sous-suite $\{n_j\}_{j \geq 0}$ telle que $(\mathbb{E} [g_k(X_{n_j})])_{j \geq 0}$ converge pour tout k .

Choisissons pour cette suite une suite dense dans $C_0(\mathbb{R}^d)$ l'ensemble des fonctions continues à support compact muni de la norme uniforme (voir le lemme B.2.3 ci-dessous). Il s'en suit que pour tout $f \in C_0(\mathbb{R}^d)$, $\{\mathbb{E}[f(X_{n_j})]\}_{j \geq 0}$ converge. Notons $u(f)$ sa limite. Il vient par passage à la limite que $f \mapsto u(f)$ est une fonctionnelle linéaire continue de $C_0(\mathbb{R}^d)$ muni de la norme uniforme dans \mathbb{R} . Elle est de plus positive ($u(f) \geq 0$ si f est une fonction positive). Une telle forme linéaire est une mesure de Radon et l'on sait par le célèbre théorème de représentation de Riesz qu'il existe une mesure positive μ finie sur tout compact de \mathbb{R}^d telle que $u(f) = \mu f$.

Remarquons à ce stade que nous n'avons toujours pas utilisé d'hypothèses sur la suite $\{X_n\}$. La tension uniforme de cette suite va en fait nous permettre de montrer que, d'une part, μ est une mesure de probabilité et d'autre part que $(\mathbb{E}[f(X_{n_l})])_l$ converge aussi vers μf si f est continue mais pas nécessairement à support compact, ce qui conclura la preuve de ce théorème.

Supposons donc que $\{X_n, n \in \mathbb{N}\}$ est uniformément tendue. Soit $(\phi_m)_{m \geq 1}$ une suite de fonction continue à valeur dans $[0, 1]$, à support dans $[-m-1, m+1]^d$ et qui vaut 1 sur $[-m, m]^d$. Alors, d'après ce qui précède, $\lim \mathbb{E}[\phi_m(X_{n_l})] = \mu \phi_m \leq 1$. D'après l'hypothèse de tension uniforme, pour tout ϵ on peut donc trouver m tel que $\mu \phi_m \geq 1 - \epsilon$. Par convergence monotone, on obtient $\mu(\mathbb{R}^d) = 1$ et donc μ est une mesure de probabilité. Soit X une v.a. de loi μ . Pour toute fonction f continue bornée de \mathbb{R}^d et tout entier $m > 0$, on peut écrire, pour toute variable Y ,

$$\mathbb{E}[f(Y)] = \mathbb{E}[(f(1 - \phi_m))(Y)] + \mathbb{E}[(f\phi_m)(Y)].$$

Il suit de la définition de ϕ_m que $|\mathbb{E}[(f(1 - \phi_m))(Y)]| \leq \sup(|f|)\mathbb{P}(|Y| \geq m)$. En appliquant cette décomposition à $Y = X_{n_l}$ et $Y = X$ et en utilisant que X est tendu et que $\{X_n, n \in \mathbb{N}\}$ est uniformément tendue, il vient, pour tout ϵ et m suffisamment grand,

$$|\mathbb{E}[f(X_{n_l})] - \mathbb{E}[f(X)]| \leq 2 \sup(|f|)\epsilon + |\mathbb{E}[(f\phi_m)(X_{n_l})] - \mathbb{E}[(f\phi_m)(X)]|.$$

La démonstration est achevée en remarquant que pour tout m , $f\phi_m$ est à support compact donc que $\lim_{l \rightarrow \infty} \mathbb{E}[(f\phi_m)(X_{n_l})] = \mathbb{E}[(f\phi_m)(X)]$ puis en prenant ϵ arbitrairement petit. \square

Lemme B.2.3. *Pour tout compact K de \mathbb{R}^d , l'ensemble $C(K)$ des fonctions continues sur K muni de la norme sup est séparable, c'est-à-dire qu'elle contient une suite dénombrable dense.*

Démonstration. Nous donnons une preuve élémentaire dans le cas $d = 1$. Pour un résultat beaucoup plus général, voir par exemple le corollaire 2.13.38 de Schwartz [1991]. Considérons, pour tout entier j l'ensemble E_j des fonctions continues dont toutes les restrictions aux intervalles $[k2^{-j}, (k+1)2^{-j}]^d$ avec $k \in \mathbb{Z}$ sont linéaires et telles que les valeurs prises en les nombres $k2^{-j}$ sont rationnels. Toute fonction continue sur un compact est uniformément continue sur ce compact. Pour l'approcher au sens de la norme sup sur K il suffit donc de l'approcher par une fonction de E_j sur la grille $\{k2^{-j}, k \in \mathbb{Z}\} \cap K$ en choisissant j suffisamment grand. L'ensemble $E_j(K)$ des fonctions de E_j restreintes à K est dénombrable et l'on trouve que $\bigcup_j E_j$ est dense dans $C(K)$ pour tout compact K . En notant E_j^0 le sous-ensemble de E_j de ses fonctions à support compact, on trouve de même que $\bigcup_j E_j^0$ est dense dans $C_0(\mathbb{R}^d)$. \square

Corollaire B.2.4. *Soit $\{X_n\}$ une suite de v.a. de \mathbb{R}^d uniformément tendue et X une v.a. de \mathbb{R}^d vérifiant : pour toute suite croissante divergente d'entiers (k_n) , si $X_{k_n} \xrightarrow{\mathcal{L}} Y$, alors $Y \sim X$. Alors $X_n \xrightarrow{\mathcal{L}} X$.*

Démonstration. D'après le théorème de Prohorov, de toute sous-suite de X_n , on peut extraire une sous-suite qui converge en loi vers X . Un raisonnement par contradiction montre que $X_n \xrightarrow{\mathcal{L}} X$: en effet, si tel n'était pas le cas, il existerait une fonction continue bornée $f : \mathbb{R}^d \rightarrow \mathbb{R}$ telle que $\mathbb{E}[f(X_n)]$ ne converge pas vers $\mathbb{E}[f(X)]$. On aurait donc une sous-suite $\{X_{n_j}\}$ et $\epsilon > 0$ tels que $|\mathbb{E}[f(X_{n_j})] - \mathbb{E}[f(X)]| > \epsilon$ pour tout $n \in \mathbb{N}$. Par suite, aucune sous-suite de $\{X_{n_j}\}$ ne convergerait vers X , d'où la contradiction. \square

B.3 Caractérisations de la convergence en loi

Le lemme de Portmanteau donne des conditions équivalentes à la convergence en loi.

Lemme B.3.1 (Portmanteau). *Soit $\{X_n\}_{n \geq 0}$ et X une suite de vecteurs aléatoires à valeurs dans \mathbb{R}^d et soit \mathbf{P}_n et \mathbf{P} leurs loi respectives. Les assertions suivantes sont équivalentes.*

- (i) $X_n \xrightarrow{\mathcal{L}} X$,
- (ii) Pour toute fonction f bornée et lipschitzienne ($|f(x) - f(y)| \leq K_f \|x - y\|$ pour tout $x, y \in \mathbb{R}^d$), $\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$,
- (iii) $\liminf \mathbf{P}_n(G) \geq \mathbf{P}(G)$ pour tout ensemble ouvert,
- (iv) $\limsup \mathbf{P}_n(F) \leq \mathbf{P}(F)$ pour tout ensemble fermé,
- (v) Pour tout ensemble borélien A tel que $\mathbf{P}(A) = \mathbf{P}(\bar{A}) = \mathbf{P}(\text{int}A)$,

$$\lim_{n \rightarrow \infty} \mathbf{P}_n(A) = \mathbf{P}(A)$$

où \bar{A} est la fermeture de A (A complété par l'ensemble des valeurs d'adhérence des suites d'éléments de A).

Démonstration. On montre une succession d'implications qui démontre l'équivalence des propositions :

1. (i) \Rightarrow (ii) est trivial puisque toute fonction lipschitzienne est continue.
2. (ii) \Rightarrow (iii) : soit G un ouvert. Pour tout m , définissons $f_m(x) = \max(1, m d(x, G^c))$. La suite $\{f_m\}$ est une suite croissante de fonctions positives bornées lipschitziennes $\{f_m\}$ qui convergence simplement vers $\mathbb{1}_G$. Par construction, $\liminf \mathbb{E}[\mathbb{1}_G(X_n)] \geq \liminf \mathbb{E}[f_m(X_n)] = \mathbb{E}[f_m(X)]$ par (ii). Le théorème de Beppo-Levi montre d'autre part $\mathbb{E}[f_m(X)]$ converge vers en croissant vers $\mathbb{E}[\mathbb{1}_G(x)]$, ce qui établit le résultat.
3. (iii) \Leftrightarrow (iv) est évident en prenant les complémentaires
4. (iii) \Leftrightarrow (iv) \Rightarrow (v) : pour tout $A \in \mathcal{B}(\mathbb{R}^d)$ vérifiant $\mathbf{P}(A) = \mathbf{P}(\bar{A}) = \mathbf{P}(\text{int}A)$, nous avons

$$\begin{aligned} \mathbf{P}(A) = \mathbf{P}(\text{int}A) &\leq \liminf \mathbf{P}_n(\text{int}A) \leq \liminf \mathbf{P}_n(A) \leq \\ &\limsup \mathbf{P}_n(A) \leq \limsup \mathbf{P}_n(\bar{A}) \leq \mathbf{P}(\bar{A}) = \mathbf{P}(A). \end{aligned}$$

D'où (v) en coinçant $\liminf \mathbf{P}_n(A)$ et $\limsup \mathbf{P}_n(A)$ entre les deux mêmes bornes.

5. (v) \Rightarrow (i) : Soit $A_M = [-M, M]^d$. Alors $\phi : M \mapsto \mathbf{P}(A_M)$ est une fonction croissante $\mathbb{R}_+ \rightarrow [0, 1]$, continue à droite, qui a un nombre fini de discontinuités plus grande qu'un $\epsilon > 0$ donné. L'ensemble de ses discontinuités de ϕ sur \mathbb{R}_+ est donc au plus dénombrable. Soit (M_n) une suite croissante divergente de points de continuité de ϕ : pour tout $\epsilon > 0$ il existe n tel que $\phi(M) = \mathbf{P}([-M, M]^d) > 1 - \epsilon$. Comme, en tout point de continuité M de ϕ , $\mathbf{P}(A_M) = \mathbf{P}(\bar{A}_M) = \mathbf{P}(\text{int}A_M)$, en utilisant (v), on obtient que $\limsup \mathbf{P}_n(A_{M_n}) > 1 - \epsilon$, puis par suite que (\mathbf{P}_n) est une suite uniformément tendue. Utilisons le corollaire B.2.4. Soit (k_n) une suite croissante divergente d'entiers telle que $X_{k_n} \xrightarrow{\mathcal{L}} Y$. Alors Pour tout $A \in \mathcal{B}(\mathbb{R}^d)$ vérifiant $\mathbf{P}(A) = \mathbf{P}(\bar{A}) = \mathbf{P}(\text{int}A)$, on a $\mathbb{P}(Y \in A) = \mathbf{P}(A)$. On définit maintenant, pour $\mathbf{x} = (x_1, \dots, x_d)$, le borélien $B_{\mathbf{x}} = (-\infty, x_1] \times \dots \times (-\infty, x_d]$. En raisonnant comme pour les A_M , on montre aisément qu'il existe un ensemble $\mathcal{C} \subseteq \mathbb{R}^d$ dense dans \mathbb{R}^d tel que, pour tout $\mathbf{x} \in \mathcal{C}$, $\mathbf{P}(B_{\mathbf{x}}) = \mathbf{P}(\text{int}B_{\mathbf{x}})$. Comme \mathcal{C} est dense, on a $\sigma(B_{\mathbf{x}}, \mathbf{x} \in \mathcal{C}) = \mathcal{B}(\mathbb{R}^d)$. On note

$$\mathcal{C}' = \{\mathbf{x} : \mathbf{P}(B_{\mathbf{x}}) = \mathbf{P}(\text{int}B_{\mathbf{x}})\}.$$

Comme \mathcal{C}' contient \mathcal{C} , on a $\sigma(B_{\mathbf{x}}, \mathbf{x} \in \mathcal{C}') = \mathcal{B}(\mathbb{R}^d)$. Comme $B_{\mathbf{x}} \cap B_{\mathbf{y}}$ a sa frontière incluse dans l'union des des frontières de $B_{\mathbf{x}}$ et $B_{\mathbf{y}}$, on voit que $\{B_{\mathbf{x}}, \mathbf{x} \in \mathcal{C}'\}$ est un π -système (*i.e.* est stable par intersection finie). Le π - λ -théorème permet donc de conclure que, la loi de Y coïncidant avec \mathbf{P} sur ce π -système, elle coïncide avec sur tous les boréliens. □

La proposition (v) implique que $\lim_n F_n(\mathbf{x}) = F(\mathbf{x})$ en tout point de continuité de F , où l'on a noté $F_n(\mathbf{x}) = \mathbb{P}(X_n \leq \mathbf{x})$ et $F(\mathbf{x}) = \mathbb{P}(X \leq \mathbf{x})$ les fonctions de répartition de X_n et X . D'après la preuve de (v) \Rightarrow (i), on voit que la réciproque est en fait vraie : si $\lim_n F_n(\mathbf{x}) = F(\mathbf{x})$ en tout point de continuité de F , alors $X_n \xrightarrow{\mathcal{L}} X$. Une conséquence immédiate est que, si la fonction de répartition de X est continue, alors $X_n \xrightarrow{\mathcal{L}} X$ est équivalent à $\mathbb{P}(X_n \leq \mathbf{x}) \rightarrow \mathbb{P}(X \leq \mathbf{x})$ pour tout $\mathbf{x} \in \mathbb{R}^d$. Pour $d = 1$, la convergence est en fait uniforme en \mathbf{x} comme l'indique le résultat suivant.

Lemme B.3.2. *Supposons, pour $d = 1$, que $X_n \xrightarrow{\mathcal{L}} X$ et que la fonction de répartition de X est continue. Alors,*

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |\mathbb{P}(X_n \leq x) - \mathbb{P}(X \leq x)| = 0.$$

Démonstration. Soit $F_n(x) = \mathbb{P}(X_n \leq x)$ et $F(x) = \mathbb{P}(X \leq x)$. F étant continue, il existe des points $-\infty^d = x_0 < x_1 < \dots < x_k = \infty^d$ tels que $F(x_i) = i/k$. F_n et F étant croissantes, nous avons, pour tout $x \in \mathbb{R}^d$, en choisissant i tel que $x_{i-1} \leq x \leq x_i$:

$$\begin{aligned} F_n(x) - F(x) &\leq F_n(x_i) - F(x_{i-1}) = F_n(x_i) - F(x_i) + 1/k \\ F_n(x) - F(x) &\geq F_n(x_{i-1}) - F(x_i) = F_n(x_{i-1}) - F(x_{i-1}) - 1/k. \end{aligned}$$

Donc $|F_n(x) - F(x)|$ est borné par $\sup_i |F_n(x_i) - F(x_i)| + 1/k$ pour tout x . Par conséquent,

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq \lim_{n \rightarrow \infty} \sup_{i \in \{0, \dots, k\}} |F_n(x_i) - F(x_i)| + 1/k = 1/k,$$

ce qui permet de conclure, en choisissant k arbitrairement grand. □

Le raisonnement s'étend sans difficulté au cas multidimensionnel mais nous omettons cet énoncé, la fonction de répartition étant mieux adaptée à la dimension un.

Comme nous l'avons déjà vu dans le lemme de Portmanteau, pour démontrer la convergence en loi, il suffit de s'intéresser à un sous-ensemble des fonctions continues bornées, par exemple, les fonctions lipschitziennes bornées, mais cette classe peut encore être réduite. Nous allons en fait démontrer dans cette partie qu'il suffit de s'intéresser à une seule fonction, la *fonction caractéristique*,

$$t \rightarrow \mathbb{E} \left[e^{it^T X} \right], \quad t \in \mathbb{R}^d.$$

Pour tout t , la fonction $x \rightarrow e^{it^T X}$ est continue et bornée. Par conséquent, si $X_n \xrightarrow{\mathcal{L}} X$, $\mathbb{E} \left[e^{it^T X_n} \right] \rightarrow \mathbb{E} \left[e^{it^T X} \right]$. Le théorème de continuité de Levy montre que la réciproque est vraie.

Théorème B.3.3 (Lévy). *Soit X_n et X des v.a. à valeurs dans \mathbb{R}^d . Alors $X_n \xrightarrow{\mathcal{L}} X$ si et seulement si $\mathbb{E} \left[e^{it^T X_n} \right] \rightarrow \mathbb{E} \left[e^{it^T X} \right]$. De plus si, pour tout t , $\mathbb{E} \left[e^{it^T X_n} \right]$ converge vers une fonction $\phi(t)$ qui est continue en 0, alors $\phi(t)$ est la fonction caractéristique d'une v.a. X et $X_n \xrightarrow{\mathcal{L}} X$.*

Démonstration. L'implication directe est évidente. Considérons la réciproque. Supposons tout d'abord que la famille $\{X_n\}$ est bornée en probabilité et utilisons le corollaire B.2.4. Soit (k_n) une suite croissante divergente d'entiers telle que $X_{k_n} \xrightarrow{\mathcal{L}} Y$. La fonction caractéristique de $\mathbb{E} \left[e^{it^T Y} \right]$ est alors la limite de $\mathbb{E} \left[e^{it^T X_{k_j}} \right]$ et donc, pour tout t , $\mathbb{E} \left[e^{it^T Y} \right] = \mathbb{E} \left[e^{it^T X} \right]$ (ou $\phi(t)$ dans le cas où l'on n'a pas supposé l'existence de X). Cela montre X et Y ont la même loi.

Il suffit maintenant de démontrer que l'hypothèse $\mathbb{E} \left[e^{it^T X_n} \right] \rightarrow \mathbb{E} \left[e^{it^T X} \right]$ implique que $\{X_n, n \in \mathbb{N}\}$ est bornée en probabilité. Pour tout x et tout $\delta > 0$ nous avons :

$$\mathbb{1}(|x| \geq 2/\delta) \leq 2 \left(1 - \frac{\sin(\delta x)}{\delta x} \right) = \frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \cos(tx)) dt.$$

En remplaçant x par X_n , en utilisant le théorème de Fubini :

$$\mathbb{P}(|X_n| \geq 2/\delta) \leq \frac{1}{\delta} \int_{-\delta}^{\delta} \operatorname{Re} \left(1 - \mathbb{E} \left[e^{it^T X_n} \right] \right) dt.$$

Par hypothèse, l'intégrande converge pour tout t vers $\operatorname{Re} \left(1 - \mathbb{E} \left[e^{it^T X} \right] \right)$, quand $n \rightarrow \infty$. Le théorème de convergence dominée implique donc :

$$\limsup \mathbb{P}(|X_n| \geq 2/\delta) \leq \frac{1}{\delta} \int_{-\delta}^{\delta} \operatorname{Re}(1 - \phi(t)) dt.$$

Comme la fonction ϕ est continue en zéro, pour tout $\epsilon > 0$, il existe δ tel que, pour tout $|t| \leq \delta$, $|\phi(t)| \leq \epsilon$. Par suite,

$$\limsup \mathbb{P}(|X_n| \geq 2/\delta) \leq 2\epsilon,$$

et donc la famille $\{X_n, n \in \mathbb{N}\}$ est bornée en probabilité. \square

La fonction caractéristique d'un vecteur aléatoire $X = (X_1, \dots, X_k)$ en $t \in \mathbb{R}^d$ peut être vue comme la fonction caractéristique de la v.a. $Y = t^T X$ évaluée au point 1, $\psi(t) = \mathbb{E} \left[e^{it^T X} \right] = \phi(1)$ où $\phi : u \in \mathbb{R} \rightarrow \mathbb{E} \left[e^{iu^T Y} \right]$. Supposons que $X_n \xrightarrow{\mathcal{L}} X$. Alors, pour tout $t \in \mathbb{R}^d$ et tout $u \in \mathbb{R}$, $\mathbb{E} \left[e^{iut^T X_n} \right] \rightarrow \mathbb{E} \left[e^{iu(t^T X)} \right]$, et donc $t^T X_n \xrightarrow{\mathcal{L}} t^T X$ par application du théorème de Levy B.3.3. Réciproquement, supposons que pour tout $t \in \mathbb{R}^d$, $t^T X_n \xrightarrow{\mathcal{L}} t^T X$. Alors $\mathbb{E} \left[e^{i(t^T X_n)} \right] \rightarrow \mathbb{E} \left[e^{i(t^T X)} \right]$ et donc $X_n \xrightarrow{\mathcal{L}} X$, encore par application du théorème de Levy. Par conséquent :

Proposition B.3.4 (Procédé de Cramér-Wold). *Soit $(X_n, n \in \mathbb{N})$ une suite de vecteurs aléatoires à valeurs dans \mathbb{R}^d . Alors :*

$$X_n \xrightarrow{\mathcal{L}} X \iff t^T X_n \xrightarrow{\mathcal{L}} t^T X, \quad \forall t \in \mathbb{R}^d.$$

Cette approche est très utilisée pour prouver la convergence en loi de vecteurs. Elle est connue sous le nom de *procédé de Cramér-Wold*. Elle permet de réduire les problèmes de convergence de vecteurs aléatoires à des problèmes de convergence de variables aléatoires.

B.4 théorème de continuité

Une propriété fondamentale des notions de convergence que l'on a introduite est qu'elles sont conservées par une transformation continue : si la suite de v.a. $\{X_n\}$ converge vers X et que g est continue, alors $g(X_n)$ converge vers $g(X)$ et le résultat est vrai pour les trois types de convergence considérés.

Théorème B.4.1 (Transformation continue). *Soit $g : \mathbb{R}^d \mapsto \mathbb{R}^m$ est continue en tout point d'un ensemble C tel que $\mathbb{P}(X \in C) = 1$. Alors,*

- (i) Si $X_n \xrightarrow{\mathcal{L}} X$, alors $g(X_n) \xrightarrow{\mathcal{L}} g(X)$,
- (ii) Si $X_n \xrightarrow{\mathbb{P}} X$, alors $g(X_n) \xrightarrow{\mathbb{P}} g(X)$,
- (iii) Si $X_n \xrightarrow{\mathbb{P}\text{-p.s.}} X$, alors $g(X_n) \xrightarrow{\mathbb{P}\text{-p.s.}} g(X)$.

Démonstration. (i) : Par définition, $\{g(X_n) \in F\} = \{X_n \in g^{-1}(F)\}$. Pour tout fermé F , on a :

$$g^{-1}(F) \subset \overline{g^{-1}(F)} \subset (g^{-1}(F) \cup C^c).$$

En effet, seule la seconde inclusion est non-triviale ; soit $x \in \overline{g^{-1}(F)}$ et montrons qu'ou bien $x \in C^c$ (le complémentaire de C) ou bien $x \in g^{-1}(F)$. Il existe une suite x_m d'éléments de $g^{-1}(F)$ telle que $x_m \rightarrow x$. Si $x \in C$, $g(x_m) \rightarrow g(x)$, car g est continue au point x , et comme $g(x_m) \in F$ et F est fermé, $g(x) \in F$, ce qui implique que $x \in g^{-1}(F)$. On déduit de cette série d'inclusions et du lemme de Portmanteau, comme $X_n \xrightarrow{\mathcal{L}} X$ et $\mathbb{P}(X \in C^c) = 0$,

$$\begin{aligned} \limsup \mathbb{P}(g(X_n) \in F) &\leq \limsup \mathbb{P}(X_n \in \overline{g^{-1}(F)}) \leq \mathbb{P}(X \in \overline{g^{-1}(F)}) \\ &\leq \mathbb{P}(X \in g^{-1}(F) \cup C^c) = \mathbb{P}(X \in g^{-1}(F)) = \mathbb{P}(g(X) \in F), \end{aligned}$$

et, donc, en appliquant de nouveau le lemme de Portemanteau, $g(X_n) \xrightarrow{\mathcal{L}} g(X)$.

(ii) : Soit $\epsilon > 0$. Pour tout $\delta > 0$, soit B_δ l'ensemble des points x tels qu'il existe y tel que $\|x - y\| \leq \delta$ mais $\|g(x) - g(y)\| \geq \epsilon$. Si $X \notin B_\delta$ et $\|g(X_n) - g(X)\| \geq \epsilon$, alors $\|X_n - X\| \geq \delta$. Nous avons donc :

$$\mathbb{P}(\|g(X_n) - g(X)\| \geq \epsilon) \leq \mathbb{P}(X \in B_\delta) + \mathbb{P}(\|X_n - X\| \geq \delta).$$

Le second terme du membre de droite tend vers 0 car $X_n \xrightarrow{\mathbb{P}} X$. Nous avons $\mathbb{P}(X \in B_\delta \cap C^c) = 0$ et $\lim_{\delta \searrow 0} \mathbb{P}(X \in B_\delta \cap C) = 0$ par continuité de g .

(iii) : dans ce cas, une simple application de la définition donne le résultat. \square

B.5 Loi des grands nombres

Proposition B.5.1 (Loi faible des grands nombres). *Soit $\{X_n\}_{n>0}$ une suite de v.a. i.i.d. telle que $E\|X_n\| < \infty$ et $E[X_n] = \mu$. Alors*

$$\bar{X}_n = n^{-1} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} \mu.$$

Proposition B.5.1. Notons $\psi_n(t)$ la fonction caractéristique de \bar{X}_n et $\phi(t)$ la fonction caractéristique de Z_k . Les variables étant i.i.d., :

$$\psi_n(t) = \mathbb{E} \left[\exp(itn^{-1} \sum_{k=1}^n Z_k) \right] = \prod_{k=1}^n \mathbb{E} [\exp(itn^{-1} Z_k)] = \phi(n^{-1}t)^n.$$

Comme $\mathbb{E}\|Z\| < \infty$ existe, ϕ est dérivable en 0 et

$$\phi(n^{-1}t)^n = \left(1 + \frac{it}{n} \mathbb{E}[Z] + o(n^{-1}) \right)^n \rightarrow e^{it\mu}.$$

Le membre de droite est la fonction caractéristique de la constante μ . Le théorème de Levy B.3.3 montre que $\bar{Z}_n \xrightarrow{\mathcal{L}} \mu$ et donc $\bar{X}_n \xrightarrow{\mathbb{P}} \mu$. \square

Théorème B.5.2 (Loi forte des grands nombres). *Soit $\{X_n\}_{n>0}$ une suite de v.a. i.i.d. telle que $E\|X_n\| < \infty$ et $E[X_n] = \mu$. Alors*

$$\bar{Z}_n = n^{-1} \sum_{i=1}^n Z_i \xrightarrow{\mathbb{P}\text{-p.s.}} \mu.$$

Comme son nom l'indique, la loi forte des grands nombres est un résultat plus puissant que la loi faible, puisque la convergence p.s. implique la convergence en probabilité, comme nous le verrons au théorème B.7.

B.6 théorème de la limite centrale

Le théorème de limite centrale (T.L.C.) donne des conditions sous lesquelles des sommes normalisées de v.a. indépendantes de moyenne nulle converge en loi vers une gaussienne. Ce résultat joue un rôle majeur en statistique (voir Le Cam, 1986, pour une histoire de ce théorème).

Théorème B.6.1. Soit $\{X_n\}$ une suite de vecteurs aléatoires de \mathbb{R}^d i.i.d. de moyenne μ et de matrice de covariance $\Sigma > 0$. Alors :

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \xrightarrow{\mathcal{L}} \mathbf{N}(0, \Sigma),$$

où l'on a noté $\mathbf{N}(0, \Sigma)$ le vecteur gaussien centré de matrice de covariance Σ .

Démonstration. Il suffit d'établir le résultat pour des v.a. scalaires. Le procédé de Cramér-Wold (voir section B.3) nous permet d'étendre le résultat au cas vectoriel.

Posons : $Y_n = \sigma^{-1}(X_n - \mu)$. La variable aléatoire Y_n est de moyenne nulle et de variance unité : $\mathbb{E}[Y_n] = 0$ et $\text{Var}(Y_n) = 1$. Notons :

$$\bar{Y}_n = n^{-1} \sum_{i=1}^n Y_i.$$

Pour $t \in \mathbb{R}$, notons $\phi_n(t)$ la fonction caractéristique associée à la variable aléatoire $\sqrt{n}\bar{Y}_n$.

$$\phi_n(t) = \mathbb{E} [\exp(it\sqrt{n}\bar{Y}_n)] .$$

Nous allons démontrer que pour tout t :

$$\lim_{n \rightarrow \infty} \phi_n(t) = \exp(-t^2/2)$$

et nous concluons en utilisant le théorème de Lévy, en remarquant que $\exp(-t^2/2)$ est la fonction caractéristique de la loi gaussienne de moyenne nulle et de variance unité.

Les variables aléatoires $\{Y_i\}_{1 \leq i \leq n}$ étant indépendantes, nous avons, pour tout $t \in \mathbb{R}$:

$$\phi_n(t) = \psi(n^{-1/2}t)^n ,$$

où $\psi(t) = \mathbb{E} [\exp(itY)]$. Comme $\mathbb{E}[Y] = 0$ et $\mathbb{E}[Y^2] = 1$, on a, quand $n \rightarrow \infty$,

$$\psi(n^{-1/2}t) = 1 - \frac{1}{2}(n^{-1/2}t)^2 + o(n^{-1}) ,$$

et par suite,

$$\lim_{n \rightarrow \infty} \phi_n(t) = \exp(-(1/2)t^2),$$

ce qui conclut la preuve du T.L.C. scalaire. □

Il existe une autre méthode de preuve du théorème de la limite centrale (T.L.C.), due à Lindeberg (1922), qui permet de généraliser le T.L.C. à des variables aléatoires indépendantes mais qui ne sont pas nécessairement identiquement distribuées. Ce résultat s'applique donc en toute généralité à des tableaux triangulaires de v.a. indépendantes.

Théorème B.6.2 (Lindeberg-Feller). Soit $(k_n, n \geq 0)$ une suite d'entiers croissante. Soit $(Y_{n,1}, \dots, Y_{n,k_n})$ un tableau triangulaire de vecteurs aléatoires indépendants centrés tels que

$\mathbb{E} [\|Y_{n,i}\|^2] < \infty$ pour $i \in \{1, \dots, k_n\}$. Supposons les conditions de Lindeberg–Feller vérifiées :

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{k_n} \mathbb{E} [\|Y_{n,i}\|^2 \mathbb{1}(\|Y_{n,i}\| > \epsilon)] = 0, \quad \forall \epsilon > 0, \quad (\text{B.2})$$

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{k_n} \text{Var}(Y_{n,i}) = \Sigma. \quad (\text{B.3})$$

Alors, la suite $\sum_{i=1}^{k_n} Y_{n,i}$ converge vers une loi normale $\mathbf{N}(0, \Sigma)$.

Démonstration. Par le procédé de Cramér–Wold, il suffit de montrer ce résultat en dimension un. Ce résultat repose tout d’abord sur le fait que (B.3) permettrait de conclure directement si on supposait les v.a. $Y_{n,i}$ gaussiennes (voir proposition B.9.3). La méthode de Lindeberg repose alors sur la comparaison entre les sommes partielles $\sum_{i=1}^{k_n} Y_{n,i}$ et $\sum_{i=1}^{k_n} X_{n,i}$ où $(X_{n,i})_{n \in \mathbb{N}, 1 \leq k \leq k_n}$ est un tableau triangulaires de v.a. *gaussiennes* indépendantes, telles que,

— Pour tout n et pour tout $i \in \{1, \dots, k_n\}$, $\text{var} Y_{n,i} = \text{var} X_{n,i}$,

— Pour tout n et tout $i, j \in \{1, \dots, k_n\}$, les v.a. $X_{n,i}$ et $Y_{n,j}$ sont indépendantes.

Nous allons tout d’abord que, sous ces deux conditions, il est possible de contrôler la différence entre les sommes partielles construites à partir du tableau triangulaire $(X_{n,i}, i \in \{1, \dots, k_n\})$ et $(Y_{n,i}, i \in \{1, \dots, k_n\})$ de telle sorte que la convergence en loi de la somme partielle $S_n = \sum_{i=1}^{k_n} X_{n,i}$ implique la convergence en loi de la somme partielle $T_n = \sum_{i=1}^{k_n} Y_{n,i}$.

Considérons deux sommes $S_n = X_{n,1} + X_{n,2} + \dots + X_{n,k_n}$ et $T_n = Y_{n,1} + Y_{n,2} + \dots + Y_{n,k_n}$. Soit f une fonction deux fois différentiable avec une dérivée seconde bornée et Lipschitzienne, *i.e.*,

$$|f''|_{\text{Lip}} = \sup_{(x,y) \in \mathbb{R} \times \mathbb{R}, x \neq y} \frac{|f''(x) - f''(y)|}{|x - y|} < \infty. \quad (\text{B.4})$$

où f'' est la dérivée seconde de f . On a la décomposition

$$\mathbb{E} [f(S_n)] - \mathbb{E} [f(T_n)] = \sum_{k=1}^{k_n} [\mathbb{E} [f(R_{n,k} + X_{n,k})] - \mathbb{E} [f(R_{n,k} + Y_{n,k})]],$$

où $R_{n,k} = \left(\sum_{j < k} X_{n,j} \right) + \left(\sum_{j > k} Y_{n,j} \right)$. Développons $f(R_{n,k} + X_{n,k})$ au voisinage de $R_{n,k}$:

$$f(R_{n,k} + X_{n,k}) = f(R_{n,k}) + X_{n,k} f'(R_{n,k}) + \frac{X_{n,k}^2}{2} f''(R_{n,k}) + \frac{X_{n,k}^2}{2} [f''(R_{n,k} + \theta_{n,k} X_{n,k}) - f''(R_{n,k})],$$

où $\theta_{n,k} \in [0, 1]$. Développons de même $f(R_{n,k} + Y_{n,k})$. Notons que, la v.a. $R_{n,k}$ est indépendante, par construction, des v.a. $X_{n,k}$ et $Y_{n,k}$. Par conséquent, nous avons pour tout $k \in \{1, \dots, n\}$,

$$\mathbb{E} [f'(R_{n,k})(X_{n,k} - Y_{n,k})] = \mathbb{E} [f'(R_{n,k})] (\mathbb{E} [X_{n,k}] - \mathbb{E} [Y_{n,k}]) = 0,$$

en utilisant que $\mathbb{E} [X_{n,k}] = \mathbb{E} [Y_{n,k}] = 0$. De la même façon, comme par construction des v.a. $X_{n,k}$ et $Y_{n,k}$ nous avons $\mathbb{E} [X_{n,k}^2] = \mathbb{E} [Y_{n,k}^2]$, nous avons

$$\mathbb{E} [f''(R_{n,k})(X_{n,k}^2 - Y_{n,k}^2)] = \mathbb{E} [f''(R_{n,k})] (\mathbb{E} [X_{n,k}^2] - \mathbb{E} [Y_{n,k}^2]) = 0,$$

Comme f'' est une fonction Lipschitzienne, pour tout $\epsilon > 0$, nous avons, pour tout $\theta \in [0, 1]$,

$$\begin{aligned} |f''(R_{n,k} + \theta X_{n,k}) - f''(R_{n,k})| &\leq |f''|_{\text{Lip}} |X_{n,k}|, \\ |f''(R_{n,k} + \theta Y_{n,k}) - f''(R_{n,k})| &\leq |f''|_{\text{Lip}} |Y_{n,k}| \mathbb{1}(|Y_{n,k}| \leq \epsilon) + 2|f''|_{\infty} \mathbb{1}(|Y_{n,k}| > \epsilon), \end{aligned}$$

où $|f''|_{\infty} = \sup_{x \in \mathbb{R}} |f''(x)|$ (qui est fini par hypothèse). Nous utilisons ici deux majorations différentes pour des raisons qui deviendront transparentes dans la suite de la preuve. Par conséquent, $\mathbb{E}[f(R_{n,k} + X_{n,k})] - \mathbb{E}[f(R_{n,k} + Y_{n,k})]$ est majoré en valeur absolue par

$$\frac{1}{2} |f''|_{\text{Lip}} (\mathbb{E}[|X_{n,k}|^3] + \mathbb{E}[|Y_{n,k}| \mathbb{1}(|Y_{n,k}| \leq \epsilon)|^3]) + |f''|_{\infty} \mathbb{E}[|Y_{n,k}| \mathbb{1}(|Y_{n,k}| > \epsilon)|^2],$$

pour tout $\epsilon > 0$. Remarquons que

$$\mathbb{E}[|X_{n,k}|^3] = \sigma_{n,k}^3 \mathbb{E}[|X_{n,k}/\sigma_{n,k}|^3] = \sigma_{n,k}^3 m_3,$$

où m_3 est le moment d'ordre trois d'une loi gaussienne centrée réduite et $\sigma_{n,k} \stackrel{\text{def}}{=} \mathbb{E}[X_{n,k}^2] = \mathbb{E}[Y_{n,k}^2]$. Remarquons aussi que

$$\mathbb{E}[|Y_{n,k}| \mathbb{1}(|Y_{n,k}| \leq \epsilon)|^3] \leq \epsilon \mathbb{E}[|Y_{n,k}|^2] = \epsilon \sigma_{n,k}^2.$$

Ces inégalités conduisent à la majoration :

$$|\mathbb{E}[f(S_n)] - \mathbb{E}[f(T_n)]| \leq \frac{1}{2} |f''|_{\text{Lip}} \left(m_3 \sum_{k=1}^{k_n} \sigma_{n,k}^3 + \epsilon \sum_{k=1}^{k_n} \sigma_{n,k}^2 \right) + |f''|_{\infty} \sum_{k=1}^{k_n} \mathbb{E}[|Y_{n,k}| \mathbb{1}(|Y_{n,k}| > \epsilon)|^2].$$

On a d'autre part

$$\sum_{k=1}^{k_n} \sigma_{n,k}^3 \leq \sum_{k=1}^{k_n} \sigma_{n,k}^2 \max_{k \in \{1, \dots, k_n\}} (\sigma_{n,k}).$$

Or (B.3) implique que pour tout $1 \leq k \leq k_n$, et tout $\epsilon > 0$, nous avons

$$\sigma_{n,k}^2 \leq \epsilon^2 + \mathbb{E}[|Y_{n,k}| \mathbb{1}(|Y_{n,k}| > \epsilon)|^2] \leq \epsilon^2 + \sum_{j=1}^{k_n} \mathbb{E}[|Y_{n,j}| \mathbb{1}(|Y_{n,j}| > \epsilon)|^2].$$

et donc

$$\limsup_{n \rightarrow \infty} \max_{k \in \{1, \dots, k_n\}} \sigma_{n,k}^2 = 0.$$

D'où, pour tout $\epsilon > 0$,

$$\begin{aligned} |\mathbb{E}[f(S_n)] - \mathbb{E}[f(T_n)]| &\leq \frac{1}{2} |f''|_{\text{Lip}} \left(m_3 \sum_{k=1}^{k_n} \sigma_{n,k}^2 \left(\epsilon^2 + \sum_{j=1}^{k_n} \mathbb{E}[|Y_{n,j}| \mathbb{1}(|Y_{n,j}| > \epsilon)|^2] \right)^{1/2} + \epsilon \sum_{k=1}^{k_n} \sigma_{n,k}^2 \right) \\ &\quad + |f''|_{\infty} \sum_{k=1}^{k_n} \mathbb{E}[|Y_{n,k}| \mathbb{1}(|Y_{n,k}| > \epsilon)|^2]. \quad (\text{B.5}) \end{aligned}$$

Par hypothèse la série $\sum_k \mathbb{E} [|Y_{n,k}| \mathbb{1}(|Y_{n,k}| > \epsilon)]^2$ tend vers zéro et $\sum_k \sigma_{n,k}^2$ tend vers une constante σ^2 respectivement quand n tend vers l'infini. D'où, pour tout $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} |\mathbb{E} [f(S_n)] - \mathbb{E} [f(T_n)]| \leq \frac{1}{2} |f''|_{\text{Lip}} (m_3 + 1) \sigma^2 \epsilon.$$

En faisant tendre, dans un deuxième temps, ϵ vers zéro, on obtient donc une limite nulle. Comme les $(X_{n,j}, j \in \{1, \dots, k_n\})$ sont des v.a. gaussiennes, la proposition B.9.3 permet de conclure. Comme ces majorations sont en particulier valables pour $f(x) = \exp(-itx)$, le théorème B.3.3 de Levy permet de conclure. La majoration (B.5) est plus précise que le résultat de convergence en loi car elles fournissent des vitesses de convergences de $\mathbb{E} [f(S_n)]$ en fonction de normes appliquées au tableau $Y_{n,k}$ et de normes appliquées à f pour des f convenablement choisis. \square

B.7 Symboles o et O stochastiques

Le théorème suivant clarifie les relations entre les différentes définitions de convergence et celles entre convergence d'un vecteur et convergence de ses coordonnées.

Théorème B.7.1. *Soient X_n, X et Y_n des v.a. Nous avons :*

- (i) $X_n \xrightarrow{\mathbb{P}\text{-p.s.}} X$ implique $X_n \xrightarrow{\mathbb{P}} X$,
- (ii) $X_n \xrightarrow{\mathbb{P}} X$ implique $X_n \xrightarrow{\mathcal{L}} X$,
- (iii) $X_n \xrightarrow{\mathbb{P}} c$, où c est une constante, si et seulement si $X_n \xrightarrow{\mathcal{L}} c$,
- (iv) Si $X_n \xrightarrow{\mathcal{L}} X$ et $X_n - Y_n \xrightarrow{\mathbb{P}} 0$, alors $Y_n \xrightarrow{\mathcal{L}} X$,
- (v) Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathbb{P}} c$ pour une constante c , alors, $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, c)$,
- (vi) Si $X_n \xrightarrow{\mathbb{P}} X$ et $Y_n \xrightarrow{\mathbb{P}} Y$ alors $(X_n, Y_n) \xrightarrow{\mathbb{P}} (X, Y)$.

Démonstration. (i) : Si $X_n \xrightarrow{\mathbb{P}\text{-p.s.}} X$, pour tout $\epsilon > 0$, $\mathbb{P}(\bigcap_n A_n) = 0$ avec $A_n = \bigcup_{k \geq n} \{\|X_k - X\| \geq \epsilon\}$. Par conséquent :

$$\lim_n \mathbb{P}(\|X_n - X\| \geq \epsilon) \leq \lim_n \mathbb{P}(A_n) = 0$$

(ii) : Soit f Lipschitzienne bornée, notons $|f|_{\infty} = \sup_x |f(x)|$ et $|f|_{\text{Lip}} = \sup_{x \neq y} |f(x) - f(y)| / \|x - y\|$. Pour tout $\epsilon > 0$,

$$|\mathbb{E} [f(X_n)] - \mathbb{E} [f(X)]| \leq \epsilon |f|_{\text{Lip}} + 2|f|_{\infty} \mathbb{P}(\|X_n - X\| \geq \epsilon).$$

Le second terme du membre de droite tend vers 0 et le premier peut être rendu arbitrairement petit. Donc, $\lim_n \mathbb{E} [f(X_n)] = \mathbb{E} [f(X)]$.

(iii) : (ii) montre que si $X_n \xrightarrow{\mathbb{P}} c$, alors $X_n \xrightarrow{\mathcal{L}} c$. Pour la réciproque, soit $\epsilon > 0$ et soit $B(c, \epsilon) = \{x : \|x - c\| < \epsilon\}$ la boule ouverte de centre c et de rayon ϵ . Nous avons $\mathbb{P}(\|X_n - c\| \geq \epsilon) = \mathbb{P}(X_n \in B(c, \epsilon)^c)$.

Si $X_n \xrightarrow{\mathcal{L}} c$, le lemme de Portmanteau montre que $\limsup \mathbb{P}(X_n \in B(c, \epsilon)^c) \leq \mathbb{P}(c \in B(c, \epsilon)^c) = 0$

(iv) : Soit f Lipschitzienne bornée, et $\epsilon > 0$,

$$|\mathbb{E} [f(X_n)] - \mathbb{E} [f(Y_n)]| \leq |f|_{\text{Lip}} \epsilon + |f|_{\infty} \mathbb{P}(\|X_n - Y_n\| \geq \epsilon).$$

Le second terme tend vers 0 et le premier peut être rendu arbitrairement petit, donc $\mathbb{E}[f(X_n)]$ et $\mathbb{E}[f(Y_n)]$ ont la même limite.

(v) : Remarquons, en notant $d(x, y)$ la distance euclidienne, que $d((X_n, Y_n), (X_n, c)) = d(Y_n, c) \rightarrow 0$. Donc, en utilisant (iv), il suffit de prouver que $(X_n, c) \xrightarrow{\mathcal{L}} (X, c)$. Pour toute fonction continue bornée $f : (x, y) \rightarrow f(x, y)$ la fonction $f(\cdot, c) : x \rightarrow f(x, c)$ est continue et bornée et $|\mathbb{E}[f(X_n, c)] - \mathbb{E}[f(X, c)]| \rightarrow 0$, car $X_n \xrightarrow{\mathcal{L}} X$.

(vi) : conséquence directe de $d((x_1, y_1), (x_2, y_2)) \leq d(x_1, y_1) + d(x_2, y_2)$. □

Le théorème B.7 a quelques applications immédiates et utiles en pratique.

La propriété (vi) montre que la convergence en probabilité d'une suite de vecteurs aléatoires $X_n = (X_{n,1}, \dots, X_{n,k})$ est équivalente à la convergence de chacune de ses composantes individuellement. Le résultat analogue pour la convergence en distribution est faux : la convergence en distribution d'une suite de vecteurs aléatoires est une propriété plus forte que la convergence en distribution de chacune de ses composantes $X_{n,i}$.

La propriété (v) du théorème implique que si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathbb{P}} c$, alors $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, c)$. Le théorème de continuité (théorème B.4.1) montre donc que pour toute fonction $g : (x, y) \rightarrow g(x, y)$ continue sur un ensemble $C \times \{c\}$, $\mathbb{P}(X \in C) = 1$, $g(X_n, Y_n) \xrightarrow{\mathcal{L}} g(X, c)$. Des applications particulières de ce principe sont souvent regroupées sous la forme du lemme suivant, connue sous le nom de lemme de Slutsky.

Lemme B.7.2 (Lemme de Slutsky). *Soit X_n, X et Y_n des v.a. réelles. Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} c$ où c est une constante, alors*

$$(i) \quad X_n + Y_n \xrightarrow{\mathcal{L}} X + c;$$

$$(ii) \quad Y_n X_n \xrightarrow{\mathcal{L}} cX;$$

$$(iii) \quad Y_n^{-1} X_n \xrightarrow{\mathcal{L}} c^{-1}X \text{ pour } c \neq 0.$$

Le procédé de Cramér-Wold (voir section B.3) permet d'étendre ce résultat au cas vectoriel/matriciel, pour peu que, dans (i), c soit un vecteur de même dimension que X , et dans (ii) et (iii), $\{Y_n\}_n$ et c soient des matrices (avec c inversible pour (iii)) de même dimension adaptée à celle des vecteurs $\{X_n\}$.

Il est pratique de disposer de notations simples pour exprimer qu'une suite tend vers 0 en probabilité ou est bornée en probabilité. Nous dirons qu'une suite de v.a. (scalaire ou vecteur) $X_n = o_P(1)$ si $X_n \xrightarrow{\mathbb{P}} 0$; nous noterons $X_n = O_P(1)$ si X_n est bornée en probabilité. Plus généralement, pour R_n une suite de v.a.,

$$X_n = o_P(R_n) \quad \text{signifie} \quad X_n = Y_n R_n \quad \text{avec } Y_n = o_P(1), \tag{B.6}$$

$$X_n = O_P(R_n) \quad \text{signifie} \quad X_n = Y_n R_n \quad \text{avec } Y_n = O_P(1), \tag{B.7}$$

$$\tag{B.8}$$

Pour X_n et R_n des suites déterministes, les symboles o_P et O_P coïncident avec les symboles o et O de l'analyse. Les symboles o_P et O_P se manipulent d'ailleurs exactement de la même façon que les

symboles o et O . Par exemple,

$$\begin{aligned} o_P(1) + o_P(1) &= o_P(1), \\ o_P(1) + O_P(1) &= O_P(1), \\ O_P(1)o_P(1) &= o_P(1), \\ (1 + o_P(1))^{-1} &= O_P(1), \\ o_P(R_n) &= R_n o_P(1), \quad O_P(R_n) = R_n O_P(1) \end{aligned}$$

Pour s'assurer de la validité de ces règles, il suffit de les ré-écrire explicitement avec des suites et d'utiliser les résultats classiques énoncés ci-dessus. Par exemple, si $X_n \xrightarrow{\mathbb{P}} 0$ et $Y_n \xrightarrow{\mathbb{P}} 0$, le théorème B.7 (vi) implique que $(X_n, Y_n) \xrightarrow{\mathbb{P}} (0, 0)$ qui équivaut à $(X_n, Y_n) \xrightarrow{\mathcal{L}} (0, 0)$. Le théorème de continuité (appliqué à $f : (x, y) \mapsto x + y$) implique $X_n + Y_n \xrightarrow{\mathcal{L}} 0$, qui équivaut à $X_n + Y_n \xrightarrow{\mathbb{P}} 0$ (théorème B.7 (iii)). La troisième règle est une façon concise d'écrire : si X_n est borné en probabilité et $Y_n \xrightarrow{\mathbb{P}} 0$, alors $X_n Y_n \xrightarrow{\mathbb{P}} 0$. Si $X_n \xrightarrow{\mathcal{L}} X$, alors ce résultat découle du lemme de Slutsky (car $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathbb{P}} c$ implique que $Y_n X_n \xrightarrow{\mathcal{L}} cX$, donc si $c = 0$, $Y_n X_n \xrightarrow{\mathcal{L}} 0$ qui équivaut à $Y_n X_n \xrightarrow{\mathbb{P}} 0$). Dans le cas où X_n ne converge pas en probabilité, on peut soit donner une preuve directe, soit utiliser le théorème de Prohorov.

La règle de calcul suivante est utile pour les développements asymptotiques.

Lemme B.7.3. *Soit R une fonction $D \subset \mathbb{R}^k$ telle que $R(0) = 0$. Soit X_n une suite de v.a. à valeurs dans D tels que $X_n \rightarrow 0$. Alors, pour tout $p > 0$,*

- (i) *Si $R(h) = o(\|h\|^p)$ quand $h \rightarrow 0$, alors $R(X_n) = o_P(\|X_n\|^p)$,*
- (ii) *Si $R(h) = O(\|h\|^p)$ quand $h \rightarrow 0$, alors $R(X_n) = O_P(\|X_n\|^p)$.*

Démonstration. Définissons $g(h) = R(h)/\|h\|^p$ pour $h \neq 0$ et $g(0) = 0$: $R(X_n) = g(X_n)\|X_n\|^p$.

(i) La fonction g est continue en 0 et le théorème de continuité montre que $g(X_n) \xrightarrow{\mathbb{P}} g(0) = 0$. La deuxième assertion se démontre de façon similaire. \square

B.8 Delta-méthode

Supposons que nous disposions d'un estimateur T_n d'un paramètre ν , mais que la quantité d'intérêt soit $g(\nu)$. Un estimateur naturel de cette quantité est $g(T_n)$. Pouvons nous déduire le comportement asymptotique de $g(T_n)$ à partir de celui de T_n ?

Nous savons déjà que si T_n converge en probabilité vers ν et que g est continue au point ν , alors $g(T_n)$ converge en probabilité vers $g(\nu)$.

Si nous savons $\sqrt{n}(T_n - \nu)$ converge en loi vers une distribution limite, pouvons nous affirmer qu'il en est de même pour $\sqrt{n}(g(T_n) - g(\nu))$? La réponse est affirmative si la fonction g est différentiable au point ν : de façon heuristique, nous avons :

$$\sqrt{n}(g(T_n) - g(\nu)) \simeq g'(\nu)\sqrt{n}(T_n - \nu),$$

et donc, si $\sqrt{n}(T_n - \nu) \rightarrow_d T$, alors $\sqrt{n}(g(T_n) - g(\nu)) \rightarrow_d [g'(\nu)]T$. En particulier, si $\sqrt{n}(T_n - \nu)$ est asymptotiquement $\mathbf{N}(0, \sigma^2)$ alors $\sqrt{n}(g(T_n) - g(\nu))$ est asymptotiquement normal $\mathbf{N}(0, [g'(\nu)]^2 \sigma^2)$.

La même question se pose lorsque $\mathbf{T}_n = (T_{n,1}, \dots, T_{n,k})$ est un vecteur aléatoire et g est une fonction de $\mathbb{R}^k \mapsto \mathbb{R}^m$. Nous connaissons la loi de $\sqrt{n}(\mathbf{T}_n - \nu)$ et nous nous intéressons à la loi de $\sqrt{n}(g(\mathbf{T}_n) - g(\nu))$. Le résultat ci-dessus s'étend directement en remplaçant la dérivée par la différentielle $\nabla_\nu g(\nu)$.

Proposition B.8.1. *Soit $g : D_g \subset \mathbb{R}^k \mapsto \mathbb{R}^m$ une fonction définie sur un sous ensemble D_g de \mathbb{R}^k et différentiable au point ν . Soit \mathbf{T}_n des v.a. à valeurs dans D_g . Si $r_n(\mathbf{T}_n - \nu) \xrightarrow{\mathcal{L}} T$, pour une suite $r_n \nearrow \infty$, alors $r_n(g(\mathbf{T}_n) - g(\nu)) \xrightarrow{\mathcal{L}} \nabla_\nu g(\nu) \mathbf{T}$. De plus $r_n(g(\mathbf{T}_n) - g(\nu)) - \nabla_\nu g(\nu)(r_n(\mathbf{T}_n - \nu)) \xrightarrow{\mathbb{P}} 0$.*

Démonstration. $r_n(\mathbf{T}_n - \nu)$ converge en distribution, la suite $r_n(\mathbf{T}_n - \nu)$ est bornée en probabilité et $\mathbf{T}_n - \nu$ tend vers 0. La différentiabilité de la fonction g implique que $g(\nu + h) = g(\nu) + \nabla_\nu g(\nu)h + R(h)$ et $R(h) = o(\|h\|)$. Le lemme B.7.3 montre que :

$$g(T_n) - g(\nu) - \nabla_\nu g(\nu)(T_n - \nu) = R(T_n - \nu) = o_P(\|T_n - \nu\|).$$

En multipliant les deux membres de l'identité précédente par r_n , nous avons donc :

$$r_n(g(T_n) - g(\nu)) = \nabla_\nu g(\nu)(r_n(T_n - \nu)) + o_P(1).$$

Nous concluons en appliquant le lemme de Slutsky. □

B.9 Convergence des moments

Par définition $X_n \xrightarrow{\mathcal{L}} X$ implique que pour toute fonction continue bornée f , $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$. La condition bornée n'est pas superflue, et il est très facile de trouver des exemples de suite de variables vérifiant $X_n \xrightarrow{\mathcal{L}} X$ et pour lesquelles nous n'avons pas $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ pour f une fonction continue non bornée.

Définition B.9.1 (Uniforme Intégrabilité). *Une suite de v.a. Y_n est dite uniformément intégrable si*

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E}[\|Y_n\| \mathbb{1}(\|Y_n\| \geq M)] = 0.$$

Remarquons tout d'abord que l'uniforme intégrabilité implique $\mathbb{E}[\|Y_n\|]$ est majoré indépendamment de n . En effet, il existe M et $C > 0$ tels que, pour tout n

$$\mathbb{E}[\|Y_n\| \mathbb{1}(\|Y_n\| \geq M)] \leq C,$$

ce qui implique

$$\mathbb{E}[\|Y_n\|] = \mathbb{E}[\|Y_n\| \mathbb{1}(\|Y_n\| \leq M)] + \mathbb{E}[\|Y_n\| \mathbb{1}(\|Y_n\| \geq M)] \leq M + C$$

pour tout n . Comme le montre le théorème ci-dessous, l'uniforme intégrabilité permet de relier la convergence en loi et la convergence des moments.

Théorème B.9.2. *Soit $f : \mathbb{R}^k \mapsto \mathbb{R}$ une fonction borélienne continue en tout point de $C \in \mathcal{B}(\mathbb{R}^k)$. Supposons que $X_n \xrightarrow{\mathcal{L}} X$ et $\mathbb{P}(X \in C) = 1$. Alors, $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ si et seulement si la suite $f(X_n)$ est uniformément intégrable.*

Démonstration. Nous ne montrons ici que la réciproque. Posons $Y_n = f(X_n)$ et supposons que Y_n est uniformément intégrable. Nous allons montrer que $\mathbb{E}[Y_n] \rightarrow \mathbb{E}[Y]$, où $Y = f(X)$. Nous supposons sans perte de généralité que Y_n est positive (il suffit autrement de raisonner sur les parties positives et négatives séparément). Le théorème de continuité montre que $Y_n \xrightarrow{\mathcal{L}} Y$. Nous notons $a \wedge b = \inf(a, b)$. L'inégalité triangulaire donne d'une part

$$\mathbb{E}[Y \wedge M] \leq |\mathbb{E}[Y_n \wedge M] - \mathbb{E}[Y \wedge M]| + \mathbb{E}[Y_n \wedge M]$$

et d'autre part

$$|\mathbb{E}[Y_n] - \mathbb{E}[Y]| \leq |\mathbb{E}[Y_n] - \mathbb{E}[Y_n \wedge M]| + |\mathbb{E}[Y_n \wedge M] - \mathbb{E}[Y \wedge M]| + |\mathbb{E}[Y \wedge M] - \mathbb{E}[Y]|.$$

Comme la fonction $y \rightarrow y \wedge M$ est continue et bornée, $|\mathbb{E}[Y_n \wedge M] - \mathbb{E}[Y \wedge M]| \rightarrow 0$ quand $n \rightarrow \infty$. Le second terme de la partie droite de la première inégalité est majorée indépendamment de M (voir ci-dessus), donc $\mathbb{E}[Y] < \infty$. Le premier terme et le troisième termes de la partie droite de la seconde inégalité peuvent être rendus arbitrairement petits en utilisant respectivement l'uniforme intégrabilité et $\mathbb{E}[Y] < \infty$, ce qui achève la démonstration de l'implication réciproque. \square

Une conséquence élémentaire de l'exemple précédent est donnée ci-après.

Proposition B.9.3. *Si $\{X_k, k \in \mathbb{N}\}$ est une famille de vecteurs aléatoires gaussiens, alors les deux assertions suivantes sont équivalentes.*

- (i) $\lim_{k \rightarrow \infty} \mathbb{E}[X_k] = \mu$ et $\lim_{k \rightarrow \infty} \text{Var}(X_k) = \Sigma$
- (ii) $X_k \xrightarrow{\mathcal{L}} \mathbf{N}(\mu, \Sigma)$.

Démonstration. Par le procédé de Cramér-Wold, il suffit de montrer ce résultat en dimension un. En dimension un, on note $\Sigma = \sigma^2$. (i) \Rightarrow (ii) est alors une simple application du lemme B.7.2 de Slutsky puisque $(X_k - \mathbb{E}[X_k])/\sqrt{\text{var} X_k}$ suit une loi $\mathbf{N}(0, 1)$ pour tout k . D'autre part, le théorème de convergence dominé appliqué à

$$\mathbb{E}[f(X_k)] = \frac{1}{\sqrt{2\pi\text{var}(X_k)}} \int f(t) \exp(-(t - \mathbb{E}[X_k])^2/(2\text{var}(X_k))) dt$$

pour tout f continue bornée montre que si $\text{var}(X_k) \rightarrow \infty$, $\mathbb{E}[f(X_k)] \rightarrow 0$, ce qui est impossible sous (ii). Il en est de même pour toute sous-suite $\text{var}(X_{\alpha_k})$. On en conclut que (ii) implique que $\text{var}(X_k)$ est majoré indépendamment de k . Dans ce cas il est facile de voir que $\mathbb{E}[X_k]$ a aussi une limite car le contraire impliquerait que $\mathbb{P}\{|X| \leq M\} = 0$ pour tout $M > 0$. Par un simple changement de variable, on montre que pour tout $p > 0$, il existe une constante c_p telle que, pour toute variable gaussienne centrée Y , $\mathbb{E}[|Y|^p] = c_p(\mathbb{E}[Y^2])^{\frac{p}{2}}$. Il s'en suit que

$$(\mathbb{E}[|X_k|^p])^{\frac{1}{p}} \leq (\mathbb{E}[|X_k - \mathbb{E}[X_k]|^p])^{\frac{1}{p}} + |\mathbb{E}[X_k]| \leq c_p \text{var}^{1/2}(X_k) + |\mathbb{E}[X_k]|$$

est majoré indépendamment de k pour tout $p > 0$. On obtient par uniforme intégrabilité que les moments $\mathbb{E}[X_k^q]$ converge vers $\mathbb{E}[X^q]$ pour tout $q > 0$. Les cas $q = 1$ et $q = 2$ donnent (i), ce qui achève la démonstration. \square

Bibliographie

Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002. ISBN 0-387-95313-2.

L. Schwartz. *Analyse. I*, volume 42 of *Collection Enseignement des Sciences [Collection : The Teaching of Science]*. Hermann, Paris, 1991. ISBN 2-7056-6161-8. Théorie des ensembles et topologie. [Set theory and topology], With the collaboration of K. Zizi.