

Université Pierre et Marie Curie

Master de Mathématiques

année 2004-05

Processus stationnaires et prévision

Laboratoire de Probabilités et Modèles Aléatoires

Processus stationnaires et prévision.

Mode d'emploi

Ce polycopié est destiné aux étudiants de l'U.E. "Processus stationnaires et prévision" du Master de Mathématiques de l'Université Pierre et Marie Curie. En principe il s'adresse donc à des étudiants ayant suivi un premier cours de probabilités. Cependant le chapitre 2 contient un rappel de tous les résultats probabilistes utilisés par la suite. Il est relativement autonome et peut éventuellement être abordé par un étudiant n'ayant jamais suivi de cours de probabilités. Le chapitre 1 introduit les principales notions de séries et transformations de Fourier, et comporte, en annexe, quelques éléments de la théorie classique du signal déterministe. Cette annexe qui, en principe, ne fait pas partie de ce cours, concerne néanmoins tous les étudiants désirant acquérir quelques notions de base de traitement du signal. Les chapitres 3, 5 et 6 sont consacrés aux deux sujets essentiels traités dans ce polycopié, à savoir l'étude spectrale des processus du second ordre et l'étude des séries chronologiques classiques: Ar, Ma, Arma ainsi que quelques éléments d'étude statistique de ces processus.

Un certain nombre de résultats sur la "petite" intégrale stochastique, figurant classiquement dans un cours sur les processus du second ordre, ont été rassemblés dans le chapitre 4 "L'intégrale stochastique", car ils ne sont pas vraiment nécessaires pour la compréhension des trois chapitres principaux, à savoir les chapitres 3, 5 et 6. Il est néanmoins vivement recommandé au lecteur d'en prendre connaissance.

Ce polycopié est divisé en chapitres, sections et sous-sections. Ainsi 3.2.4 renvoie au chapitre 3, section 2, sous-section 4 et 5.4 renvoie au chapitre 5, section 4. A l'intérieur d'une même section, les énoncés sont numérotés en continu. Ainsi "d'après le th. 5.4.6" renvoie au chapitre 5, section 4, énoncé 6. Quant aux égalités, elles sont numérotées entre parenthèses et en continu au sein d'un même chapitre. Ainsi "vu (3.5)" réfère à la cinquième égalité numérotée du chapitre 3. Le signe " \square " indique la fin d'une preuve.

Paolo Baldi qui a enseigné ce cours en 04-05, s'est tout particulièrement attaché à le compléter par de procédures d'estimation dans les modèles Arma.

Jean Lacroix

Table des Matières

1	Analyse de Fourier	7
1.1	Rappels de cours	7
1.1.1	Rappels sur les espaces de Hilbert	7
1.1.2	Rappels d'Intégration	9
1.2	Séries de Fourier	10
1.2.1	Coefficients de Fourier d'une fonction de $L^2(\mathbb{T}, dx)$	11
1.2.2	Transformation de Fourier d'une suite de $\ell^2(\mathbb{Z})$	14
1.2.3	Coefficients de Fourier d'une fonction de $L^1(\mathbb{T}, dx)$	14
1.3	Transformation de Fourier sur \mathbb{R}	16
1.3.1	Transformation de Fourier sur $L^1(\mathbb{R}, dx)$	17
1.3.2	Transformation de Fourier sur $L^2(\mathbb{R}, dx)$	19
1.3.3	Exemples	21
1.4	Convolution	23
1.5	Autocorrélation	24
1.6	Transformée en z	26
1.7	Annexe: Signaux déterministes	27
1.7.1	Echantillonnage et formule de Poisson	28
1.7.2	Transformée de Fourier sur un ensemble fini	31
1.7.3	Transformation de Laplace	33
2	Rappels de calcul des probabilités	35
2.1	Généralités	35
2.2	Convergences	36
2.3	Indépendance	38
2.4	V.a. complexes	40
2.5	Fonctions caractéristiques	41
2.6	Processus Gaussiens	42
2.7	Le Modèle linéaire	47
3	Théorie spectrale des processus du second ordre	49
3.1	Processus du second ordre	49
3.2	Théorèmes de Herglotz et Bochner	51
3.3	Filtrage des processus	57
3.4	Représentation spectrale des p.s.c.	59

4	L'intégrale stochastique	63
4.1	Processus à accroissements orthogonaux et intégrale stochastique	63
4.2	L'intégrale stochastique et les processus stationnaires	65
4.3	Les exemples fondamentaux	67
4.3.1	Le mouvement brownien.	67
4.3.2	Le processus de Poisson.	71
5	Processus stationnaires aléatoires	73
5.1	Inversion des filtres	73
5.2	Prédiction	74
5.3	Processus AR, MA et ARMA	74
5.3.1	Processus AR	74
5.3.2	Processus MA	77
5.3.3	Processus ARMA	79
5.3.4	La fonction de covariance des processus ARMA	82
5.4	Matrices de Toeplitz	83
5.4.1	L'Algorithme de Levinson	85
5.4.2	Prolongement AR d'un segment de covariance	87
5.4.3	Prolongement de Pisarenko d'un segment de covariance	88
6	Statistique des processus du second ordre	91
6.1	Théorèmes limites	92
6.2	Étude statistique d'un processus stationnaire	95
6.2.1	Estimation de la moyenne	95
6.2.2	Estimation de la covariance	95
6.3	Estimation des paramètres d'un processus ARMA réel	97
6.3.1	Estimation dans les modèles AR: les équations de Yule-Walker	97
6.3.2	Les estimateurs du maximum de vraisemblance	100
6.4	Estimation du spectre d'un p.s.c.	102
6.4.1	Estimation de la densité spectrale	102
6.4.2	Estimateurs des paramètres basés sur le spectre	104

Chapitre 1

Analyse de Fourier

1.1 Rappels de cours

1.1.1 Rappels sur les espaces de Hilbert

Soit H un espace de Hilbert complexe muni du produit scalaire $\langle \cdot, \cdot \rangle$. L'inégalité suivante, appelée inégalité de Schwarz, sera très souvent utilisée dans la suite du cours:

Proposition 1.1. *Soit x et y des éléments de H . On a alors:*

$$|\langle x, y \rangle| \leq \|x\| \|y\|$$

Avec égalité si et seulement si x et y sont colinéaires.

La propriété suivante porte le nom de "bicontinuité du produit scalaire":

Proposition 1.2. *Soit x_n et y_n deux suites dans H qui convergent respectivement vers x et y . Alors la suite numérique $\langle x_n, y_n \rangle$ converge vers $\langle x, y \rangle$.*

Définition 1.3. *Soit A une partie de H .*

- *On désigne par A^\perp l'ensemble des éléments de H orthogonaux à tous les éléments de A . C'est un sous espace fermé de H .*
- *On désigne par $\sigma(A)$ l'espace vectoriel fermé engendré par A , et on dit que la partie A est totale dans H si $\sigma(A) = H$.*

Proposition 1.4. *Soit A une partie de H . Alors:*

$$A^\perp = (\sigma(A))^\perp, \text{ et } H = \sigma(A) \oplus A^\perp$$

Cette proposition sera souvent utilisée sous la forme suivante:

Corollaire 1.5. *Soit A une partie de H .*

1. *A est totale dans H si et seulement si $A^\perp = \{0\}$.*
2. *Soit $x \in H$. Alors $x \in \sigma(A)$ si et seulement si pour tout $y \in A^\perp$ on a $\langle x, y \rangle = 0$.*

La proposition suivante est généralement attribuée à Bessel et Parseval:

Proposition 1.6. Soit $\{e_\lambda\}_{\lambda \in \Lambda}$ une famille orthonormée dans H . Pour $x \in H$ on pose $x_\lambda = \langle x, e_\lambda \rangle$. On a alors:

$$\sum_{\lambda \in \Lambda} |x_\lambda|^2 \leq \|x\|^2$$

Si de plus la famille $\{e_\lambda\}_{\lambda \in \Lambda}$ est totale dans H (on dit alors que c'est une base orthonormée) on a les égalités:

$$\sum_{\lambda \in \Lambda} \|x_\lambda\|^2 = \|x\|^2, \quad \sum_{\lambda \in \Lambda} x_\lambda \overline{y_\lambda} = \langle x, y \rangle$$

Dans le cas d'un espace de Hilbert séparable (de dimension infinie) toute base est dénombrable et peut donc être indexée par l'ensemble \mathbb{Z} . On a alors:

Proposition 1.7. Soit $\{e_n\}_{n \in \mathbb{Z}}$ une base orthonormée de H . L'application $x \rightarrow \{x_n\}_{n \in \mathbb{Z}}$ est une application linéaire bijective et isométrique de H sur $\ell^2(\mathbb{Z})$, dont l'application réciproque est donnée par $\{x_n\}_{n \in \mathbb{Z}} \rightarrow \sum_{n \in \mathbb{Z}} x_n e_n$.

Autrement dit, tout espace de Hilbert séparable est isomorphe à $\ell^2(\mathbb{Z})$. La notion de projection orthogonale sera fondamentale dans la suite du cours et résulte de la décomposition orthogonale de l'espace H donnée en 1.4 (2).

Proposition 1.8. Soit F un sous espace fermé de H , on désigne par $P_F(x)$ (projection orthogonale de x sur F) la surjection continue de H sur F vérifiant les deux propriétés équivalentes:

1. $\langle x - P_F(x), y \rangle = 0$ pour $\forall y \in F$.
2. $\|x - P_F(x)\| = \inf_{y \in F} \|x - y\|$.

Dans la plupart des cas, on calcule une projection orthogonale en utilisant une base orthonormée:

Proposition 1.9. Soit $\{e_n\}_{n \in \mathbb{Z}}$ une base orthonormée de H , telle que $\{e_n\}_{n \leq p}$ soit une base orthonormée de F , alors $P_F(x) = \sum_{-\infty}^p \langle x, e_n \rangle e_n$

Il est facile de prouver la propriété de "continuité" suivante:

Proposition 1.10. Soit F_n une suite de sous espaces de l'espace de Hilbert H et $x \in H$. On pose $x_n = \text{Proj}(x|F_n)$

1. Si la suite F_n est décroissante et $F = \bigcap_n F_n$ alors x_n converge vers $\text{Proj}(x|F)$.
2. Si la suite F_n est croissante et $F = \overline{\bigcup_n F_n}$ alors x_n converge vers $\text{Proj}(x|F)$.

Nous terminons cette section par une propriété d'extension des isométries qui nous sera très utile.

Proposition 1.11. Soit F_1 (resp F_2) un sous espace dense de l'espace de Hilbert H_1 (resp H_2). Alors si W est une application linéaire isométrique et surjective de F_1 sur F_2 elle se prolonge de façon unique en une isométrie de H_1 sur H_2

Les espaces de Hilbert les plus utilisés dans ce cours seront les espaces de suites ou de fonctions complexes de carré sommable.

1.1.2 Rappels d'Intégration

Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, μ étant une mesure positive que l'on supposera toujours σ finie. Pour $1 \leq p \leq \infty$ on désigne par L^p l'espace de Banach des classes d'équivalence (pour l'égalité μ presque sûre) des fonctions complexes mesurables telles que les quantités

- $\|f\|_p = \left(\int_{\Omega} |f|^p d\mu \right)^{1/p}$ si $p < \infty$
- $\|f\|_{\infty} = \inf\{c; |f| \leq c \mu\text{-p.p.}\}$

soient finies (ce sont alors des normes sur L^p).

Proposition 1.12. *L'espace L^2 muni du produit scalaire*

$$\langle f, g \rangle = \int_{\Omega} f(\omega) \overline{g(\omega)} d\mu(\omega)$$

est un espace de Hilbert.

Proposition 1.13. (Inégalité de Hölder) *Soit $p \geq 1, q \geq 1$ avec $1/p + 1/q = 1$ et $f \in L^p, g \in L^q$. Alors $fg \in L^1$ et $\|fg\|_1 \leq \|f\|_p \|g\|_q$*

Proposition 1.14. (Lien avec la convergence presque partout) *Soit f_n une suite de L^p qui converge dans L^p vers $f \in L^p$. Alors il existe une sous suite extraite de f_n qui converge μ -p.p. vers f .*

En particulier ceci signifie que si une suite f_n converge vers f_1 dans L^p et f_n converge vers f_2 μ -p.p. alors $f_1 = f_2$ μ -p.p.

Proposition 1.15. (Théorème de Lebesgue) *Soit f_n une suite de fonctions de L^1 qui converge μ -p.p. vers f et telle qu'il existe $g \in L^1$ avec $|f_n| \leq g$ pour tout n . Alors $f \in L^1$ et f_n converge vers f dans L^1 .*

Nous utiliserons très souvent le théorème de Fubini. Soit $(\Omega_i, \mathcal{A}_i, \mu_i), i = 1, 2$ des espaces mesurés, les deux mesures μ_i étant positives et σ finies.

Théorème 1.16. *Soit $f(\omega_1, \omega_2)$ une fonction $\mathcal{A}_1 \otimes \mathcal{A}_2$ mesurable.*

1. Si $f \geq 0$ alors

$$\int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) \right) d\mu_1(\omega_1) = \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1) \right) d\mu_2(\omega_2)$$

Cette valeur commune étant finie ou non.

2. Pour f non nécessairement positive la formule ci-dessus est toujours valable si l'on s'assure au préalable que cette formule appliquée à la fonction positive $|f|$ fournit une valeur finie. Toutes les intégrales écrites sont alors absolument convergentes et définissent l'intégrale de f par rapport à la mesure produit $\mu_1 \otimes \mu_2$.

On supposera $1 \leq p < \infty$ dans les énoncés suivants de cette section.

Proposition 1.17. *Soit f_n une suite de fonctions de L^p telle que $\sum_n \|f_n\|_p < \infty$. Alors la série $\sum_n f_n$ converge μ -p.p. absolument et dans L^p (vers une fonction de L^p notée $\sum_n f_n$). En particulier pour $p = 1$ on aura donc $\int \sum_n f_n = \sum_n \int f_n$.*

Démonstration. On pose $S_n = \sum_{k=1}^n f_k$ et $h_n = \sum_{k=1}^n |f_k|$, $h = \sum_{k=1}^{\infty} |f_k|$. On a

$$\|h\|_p = \lim_n \uparrow \|h_n\|_p \leq \sum_{k=1}^{\infty} \|f_k\|_p < \infty$$

On en déduit $h \in L^p$ et par conséquent que h est finie presque partout (et donc $\sum_n f_n$ converge μ -p.p. absolument). D'autre part pour $m > n$, $\|S_m - S_n\|_p \leq \sum_{k=n+1}^m \|f_k\|_p$ ce qui prouve que la suite S_n est de Cauchy dans L^p en utilisant la propriété de Cauchy de la série convergente $\sum_k \|f_k\|_p$. L'espace L^p étant complet la série $\sum_n f_n$ est donc convergente dans L^p . \square

Corollaire 1.18. Soit $a_n \in \ell^1(\mathbb{Z})$ et f_n une suite de fonctions bornée dans L^p . Alors la série $\sum_n a_n f_n$ converge μ -p.p. absolument et dans L^p

Dans le cas de séries orthogonales dans L^2 on a un résultat voisin (il s'agit de la somme des carrés des normes!) mais on perd en général la convergence presque sûre:

Proposition 1.19. Soit f_n une suite de fonctions orthogonales dans L^2 .

1. Si $\sum_n \|f_n\|_2^2 < \infty$ la série $\sum_n f_n$ converge dans L^2 .
2. Soit $a_n \in \ell^2(\mathbb{Z})$ et f_n une suite bornée dans L^2 . Alors la série $\sum_n a_n f_n$ converge dans L^2 . Si de plus $a_n \in \ell^1(\mathbb{Z})$ alors la série converge μ -p.p.

Démonstration. On vérifie immédiatement que la série est de Cauchy dans L^2 car en adoptant les notations de la preuve de 1.17 on a d'après le "théorème de Pythagore":

$$\|S_m - S_n\|_2^2 = \sum_{k=n+1}^m \|f_k\|_2^2$$

Il suffit alors d'utiliser la propriété de Cauchy de la série convergente $\sum_k \|f_k\|_2^2$. La seconde assertion est un corollaire immédiat de 1.18. \square

Proposition 1.20. L'ensemble des combinaisons linéaires à coefficients complexes de fonctions indicatrices d'ensembles mesurables de μ mesure finie est dense dans L^p . Dans le cas de $\Omega = \mathbb{R}$ ou \mathbb{T} muni de leur tribu Borélienne on peut se contenter d'indicatrices d'intervalles. De plus dans le cas de \mathbb{R} , l'ensemble des fonctions continues à support compact est dense dans L^p et de même pour l'ensemble des fonctions continues dans le cas de \mathbb{T} .

On remarquera que:

- Si μ est une mesure finie on a les inclusions: $L^\infty \subset L^2 \subset L^1$
- Par contre $\ell^1(\mathbb{Z}) \subset \ell^2(\mathbb{Z}) \subset \ell^\infty(\mathbb{Z})$

mais qu'aucune inclusion de ce genre n'est valide sur \mathbb{R} muni de la mesure de Lebesgue.

1.2 Séries de Fourier

On identifie les fonctions sur $[0, 1[$ aux fonctions sur \mathbb{R} périodiques et de période 1 et on désigne par \mathbb{T} l'intervalle $[0, 1[$ muni de la topologie du tore \mathbb{R}/\mathbb{Z} qui en fait un espace compact. (Attention, la restriction à $[0, 1[$ d'une fonction f continue sur \mathbb{R} ne définit une fonction continue sur \mathbb{T} que si $f(0) = f(1)$!). On écrira donc indifféremment:

$$\int_{\mathbb{T}} f(x) dx \quad , \quad \int_0^1 f(x) dx \quad , \quad \int_a^{a+1} f(x) dx$$

1.2.1 Coefficients de Fourier d'une fonction de $L^2(\mathbb{T}, dx)$

Cette section est essentiellement une application de la proposition 1.7. Il nous faut donc exhiber une base orthonormée de l'espace de Hilbert $L^2(\mathbb{T}, dx)$.

Définition 1.21. On pose $\gamma_n(t) = \exp(-2i\pi nt)$

Tout ce qui suit est aussi valable avec le choix $\gamma_n(t) = \exp(2i\pi nt)$ sauf pour certains calculs de dérivation ou d'intégration des fonctions $\gamma_n(t)$ pour lesquels il faut changer de signe...

Définition 1.22. Soit $f \in L^2(\mathbb{T}, dx)$. On pose:

$$\hat{f}(n) = \langle f, \gamma_n \rangle = \int_0^1 f(t) \gamma_{-n}(t) dt, \quad S_n f(t) = \sum_{k=-n}^n \hat{f}(k) \gamma_k(t)$$

On vérifie facilement que la famille $\{\gamma_n\}_{n \in \mathbb{Z}}$ est une famille orthonormée. Il s'en suit que la suite $\hat{f}(n)$ des coefficients de Fourier de f est de carré sommable.

Théorème 1.23. La famille $\{\gamma_n\}_{n \in \mathbb{Z}}$ est un système total dans $L^2(\mathbb{T}, dx)$ et est donc une base orthonormée de $L^2(\mathbb{T}, dx)$.

Démonstration. Soit $E = C(\mathbb{T})$ l'espace de Banach (complexe) des fonctions complexes continues sur \mathbb{T} muni de la norme uniforme et F un sous espace vectoriel de E . Le théorème de Stone Weierstrass permet d'affirmer que F est dense dans E s'il possède les propriétés suivantes:

- F contient les fonctions constantes.
- F est stable par multiplication et par conjugaison.
- Pour tout couple (x, y) de points distincts de \mathbb{T} il existe un élément $f \in F$ tel que $f(x) \neq f(y)$.

On a bien sûr l'inclusion (ensembliste) $E \subset L^2(\mathbb{T}, dx)$. Soit F l'espace des combinaisons linéaires (complexes) des fonctions $\gamma_n(t)$, il est alors facile de vérifier que F vérifie les trois conditions ci-dessus et F est donc dense dans E pour la norme uniforme. Soit alors $f \in L^2(\mathbb{T}, dx)$ et $\epsilon > 0$. D'après 1.20 il existe une fonction $g \in E$ telle que $\|f - g\|_2 \leq \epsilon/2$. Pour cette fonction g il existe $h \in F$ telle que la norme uniforme $\sup_{x \in \mathbb{T}} |g(x) - h(x)| \leq \epsilon/2$ et donc $\|h - g\|_2 \leq \epsilon/2$. \square

Remarque 1.24. On peut prouver de la même façon que l'espace des combinaisons linéaires à coefficients complexes des fonctions $\gamma_n(t)$ est dense dans $L^p(\mathbb{T}, dx)$ pour $1 \leq p < \infty$

Cette preuve du fait que les combinaisons linéaires d'exponentielles sont denses dans $L^2(\mathbb{T}, dx)$ n'est pas très constructive, c'est pourquoi l'on va fournir une seconde preuve, utilisant les propriétés du noyau de Poisson du disque. Pour ce faire nous aurons besoin des quelques lemmes énoncés ci-dessous.

Définition 1.25. On note \mathcal{T}_u l'opérateur de translation $\mathcal{T}_u f(t) = f(t - u)$

Lemme 1.26. Soit f une fonction de $L^p(\mathbb{T}, dx)$ avec $1 \leq p < \infty$. Alors l'application $u \rightarrow \|f - \mathcal{T}_u f\|_p$ est continue.

Démonstration. Nous utiliserons souvent le raisonnement suivant: Pour prouver qu'une propriété est vraie sur un espace topologique E on commence par vérifier qu'elle est vraie sur une partie dense $F \subset E$ puis on montre ensuite qu'elle "passe à la limite". Il est facile ici (par Lebesgue) de prouver que cette propriété de continuité est vraie sur l'espace vectoriel F engendré par les exponentielles. Soit $f \in L^p(\mathbb{T}, dx)$ et $\epsilon > 0$. D'après 1.24 il existe g dans F telle que $\|f - g\|_p \leq \epsilon/4$ et on peut écrire:

$$\begin{aligned} \|f - \mathcal{T}_u f\|_p &\leq \|f - g\|_p + \|g - \mathcal{T}_u g\|_p + \|\mathcal{T}_u g - \mathcal{T}_u f\|_p \\ &= 2\|f - g\|_p + \|g - \mathcal{T}_u g\|_p \leq \epsilon/2 + \|g - \mathcal{T}_u g\|_p \end{aligned}$$

et il suffit d'appliquer la propriété à g . \square

Définition 1.27. Soit $0 \leq r < 1$. On définit la fonction continue $P_r(t)$ sur \mathbb{T} (noyau de Poisson) par le développement en série uniformément convergent sur le tore \mathbb{T} soit

$$P_r(t) = \sum_{n \in \mathbb{Z}} r^{|n|} \gamma_n(t) = \sum_{n \in \mathbb{Z}} r^{|n|} \gamma_{-n}(t)$$

et pour une fonction $f \in L^1(\mathbb{T}, dx)$ on pose

$$P_r f(t) = \int_{\mathbb{T}} f(t-u) P_r(u) du = \int_{\mathbb{T}} f(u) P_r(t-u) du$$

Proposition 1.28. Soit $0 \leq r < 1$ et $f \in L^2(\mathbb{T}, dx)$.

1. $P_r(t) = \frac{1-r^2}{|1-r\gamma_1(t)|^2}$ et donc $P_r(t) \geq 0$.
2. $\int_{\mathbb{T}} P_r(t) dt = 1$

Démonstration. La propriété (1) s'obtient par sommation et pour la propriété (2) il suffit de constater que la série définissant $r(t)$ converge dans $L^1(\mathbb{T}, dx)$ (d'après 1.18). Elle est donc intégrable terme à terme sur \mathbb{T} et l'on remarque que $\int_{\mathbb{T}} \gamma_n(t) = \delta_0^n$. \square

Lemme 1.29. Soit g une fonction intégrable sur \mathbb{T} et continue à l'origine. Alors l'intégrale $\int_{\mathbb{T}} P_r(u) g(u) du$ tend vers $g(0)$ lorsque $r \rightarrow 1$.

Démonstration. On peut supposer $g(0) = 0$. Soit $\epsilon > 0$, il existe un voisinage \mathcal{V} de x dans \mathbb{T} tel que si $x \in \mathcal{V}$ alors $|g(x)| \leq \epsilon/2$. On peut donc écrire:

$$\begin{aligned} \left| \int_{\mathbb{T}} P_r(u) g(u) du \right| &\leq \int_{\mathcal{V}} P_r(u) |g(u)| du + \int_{\mathbb{T} \setminus \mathcal{V}} P_r(u) |g(u)| du \\ &\leq \epsilon/2 + \|g\|_1 \sup_{u \in \mathbb{T} \setminus \mathcal{V}} P_r(u) \end{aligned}$$

Il est maintenant facile de constater que hors d'un voisinage \mathcal{V} de zéro la fonction $P_r(u)$ converge uniformément vers 0 lorsque $r \rightarrow 1$. \square

Proposition 1.30. Soit $0 \leq r < 1$ et $f \in L^2(\mathbb{T}, dx)$.

1. $P_r f(t) = \sum_{n \in \mathbb{Z}} \widehat{f}(n) r^{|n|} \gamma_n(t)$
2. $P_r f \rightarrow f$ dans $L^2(\mathbb{T}, dx)$ lorsque $r \rightarrow 1$.

Démonstration. Pour prouver (1) on remarque que:

$$P_r f(t) = \int_{\mathbb{T}} \sum_{n \in \mathbb{Z}} r^{|n|} \gamma_n(t) \gamma_{-n}(u) f(u) du = \sum_{n \in \mathbb{Z}} r^{|n|} \gamma_n(t) \langle f, \gamma_n \rangle$$

La seconde égalité provient du fait que la série $\sum_{n \in \mathbb{Z}} \|r^{|n|} \gamma_n(t) \gamma_{-n} f\|_1$ est convergente et l'on applique 1.17.

Pour prouver (2) on peut écrire:

$$\begin{aligned} |P_r f(t) - f(t)|^2 &= \left| \int_{\mathbb{T}} (f(t-u) - f(t)) P_r(u) du \right|^2 = \\ &= \left| \int_{\mathbb{T}} (f(t-u) - f(t)) \sqrt{P_r(u)} \sqrt{P_r(u)} du \right|^2 \leq \int_{\mathbb{T}} |f(t-u) - f(t)|^2 P_r(u) du \end{aligned}$$

En utilisant le théorème de Fubini pour les fonctions positives on obtient:

$$\|P_r f - f\|_2^2 = \int_{\mathbb{T}} |P_r f(t) - f(t)|^2 dt \leq \int_{\mathbb{T}} \|f - \tau_u f\|_2^2 P_r(u) du$$

et il suffit d'appliquer le lemme 1.29 avec $g(u) = \|f - \tau_u f\|_2^2$ □

Cette Proposition permet de redémontrer le fait que les exponentielles forment un système total dans $L^2(\mathbb{T}, dx)$ en utilisant 1.4. En effet si la fonction $f \in L^2(\mathbb{T}, dx)$ est orthogonale aux exponentielles on a $\hat{f}(n) = 0$ pour tout $n \in \mathbb{Z}$ et donc $P_r(f) = 0$ pour tout $0 \leq r < 1$ et il s'en suit que f est nulle (au sens de $L^2(\mathbb{T}, dx)$).

Théorème 1.31. Soit $f \in L^2(\mathbb{T}, dx)$, l'application $f \longrightarrow \{\hat{f}(n)\}_{n \in \mathbb{Z}}$ est une application linéaire bijective et isométrique de $L^2(\mathbb{T}, dx)$ sur $\ell^2(\mathbb{Z})$. L'application réciproque est définie par

$$\{\hat{f}(n)\}_{n \in \mathbb{Z}} \longrightarrow \sum_{n \in \mathbb{Z}} \hat{f}(n) \gamma_n$$

cette dernière série étant convergente dans $L^2(\mathbb{T}, dx)$.

Remarquons enfin que l'on peut traiter des fonctions périodiques de période quelconque en se ramenant sur le tore \mathbb{T} par un simple changement de variable. Pour un intervalle $[A, B]$ on pose $\tau = \frac{1}{B-A}$ et $H = L^2([A, B], \tau dx)$.

Proposition 1.32. Soit $f \in H$

1. La famille $\gamma_{n\tau}(t)$ est une base orthonormée de H
2. Si on pose $\hat{f}(n) = \langle f, \gamma_{n\tau} \rangle_H = \tau \int_A^B f(x) \gamma_{-n\tau}(x) dx$
 - (a) $\tau \int_A^B |f(x)|^2 dx = \sum_{n \in \mathbb{Z}} |\hat{f}(n)|^2$
 - (b) $f = \sum_n \hat{f}(n) \gamma_{n\tau}$, la série étant convergente dans H .

1.2.2 Transformation de Fourier d'une suite de $\ell^2(\mathbb{Z})$

Définition 1.33. Pour une suite $a \in \ell^2(\mathbb{Z})$ on définit sa transformée de Fourier $\mathcal{F}a$ comme l'élément de $L^2(\mathbb{T}, dx)$ somme de la série:

$$\{a_n\}_{n \in \mathbb{Z}} \longrightarrow \mathcal{F}a = \sum_{n \in \mathbb{Z}} a_n \gamma_{-n}$$

cette dernière série étant convergente dans $L^2(\mathbb{T}, dx)$.

De même on définit $\tilde{\mathcal{F}}a$ par:

$$\{a_n\}_{n \in \mathbb{Z}} \longrightarrow \tilde{\mathcal{F}}a = \sum_{n \in \mathbb{Z}} a_n \gamma_n$$

Remarque 1.34. Si l'on avait décidé de noter $\mathcal{F}f(n)$ les coefficients de Fourier de $f \in L^2(\mathbb{T}, dx)$ le théorème 1.31 est alors équivalent à $\mathcal{F} \circ \tilde{\mathcal{F}} = \tilde{\mathcal{F}} \circ \mathcal{F} = \text{Identité}$.

On peut constater que si $a \in \ell^1(\mathbb{Z})$ alors la série définissant $\mathcal{F}a$ converge uniformément sur \mathbb{T} vers une fonction continue sur \mathbb{T} . On note alors $\mathcal{I}a$ la fonction continue obtenue, de préférence à $\mathcal{F}a$ pour bien indiquer qu'il s'agit d'une fonction définie point par point et non d'une classe d'équivalence (En fait cette distinction de notation est vite oubliée dans les applications...) Par exemple la série de Fourier d'une fonction f telle que $f(t) = f(0) + \int_0^t g(u) du$ avec $g \in L^2(\mathbb{T}, dx)$ et $\int_{\mathbb{T}} g(u) du = 0$ est sommable et c'est donc le cas en particulier pour des fonctions f de classe \mathbf{C}^1 sur le tore. Il existe une littérature extensive reliant les propriétés de la suite a et celles de sa transformée de Fourier, montrant "en gros" que plus la suite a décroît rapidement, plus sa transformée de Fourier est régulière.

1.2.3 Coefficients de Fourier d'une fonction de $L^1(\mathbb{T}, dx)$

Lorsque $f \in L^1(\mathbb{T}, dx)$ on peut toujours définir ses coefficients de Fourier par la formule $\hat{f}(n) = \int_0^1 f(t) \gamma_n(t) dt$. Cette suite n'est plus nécessairement dans $\ell^2(\mathbb{Z})$ mais:

Proposition 1.35. Si $f \in L^1(\mathbb{T}, dx)$ on a:

1. $|\hat{f}(n)| \leq \|f\|_1$
2. $\hat{f}(n)$ tend vers 0 à l'infini.

Démonstration. La propriété (1) est évidente. Il est facile de montrer la propriété (2) pour les indicatrices d'intervalles et donc pour l'espace vectoriel F de leurs combinaisons linéaires qui est dense dans $L^1(\mathbb{T}, dx)$ (on peut aussi choisir de le vérifier sur les exponentielles). Soit $f \in L^1(\mathbb{T}, dx)$ et $\epsilon > 0$. Il existe g dans F telle que $\|f - g\|_1 \leq \epsilon/2$ et on peut écrire:

$$|\hat{f}(p)| \leq |\hat{f}(p) - \hat{g}(p)| + |\hat{g}(p)| \leq \|f - g\|_1 + |\hat{g}(p)| \leq \epsilon/2 + |\hat{g}(p)|$$

et il suffit d'appliquer la propriété à g . □

Pour $f \in L^2(\mathbb{T}, dx)$ on a vu que la suite $S_n f$ converge dans $L^2(\mathbb{T}, dx)$. Ce n'est plus vrai si l'on remplace $L^2(\mathbb{T}, dx)$ par $L^1(\mathbb{T}, dx)$ mais nous allons prouver que dans ce cas on a la propriété plus faible de convergence des moyennes arithmétiques:

$$\sigma_n f = \frac{S_0 f + \dots + S_n f}{n+1}$$

Définition 1.36. On définit le noyau de Féjer $K_n(t)$ par la formule:

$$K_n(t) = \sum_{k=-n}^n \left(1 - \frac{|k|}{n+1}\right) \gamma_k(t) = \frac{1}{n+1} \sum_{k=1}^{n+1} \sum_{r=1}^{n+1} \gamma_k(t) \gamma_{-r}(t)$$

Pour $f \in L^1(\mathbb{T}, dx)$ on définit:

$$K_n f(t) = \int_{\mathbb{T}} f(t-u) K_n(u) du = \int_{\mathbb{T}} f(u) K_n(t-u) du$$

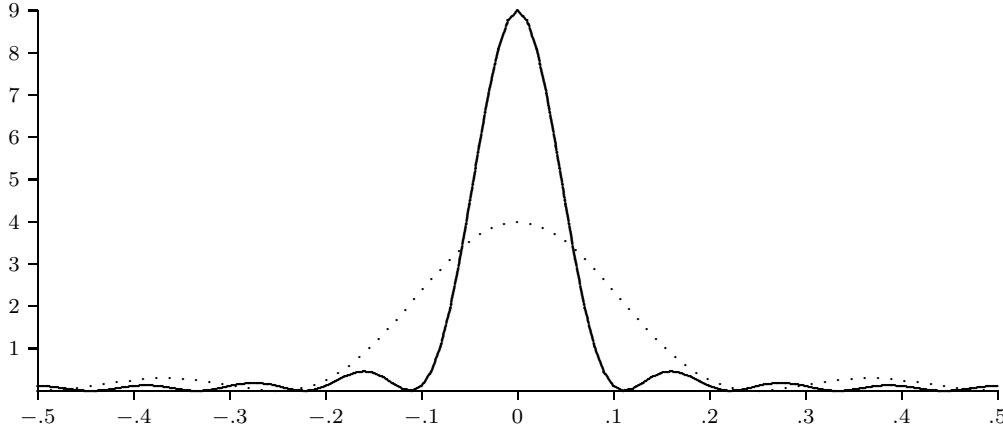


Figure 1.1 Le graphique des fonctions de Féjer pour $n = 3$ (trait solide) et $n = 9$ (pointillé). Les intégrales des deux fonctions valent 1, mais on voit que, pour n grand, la fonction est très vite près de 0 dès qu'on s'éloigne de l'origine.

Proposition 1.37. Le noyau de Féjer possède les propriétés suivantes:

1. $K_n(t) = \frac{1}{n+1} \left(\frac{\sin((n+1)\pi t)}{\sin(\pi t)} \right)^2$ et donc $K_n(t) \geq 0$.
2. $\int_{\mathbb{T}} K_n(t) dt = 1$
3. $K_n f(t) = \sigma_n f(t) = \sum_{k=-n}^n \left(1 - \frac{|k|}{n+1}\right) \widehat{f}(k) \gamma_k(t)$
4. $\sigma_n f \rightarrow f$ dans $L^1(\mathbb{T}, dx)$ lorsque $n \rightarrow \infty$.

Démonstration. Les assertions (1) et (2) et (3) s'obtiennent par un calcul facile. La preuve de (4) est tout à fait analogue à celle faite dans le cas du noyau de Poisson et l'on peut en résumer les principales étapes:

1. Soit $f \in L^1(\mathbb{T}, dx)$. Alors l'application $u \rightarrow \|f - \mathcal{T}_u f\|_1$ est continue
2. Soit g une fonction intégrable sur \mathbb{T} et continue à l'origine. Alors $\int_{\mathbb{T}} K_n(u) g(u) du \rightarrow g(0)$ lorsque $n \rightarrow \infty$.
3. On a les inégalités:

$$|K_n f(t) - f(t)| \leq \left| \int_{\mathbb{T}} |f(t-u) - f(t)| K_n(u) du \right|$$

$$\int_{\mathbb{T}} |K_n f(t) - f(t)| dt \leq \int_{\mathbb{T}} \|f - \mathcal{T}_u f\|_1 K_n(u) du$$

et il suffit d'appliquer la rubrique (2) avec $g(u) = \|f - \mathcal{T}_u f\|_1$

La Proposition est donc prouvée. \square

Remarque 1.38. Une conséquence immédiate de cette proposition est qu'une fonction de $L^1(\mathbb{T}, dx)$ est définie par la suite de ses coefficients de Fourier.

Nous n'avons établi jusqu'à présent que la convergence en moyenne des séries de Fourier. La proposition suivante permet d'obtenir une convergence ponctuelle mais vers une version "régularisée" de f plutôt que vers la fonction f elle-même.

Définition 1.39. Si f est une fonction sur \mathbb{T} on définit

$$\tilde{f}(t) = \frac{1}{2} \lim_{u \rightarrow 0} (f(t+u) + f(t-u))$$

lorsque cette limite existe.

Proposition 1.40. Soit $f \in L^1(\mathbb{T}, dx)$ et $t_0 \in \mathbb{T}$ un point où $\tilde{f}(t_0)$ existe. On a alors:

1. $\sigma_n f(t_0) \rightarrow \tilde{f}(t_0)$ lorsque $n \rightarrow \infty$.
2. Si la suite $n\hat{f}(n)$ est bornée alors $S_n f(t_0) \rightarrow \tilde{f}(t_0)$ lorsque $n \rightarrow \infty$.

Démonstration. On peut écrire:

$$\begin{aligned} \sigma_n f(t_0) &= \int_{-1/2}^0 f(t_0 - u) K_n(u) du + \int_0^{1/2} f(t_0 - u) K_n(u) du \\ &= \int_0^{1/2} f(t_0 + u) K_n(u) du + \int_0^{1/2} f(t_0 - u) K_n(u) du \\ \sigma_n f(t_0) - \tilde{f}(t_0) &= 2 \int_0^{1/2} \left[(f(t_0 - u) + f(t_0 + u))/2 - \tilde{f}(t_0) \right] K_n(u) du \end{aligned}$$

En découpant cette dernière intégrale sur un voisinage de 0 et son complémentaire on obtient la première partie de la proposition.

La seconde repose essentiellement sur le lemme suivant, dit de Hardy-Landau:

Lemme 1.41. Soit u_n est une suite de nombres complexes. On pose $s_n = u_1 + u_2 + \dots + u_n$ et $\sigma_n = \frac{s_1 + s_2 + \dots + s_n}{n}$. Si l'on suppose que la suite nu_n est bornée, alors la convergence de σ_n vers une limite implique la convergence de s_n vers la même limite.

Il existe une version "convergence" uniforme de cette propriété:

Proposition 1.42. Soit $f \in L^1(\mathbb{T}, dx)$, continue sur un intervalle $[a, b] \subset \mathbb{T}$. On a alors:

1. $\lim_{n \rightarrow \infty} \sigma_n f(t) = f(t)$ uniformément pour $t \in [a, b]$.
2. Si la suite $n\hat{f}(n)$ est bornée alors $\lim_{n \rightarrow \infty} S_n f(t) = f(t)$ uniformément pour $t \in [a, b]$.

1.3 Transformation de Fourier sur \mathbb{R}

Définition 1.43. Soit $f \in L^1(\mathbb{R}, dx)$. On définit alors:

$$\mathcal{I}f(t) = \int_{\mathbb{R}} f(x) \gamma_{-t}(x) dx, \quad \tilde{\mathcal{I}}f(t) = \int_{\mathbb{R}} f(x) \gamma_t(x) dx$$

Si l'on veut étendre par analogie au cas de $L^2(\mathbb{R}, dx)$ la notion de transformation de Fourier définie dans le cas d'une suite $a \in \ell^2(\mathbb{Z})$ par $\mathcal{F}a(t) = \sum_{n \in \mathbb{Z}} a_n \gamma_{-n}(t)$ on constate que malheureusement l'intégrale $\mathcal{I}f(t)$ n'est définie que pour $f \in L^1(\mathbb{R}, dx)$ et il faudra donc modifier cette définition pour introduire la transformation de Fourier sur $L^2(\mathbb{R}, dx)$.

1.3.1 Transformation de Fourier sur $L^1(\mathbb{R}, dx)$

On notera par \mathbf{C}_0 l'espace des fonctions complexes continues sur \mathbb{R} qui tendent vers zéro à l'infini.

Proposition 1.44. Soit $f \in L^1(\mathbb{R}, dx)$ alors $\mathcal{I}f(t) = \int_{\mathbb{R}} f(u)\gamma_{-t}(u) du$ est une fonction de \mathbf{C}_0 .

Démonstration. La continuité (et même la continuité uniforme) résulte de:

$$|\mathcal{I}f(t+h) - \mathcal{I}f(t)| \leq \int |\gamma_h(x) - 1| |f(x)| dx$$

et on applique Lebesgue. Pour la limite à l'infini, la preuve est tout à fait analogue à celle de 1.35: On commence par montrer cette propriété pour une indicatrice d'intervalle puis on montre que cette propriété "passe à la limite" dans $L^1(\mathbb{R}, dx)$ en utilisant la majoration $\sup_{x \in \mathbb{R}} |\mathcal{I}f(x)| \leq \|f\|_1$. \square

Les propositions suivantes établissent un lien entre la décroissance à l'infini d'une fonction et la régularité de son intégrale de Fourier. (ou l'inverse).

Proposition 1.45. Soit $f \in L^1(\mathbb{R}, dx)$ telle que $g(x) = xf(x) \in L^1(\mathbb{R}, dx)$.

1. $\mathcal{I}f$ est de classe \mathbf{C}^1 et $(\mathcal{I}f)' = 2i\pi\mathcal{I}g$.
2. Si $x^p f(x)$ est intégrable alors $\mathcal{I}f$ est de classe \mathbf{C}^p

Démonstration. L'assertion (1) est une simple application du théorème de dérivation "sous le signe somme". Pour obtenir (2) on utilise l'inégalité de Hölder pour montrer que si f et $x^p f(x)$ sont intégrables il en est de même de $x^r f(x)$ pour $0 \leq r \leq p$. Il suffit alors d'appliquer (1) p fois. \square

Proposition 1.46. Soit $f \in L^1(\mathbb{R}, dx)$ de classe \mathbf{C}^1 telle que $f'(x) \in L^1(\mathbb{R}, dx)$.

1. f est dans \mathbf{C}_0 .
2. $\mathcal{I}(f')(t) = -(2i\pi t)\mathcal{I}f(t)$.
3. Si f est de classe \mathbf{C}^p et si toutes ses dérivées jusqu'à l'ordre p sont intégrables alors $t^p \mathcal{I}f(t) \in \mathbf{C}_0$.

Démonstration. On peut écrire pour $x > 0$:

$$f(x) = f(0) + \int_0^x f'(u) du, \quad \lim_{x \rightarrow \infty} f(x) = f(0) + \int_0^\infty f'(u) du$$

ce qui prouve que f a une limite en $+\infty$ (et de même en $-\infty$). La fonction f étant intégrable cette limite est nécessairement nulle ce qui prouve (1). De plus en utilisant une intégration par parties:

$$\mathcal{I}(f')(t) = \left[f(x)\gamma_{-t}(x) \right]_{x=-\infty}^{x=\infty} - 2i\pi t \mathcal{I}f(t) = -2i\pi t \mathcal{I}f(t)$$

On obtient (3) en appliquant p fois (1) et (2). \square

On peut étendre au cas continu la notion de noyau de Féjer introduite dans les séries de Fourier:

Définition 1.47. Pour $T > 0$ on définit le noyau de Féjer:

$$K_T(t) = \int_{-T}^T (1 - \frac{|u|}{T}) \gamma_t(u) du = \frac{1}{T} \int_0^T \int_0^T \gamma_t(u) \gamma_{-t}(v) dudv$$

Et pour $f \in L^1(\mathbb{R}, dx)$ on pose:

$$K_T f(t) = \int f(t-u) K_T(u) du = \int f(u) K_T(t-u) du$$

Cette définition a un sens puisque K_T est une fonction bornée.

Proposition 1.48. Le noyau de Féjer possède les propriétés suivantes:

1. $K_T(t) = T \left(\frac{\sin(\pi T t)}{\pi T t} \right)^2$ et donc $K_T(t) \geq 0$.
2. $\int K_T(t) dt = \frac{1}{\pi} \int \left(\frac{\sin x}{x} \right)^2 dx = 1$.

Démonstration. L'assertion (1) résulte d'un calcul simple. La Preuve de (2) est plus difficile car l'on ne peut pas utiliser directement le théorème de Fubini, mais l'on peut se servir de la même propriété du noyau de Féjer sur le tore :

$$\begin{aligned} 1 &= \frac{1}{n+1} \int_{-1/2}^{1/2} \left(\frac{\sin((n+1)\pi t)}{\sin(\pi t)} \right)^2 dt = \int_{-1/2}^{1/2} \left(\frac{\sin((n+1)\pi t)}{(n+1)\pi t} \right)^2 \left(\frac{\pi t}{\sin(\pi t)} \right)^2 (n+1) dt = \\ &= \frac{1}{\pi} \int_{-(n+1)\pi/2}^{(n+1)\pi/2} \left(\frac{\sin x}{x} \right)^2 \left(\frac{x/(n+1)}{\sin(x/(n+1))} \right)^2 dx \end{aligned}$$

Le résultat est maintenant une conséquence du théorème de Lebesgue. \square

Lemme 1.49. Soit g une fonction bornée, continue à l'origine. Alors $\int K_T(u)g(u) du \rightarrow g(0)$ lorsque $T \rightarrow \infty$.

Démonstration. Au vu de la proposition précédente et du fait que $\lim_{T \rightarrow \infty} \int_{|t| > \alpha} K_T(t) dt = 0$ la preuve est identique à celle du noyau de Féjer sur le tore si l'on prend en compte le lemme suivant déjà démontré sur le tore. \square

Lemme 1.50. Soit f une fonction dans $L^p(\mathbb{R}, dx)$ avec $1 \leq p < \infty$. Alors l'application $u \rightarrow \|f - \mathcal{T}_u f\|_p$ est continue bornée.

Démonstration. On prouve cette propriété sur les indicatrices d'intervalles puis on montre qu'elle "passe à la limite" dans L^p et enfin on utilise 1.20 \square

Proposition 1.51. Soit $f \in L^1(\mathbb{R}, dx)$:

1. $K_T f(t) = \int_{-T}^T (1 - \frac{|u|}{T}) \mathcal{I}f(u) \gamma_t(u) du$
2. $K_T f$ converge vers f dans $L^1(\mathbb{R}, dx)$ lorsque $T \rightarrow \infty$.
3. Si $\mathcal{I}f \in L^1(\mathbb{R}, dx)$ alors une version de la limite précédente est donnée par $\widetilde{\mathcal{I}\mathcal{I}f}$. Il s'en suit que f possède une version dans \mathbf{C}_0 .

Démonstration. L'affirmation (1) est une conséquence de Fubini en utilisant la forme de K_T utilisée dans 1.48 et (2) est maintenant "classique" au vu des preuves précédentes et si l'on prend en compte le lemme 1.50. (on peut d'ailleurs axiomatiser cette situation pour éviter des preuves à répétitions). La conclusion(3) est une application directe du théorème de Lebesgue. \square

Remarque 1.52. Soit $f \in L^1(\mathbb{R}, dx)$.

- L'application $f \rightarrow \mathcal{I}f$ est injective sur $L^1(\mathbb{R}, dx)$.
- Tout ce qui précède reste vrai si l'on remplace γ_t par γ_{-t} et par conséquent si $\tilde{\mathcal{I}}f$ est intégrable on a $f = \mathcal{I}(\tilde{\mathcal{I}}f)$.

Corollaire 1.53. L'ensemble des fonctions de $L^1(\mathbb{R}, dx)$ possédant une transformée de Fourier à support compact est dense dans $L^1(\mathbb{R}, dx)$.

Démonstration. Soit $f \in L^1(\mathbb{R}, dx)$. En appliquant Fubini (car f et K_T sont intégrables) on obtient que $\mathcal{I}(K_T f) = (\mathcal{I}f)(\mathcal{I}K_T)$. Soit $\varphi(t) = \left(1 - |t|/T\right)^+$, par définition on a $\tilde{\mathcal{I}}\varphi = K_T$ qui est intégrable donc $\varphi = \mathcal{I}K_T$. La fonction $\mathcal{I}K_T$ est par conséquent continue à support compact, il en est donc de même de $\mathcal{I}(K_T f)$. La convergence de $K_T f$ vers f dans $L^1(\mathbb{R}, dx)$ termine la preuve. \square

1.3.2 Transformation de Fourier sur $L^2(\mathbb{R}, dx)$

On pourrait utiliser le fait que $L^1(\mathbb{R}, dx) \cap L^2(\mathbb{R}, dx)$ est dense dans $L^2(\mathbb{R}, dx)$ pour étendre la transformée de Fourier mais nous préférons une méthode plus constructive. On introduit les notations:

Définition 1.54. Pour $f \in L^2(\mathbb{R}, dx)$ et $T > 0$, on pose:

$$\mathcal{F}_T f(t) = \int_{-T}^T f(u)\gamma_{-t}(u) du \quad \tilde{\mathcal{F}}_T f(t) = \int_{-T}^T f(u)\gamma_t(u) du$$

Théorème 1.55. Soit $f \in L^2(\mathbb{R}, dx)$.

1. $\int_A^B |f(x)|^2 dx = \int \left| \int_A^B f(u)\gamma_{-t}(u) du \right|^2 dt = \int \left| \int_A^B f(u)\gamma_t(u) du \right|^2 dt$.
2. La famille $\{\mathcal{F}_T f\}_{T>0}$ (resp $\{\tilde{\mathcal{F}}_T f\}_{T>0}$) est de Cauchy dans $L^2(\mathbb{R}, dx)$ et converge dans cet espace vers une fonction notée $\mathcal{F}f$ (resp $\tilde{\mathcal{F}}f$).
3. Les applications \mathcal{F} et $\tilde{\mathcal{F}}$ sont des applications linéaires isométriques de $L^2(\mathbb{R}, dx)$ et par conséquent on a $\langle f, g \rangle = \langle \mathcal{F}f, \mathcal{F}g \rangle$ pour tout couple de fonctions de $L^2(\mathbb{R}, dx)$.
4. Les applications \mathcal{F} et $\tilde{\mathcal{F}}$ vérifient $\mathcal{F}\tilde{\mathcal{F}} = \tilde{\mathcal{F}}\mathcal{F} = I$ et définissent donc des isomorphismes de $L^2(\mathbb{R}, dx)$.
5. Si $f \in L^1(\mathbb{R}, dx) \cap L^2(\mathbb{R}, dx)$ alors $\mathcal{I}f(t)$ est une version continue de $\mathcal{F}f$.

Démonstration. (1),(2) Soit $[A, B]$ un intervalle et $\tau = \frac{1}{B-A}$. En appliquant 1.32 à la fonction $f(x)\gamma_{-t}(x)$ on obtient:

$$\begin{aligned} \int_A^B |f(x)|^2 dx &= \tau \sum_n \left| \int_A^B f(x)\gamma_{-t}(x)\gamma_{-n\tau}(x) dx \right|^2 \\ \int_A^B |f(x)|^2 dx &= \sum_n \int_0^\tau \left| \int_A^B f(x)\gamma_{-t-n\tau}(x) dx \right|^2 dt \\ &= \sum_n \int_{n\tau}^{(n+1)\tau} \left| \int_A^B f(x)\gamma_{-y}(x) dx \right|^2 dy \\ &= \int \left| \int_A^B f(x)\gamma_{-y}(x) dx \right|^2 dy \end{aligned}$$

Dans la troisième ligne on a fait le changement de variable $t + n\tau = y$. On en déduit la propriété de Cauchy puisque:

$$\mathcal{F}_{T+h}f(t) - \mathcal{F}_Tf(t) = \int_{-T-h}^{-T} f(u)\gamma_{-t}(u) du + \int_T^{T+h} f(u)\gamma_{-t}(u) du = a(t) + b(t)$$

avec, $\|a\|_2^2 = \int_{-T-h}^{-T} |f(u)|^2 du$, $\|b\|_2^2 = \int_T^{T+h} |f(u)|^2 du$. Le résultat est donc une conséquence de la propriété de Cauchy des intégrales convergentes.

(3) La linéarité de \mathcal{F} est évidente. Soit f_T la restriction de f à l'intervalle $[-T, T]$, nulle à l'extérieur de cet intervalle. Il est facile de constater que f_T converge vers f dans $L^2(\mathbb{R}, dx)$ et l'on a $\|f_T\|_2 = \|\mathcal{F}_Tf\|_2$. La convergence dans $L^2(\mathbb{R}, dx)$ impliquant la convergence des normes on en déduit que \mathcal{F} est une isométrie.

(4) En utilisant la bicontinuité du produit scalaire on écrit:

$$\begin{aligned} \|\tilde{\mathcal{F}}\mathcal{F}f - f\|^2 &= \lim_{T \rightarrow \infty} \|\tilde{\mathcal{F}}_T\mathcal{F}f - f_T\|^2 = \\ &= \lim_{T \rightarrow \infty} \left[\|\tilde{\mathcal{F}}_T\mathcal{F}f\|^2 + \|f_T\|^2 - 2\Re\langle \tilde{\mathcal{F}}_T\mathcal{F}f, f_T \rangle \right] = 2(\|f\|^2 - \lim_{T \rightarrow \infty} \Re\langle \tilde{\mathcal{F}}_T\mathcal{F}f, f_T \rangle) \end{aligned}$$

Or une simple application de Fubini montre que

$$\langle \tilde{\mathcal{F}}_T\mathcal{F}f, f_T \rangle = \int_{-T}^T \mathcal{F}f(u)\overline{\mathcal{F}_Tf}(u) du = \langle (\mathcal{F}f)_T, \mathcal{F}_Tf \rangle$$

Et la bicontinuité du produit scalaire implique:

$$\lim_{T \rightarrow \infty} \langle (\mathcal{F}f)_T, \mathcal{F}_Tf \rangle = \langle \mathcal{F}f, \mathcal{F}f \rangle = \|f\|_2^2$$

et on obtient donc $\tilde{\mathcal{F}}\mathcal{F}f - f = 0$.

(5) Soit f sommable et de carré sommable. On a alors pour tout t , $\mathcal{I}f(t) = \lim_{T \rightarrow \infty} \mathcal{F}_Tf(t)$ par application du théorème de Lebesgue. Or le second membre de cette égalité converge aussi dans $L^2(\mathbb{R}, dx)$ vers $\mathcal{F}f(t)$ et par conséquent $\mathcal{I}f$ est une version (continue) de $\mathcal{F}f$. \square

On obtient alors un corollaire voisin de l'affirmation 3 de la proposition 1.51:

Corollaire 1.56. *Si une fonction $f \in L^2(\mathbb{R}, dx)$ possède une transformée de Fourier dans $L^1(\mathbb{R}, dx)$, alors cette fonction est presque sûrement égale à un élément de \mathbf{C}_0 .*

Pour des raisons de commodité d'écriture, l'on confond très rapidement fonctions et classes de fonctions et la notation \hat{f} est utilisée aussi bien pour $\mathcal{F}f$ que pour $\mathcal{I}f$. C'est ce qui sera souvent fait (avec mauvaise conscience ...) dans la suite de ce polycopié

1.3.3 Exemples

Le calcul de la transformée de Fourier d'une fonction f n'est pas toujours une opération simple. Même si f est sommable, le calcul de l'intégrale

$$\hat{f}(t) = \int_{\mathbb{R}} f(x) e^{2i\pi tx} dx \quad (1.1)$$

n'est que rarement possible par calcul de primitives. Dans ce paragraphe on va voir quelques exemples et quelques techniques de calcul.

Exemple 1.57. Si g est une fonction sommable et $h(x) = g(\alpha x)$, alors les transformées de Fourier de h et de g se déduisent facilement l'une de l'autre par changement de variable:

$$\hat{h}(t) = \int_{\mathbb{R}} g(\alpha x) e^{2i\pi tx} dx = \frac{1}{\alpha} \int_{\mathbb{R}} g(y) e^{2i\pi ty/\alpha} dy = \frac{1}{\alpha} \hat{g}\left(\frac{t}{\alpha}\right)$$

Exemple 1.58. Si $f = 1_{[-\frac{1}{2}, \frac{1}{2}]}$, alors

$$\hat{f}(t) = \frac{\sin \pi t}{\pi t} \text{ pour } t \neq 0, \quad \hat{f}(0) = 1$$

cette fonction en théorie du signal s'appelle *le sinus cardinal* et s'écrit $\text{sinc}(t)$.

Exemple 1.59. Parfois, on peut obtenir \hat{f} par la formule d'inversion. Grâce à la remarque 1.52, si deux fonctions intégrables f et g satisfont à $f = \tilde{\mathcal{I}}g$, alors:

$$g = \mathcal{I}(\tilde{\mathcal{I}}g) = \mathcal{I}f = \hat{f} \quad (1.2)$$

Soit par exemple $f(x) = \frac{1}{\pi(1+x^2)}$. La fonction f est sommable, mais le calcul de l'intégrale (1.1) ne présente pas bien. Par contre, en posant $g(x) = e^{-|x|}$, on a:

$$\begin{aligned} \mathcal{I}g(t) &= \int_{\mathbb{R}} e^{-|x|} e^{2i\pi tx} dx = \int_0^{+\infty} e^{-x} e^{2i\pi tx} dx + \int_{-\infty}^0 e^x e^{2i\pi tx} dx = \\ &= \int_0^{+\infty} e^{-x} (e^{2i\pi tx} + e^{-2i\pi tx}) dx = 2 \int_0^{+\infty} e^{-x} \cos(2\pi tx) dx \end{aligned}$$

Cette dernière intégrale se calcule de façon élémentaire et on trouve

$$\mathcal{I}g(t) = \frac{2}{1 + (2\pi t)^2}$$

Donc, grâce à la formule de l'exemple 1.57, en posant $h(x) = g(2\pi x)$, alors

$$\mathcal{I}h(t) = \frac{1}{\pi(1+t^2)} = f(t)$$

Puisque $\mathcal{I}h(t)$ est une fonction paire, $\tilde{\mathcal{I}}h = \mathcal{I}h$ et la relation (1.2) donne

$$\hat{f}(t) = h(t) = e^{-2\pi|t|}$$

Exemple 1.60. On considère le noyau de Féjer K_T de la définition 1.47. On sait que

$$K_T(t) = T \left(\frac{\sin(\pi T t)}{\pi T t} \right)^2 = T \operatorname{sinc}^2(tT)$$

Mais ici il convient de rappeler que

$$K_T(t) = \int_{-T}^T \left(1 - \frac{|u|}{T}\right) \gamma_t(u) du = \int_{\mathbb{R}} \left(1 - \frac{|u|}{T}\right)^+ \gamma_t(u) du$$

Donc si on note $g(u) = \left(1 - \frac{|u|}{T}\right)^+$, on a $K_T = \tilde{\mathcal{I}}g$. L'argument de l'exemple précédent donne alors

$$\widehat{K}_T(u) = \left(1 - \frac{|u|}{T}\right)^+$$

Exemple 1.61. Soit $f(x) = e^{-x^2/2}$. Alors

$$\widehat{f}(t) = \int_{\mathbb{R}} e^{-x^2/2} e^{2i\pi x t} dx$$

La fonction $x \rightarrow x f(x) = x e^{-x^2/2}$, est sommable sur \mathbb{R} . Donc, grâce à la proposition 1.45, la fonction $g = \widehat{f}$ est dérivable et

$$g'(t) = 2\pi i \int_{\mathbb{R}} x e^{-x^2/2} e^{2i\pi x t} dx$$

et si on intègre par parties,

$$g'(t) = (2\pi i)^2 t \int_{\mathbb{R}} e^{-x^2/2} e^{2i\pi x t} dx = -4\pi^2 t g(t)$$

On a donc montré que la fonction $g = \widehat{f}$ est solution d'une équation différentielle ordinaire d'ordre 1 et à coefficients constants. Elle se résout facilement, ce qui donne que \widehat{f} est de la forme

$$\widehat{f}(t) = c e^{-2\pi^2 t^2}$$

Pour déterminer la constante on pose $t = 0$. Puisqu'il est connu que

$$\int_{\mathbb{R}} e^{-x^2/2} dx = \sqrt{2\pi}$$

on a $c = \widehat{f}(0) = (2\pi)^{-1/2}$ finalement

$$\widehat{f}(t) = \sqrt{2\pi} e^{-2\pi^2 t^2}$$

Cette méthode est parfois utilisable. On essaie de dériver \widehat{f} avec la formule de la proposition 1.45 et on espère que \widehat{f} est solution d'une équation différentielle que l'on sait résoudre.

1.4 Convolution

Soit f et g deux fonctions mesurables définies sur \mathbb{R} (resp sur \mathbb{T}, \mathbb{Z}). On définit formellement leur convolution $f * g$ par:

$$f * g(x) = \int_{\mathbb{R}} f(t)g(x-t) dt, \quad \text{resp} \quad \int_{\mathbb{T}} f(t)g(x-t) dt, \quad f * g(n) = \sum_{k \in \mathbb{Z}} f(k)g(n-k)$$

Afin de ne pas alourdir l'exposé on traite simultanément les trois cas (en tenant compte du fait que sur \mathbb{Z} presque partout signifie partout et que la continuité des fonctions est sans intérêt. . .).

Théorème 1.62. *Soit f une fonction sommable.*

1. *Si g est sommable, la fonction $f * g(x)$ est définie pour presque tout x , est elle-même sommable et $\|f * g\|_1 \leq \|f\|_1 \|g\|_1$. De plus la convolution des fonctions sommables est associative et commutative.*
2. *Si g est de carré sommable, la fonction $f * g(x)$ est définie pour presque tout x , est elle-même de carré sommable et $\|f * g\|_2 \leq \|f\|_1 \|g\|_2$.*
3. *Si g est essentiellement bornée, la fonction $f * g(x)$ est définie pour tout x , est uniformément continue et $\sup_x |(f * g)(x)| \leq \|f\|_1 \|g\|_\infty$.*

Démonstration. Nous nous contenterons de donner la preuve dans le cas le plus difficile c'est à dire celui de \mathbb{R} . Dans chaque cas il faut prouver que la définition formelle ci dessus a bien un sens, c'est à dire que l'intégrale $h(x) = \int_{\mathbb{R}} |f(t)g(x-t)| dt$ est finie pour presque tout (ou tout) x et ensuite on prouve les majorations demandées.

(1) En utilisant Fubini on a $\int h(x) dx \leq \|f\|_1 \|g\|_1 < \infty$. La fonction $h(x)$ étant intégrable est donc finie presque partout et la majoration $|\int_{\mathbb{R}} f(t)g(x-t) dt| \leq h(x)$ permet d'obtenir l'inégalité demandée.

(2) On peut écrire:

$$\begin{aligned} \int h^2(x) dx &\leq \int \left[\left(\int |f(t)| dt \right) \left(\int |f(t)g(x-t)|^2 dt \right) \right] dx = \\ &= \|f\|_1 \int \left(\int |f(t)g(x-t)|^2 dt \right) dx = \|f\|_1^2 \|g\|_2^2 \end{aligned}$$

Dans la première ligne on a appliqué l'inégalité de Schwarz à la mesure dt et au produit des fonctions $\sqrt{|f(t)|}$ par $\sqrt{|f(t)g(x-t)|}$. La dernière ligne résulte du théorème de Fubini (cas positif). La fonction h étant de carré intégrable est donc finie presque partout. Ensuite il suffit de remarquer que $|\int f(t)g(x-t) dt|^2 \leq h^2(x)$ ce qui termine la preuve de (2).

(3) La seule chose non évidente est l'uniforme continuité. Or on a:

$$|f * g(x) - f * g(x+u)| \leq \|g\|_\infty \int |f(t-x-u) - f(t-x)| dt = \|g\|_\infty \|\mathcal{T}_u f - f\|_1$$

et on conclut en utilisant 1.50 □

Théorème 1.63. *Si f et g sont de carré sommable, la fonction $(f * g)(x)$ est définie pour tout x , est dans \mathbf{C}_0 et $\sup_x |(f * g)(x)| \leq \|f\|_2 \|g\|_2$.*

Démonstration. Le produit de deux fonctions de carré sommable est sommable donc $f * g(x)$ existe partout et l'inégalité demandée résulte directement de l'inégalité de Schwarz. Si f est l'indicatrice d'un intervalle $[a, b]$ et g l'indicatrice de $[c, d]$ alors $f * g(x) = ((x - c) \wedge b - (x - d) \vee a)^+$ qui est bien une fonction de \mathbf{C}_0 . Il suffit maintenant de constater que si f_n et g_n sont deux suites de $L^2(\mathbb{R}, dx)$ qui convergent dans cet espace vers f et g alors $f_n * g_n$ converge en norme uniforme vers $f * g$ du fait de l'inégalité déjà obtenue. On termine la preuve en utilisant le fait que l'espace \mathbf{C}_0 est fermé pour la norme uniforme. \square

La convolution a des propriétés remarquables vis à vis de la transformation de Fourier, résumées dans le théorème ci-dessous:

Théorème 1.64. *On a les propriétés suivantes:*

1. Soit $a \in \ell^1(\mathbb{Z})$ et $b \in \ell^2(\mathbb{Z})$ alors $\widehat{a * b} = \hat{a}\hat{b}$
2. Soit f et g des fonctions de $L^1(\mathbb{T}, dx)$ alors $\widehat{f * g}(n) = \hat{f}(n)\hat{g}(n)$
3. Soit $f \in L^1(\mathbb{R}, dx)$ et $g \in L^2(\mathbb{R}, dx)$ (ou $g \in L^1(\mathbb{R}, dx)$) alors $\widehat{f * g} = \hat{f}\hat{g}$

Démonstration. (1) Si a et b sont dans $\ell^1(\mathbb{Z})$ le résultat est une simple application de Fubini. Ensuite soit $g^{(n)}$ une suite d'éléments de $\ell^2(\mathbb{Z})$ convergeant vers g et vérifiant la propriété. Alors la majoration $\|f * g\|_2 \leq \|f\|_1 \|g\|_2$ montre que $f * g^{(n)}$ converge dans $\ell^2(\mathbb{Z})$ vers $f * g$. D'après la continuité de la transformée de Fourier on aura $\widehat{f * g^{(n)}} = \hat{f}\hat{g}^{(n)}$ qui converge (dans $L^2(\mathbb{T}, dx)$) à la fois vers $\widehat{f * g}$ et $\hat{f}\hat{g}$ d'où le résultat en choisissant $g^{(n)}$ dans $\ell^1(\mathbb{Z})$ (par exemple la restriction de g aux indices de $[-n, n]$).

(2) Simple application de Fubini.

(3) Identique à (1). \square

Exemple 1.65. On pose $f = 1_{[-\frac{1}{2}, \frac{1}{2}]}$ et on veut calculer $g = f * f$. Ce calcul peut se faire directement, mais il est plus simple de remarquer que $\hat{f}(t) = \text{sinc } t$. Donc $\hat{g}(t) = \text{sinc}^2 t = K_1(t)$ (K_1 est le noyau de Féjer). D'après l'exemple 1.60 on a $\widehat{K_1}(t) = (1 - \frac{|t|}{1})^+$. Donc

$$f * f(t) = (1 - |t|)^+$$

1.5 Autocorrélation

Définition 1.66. *Soit f une fonction de carré sommable. On définit sa fonction d'autocorrélation C_f par la formule $C_f(x) = (f * \check{f})(x)$ où $\check{f}(x) = \check{f}(-x)$.*

Proposition 1.67. *Soit C_f la fonction d'autocorrélation d'une fonction de carré sommable.*

1. C_f est une fonction de \mathbf{C}_0 et $C_f(0) = \|f\|_2^2$.
2. C_f est une fonction de type positif (sur \mathbb{R} , \mathbb{T} , ou \mathbb{Z}).
3. C_f est l'intégrale de Fourier d'une "densité spectrale" sommable et positive:

(a) Dans le cas de \mathbb{R} on a $C_f(t) = \int_{\mathbb{R}} |\hat{f}(u)|^2 \gamma_t(u) du$.

(b) Dans le cas de \mathbb{T} on a $C_f(t) = \sum_{\mathbb{Z}} |\hat{f}(n)|^2 \gamma_t(n)$.

(c) Dans le cas de \mathbb{Z} on a $C_f(n) = \int_{\mathbb{T}} |\hat{f}(u)|^2 \gamma_n(u) du$.

Démonstration. (1) Conséquence de 1.63

(2) On remarque que $C_f(x) = \langle f, \mathcal{T}_x f \rangle$ et que $\langle f, \mathcal{T}_x f \rangle = \langle \mathcal{T}_{-x} f, f \rangle$. Soit (z_1, \dots, z_n) des nombres complexes et (x_1, \dots, x_n) des points de \mathbb{R}, \mathbb{T} , ou \mathbb{Z} . Alors

$$\begin{aligned} \sum_{r=1}^n \sum_{l=1}^n z_r \bar{z}_l C_f(x_r - x_l) &= \sum_{r=1}^n \sum_{l=1}^n z_r \bar{z}_l \langle f, \mathcal{T}_{x_r - x_l} f \rangle = \\ &= \sum_{r=1}^n \sum_{l=1}^n z_r \bar{z}_l \langle \mathcal{T}_{x_l} f, \mathcal{T}_{x_r} f \rangle = \left\| \sum_{l=1}^n \bar{z}_l \mathcal{T}_{x_l} f \right\|^2 \geq 0 \end{aligned}$$

(3) On se contente de prouver (a). Soit $f \in L^2(\mathbb{R}, dx)$ alors on a $\mathcal{F}(\mathcal{T}_x f) = \gamma_{-x} \mathcal{F} f$ (on le vérifie pour \mathcal{F}_T et on passe à la limite). On a alors

$$C_f(x) = \langle f, \mathcal{T}_x f \rangle = \langle \mathcal{F} f, \mathcal{F}(\mathcal{T}_x f) \rangle = \langle \gamma_x, |\mathcal{F} f|^2 \rangle$$

Les autres égalités s'obtiennent de la même façon. \square

La formule définissant la fonction d'autocorrélation est aussi valide pour une fonction sommable, et C_f est alors une fonction sommable, d'où le résultat suivant:

Proposition 1.68. *Si f est sommable et g de carré sommable, alors on peut calculer la fonction d'autocorrélation de $f * g$ qui est de carré sommable et l'on a $C_{f*g} = C_f * C_g$*

Démonstration. Supposons d'abord que g est de plus sommable. On a alors

$$C_f * C_g = (f * \check{f}) * (g * \check{g}) = (f * g) * (\check{f} * \check{g}) = (f * g) * f \check{g} = C_{f*g}$$

Si u_n est une suite de fonctions de L^2 convergeant dans L^2 vers u alors pour tout x on a $C_{u_n}(x) \rightarrow C_u(x)$ car:

$$C_{u_n}(x) = \langle u_n, \mathcal{T}_x u_n \rangle \text{ converge vers } \langle u, \mathcal{T}_x u \rangle = C_u(x)$$

Soit maintenant une suite de fonctions g_n sommables et de carré sommable convergeant dans L^2 vers g . Alors puisque $f * g_n$ est une suite de L^2 convergeant dans L^2 vers $f * g$ on a $C_{f*g_n}(x) = (C_f * C_{g_n})(x)$ qui converge vers C_{f*g} et aussi vers $(C_f * C_g)(x)$ par application du théorème de Lebesgue puisque $|C_{g_n}(t)| \leq \|g_n\|^2 \leq K < \infty$. \square

On peut bien sûr appliquer la proposition 1.67 à C_{f*g} et obtenir de nouvelles formules de représentation spectrale au vu du théorème 1.64 et l'on obtient:

Corollaire 1.69. *Soit f de carré sommable et h sommable. On a alors:*

1. Dans le cas de \mathbb{R} on a $C_{f*h}(t) = \int_{\mathbb{R}} |\hat{h}(u)|^2 |\hat{f}(u)|^2 \gamma_t(u) du$.

2. Dans le cas de \mathbb{T} on a $C_{f*h}(t) = \sum_{\mathbb{Z}} |\hat{h}(n)|^2 |\hat{f}(n)|^2 \gamma_t(n)$.

3. Dans le cas de \mathbb{Z} on a $C_{f*h}(n) = \int_{\mathbb{T}} |\hat{h}(u)|^2 |\hat{f}(u)|^2 \gamma_n(u) du$.

1.6 Transformée en z

Pour l'inversion de l'opération de convolution sur \mathbb{Z} , on introduit classiquement la notion de "transformée en z ".

Définition 1.70. On dira qu'une suite $\{a_n\}_{n \in \mathbb{Z}}$ admet une transformée en z soit $A(z)$ si la série de Laurent $A(z) = \sum_{n \in \mathbb{Z}} a_n z^n$ converge dans un voisinage du cercle unité. (La fonction $A(z)$ est alors holomorphe dans l'intérieur de ce voisinage).

On remarquera que si $a \in \ell^1(\mathbb{Z})$ alors cette série converge pour $|z| = 1$ mais qu'il n'est pas toujours vrai qu'elle converge dans un voisinage du cercle unité. La proposition suivante est la traduction des propriétés des fonctions holomorphes dans un anneau et de leur série de Laurent.

Proposition 1.71. Soit $H(z)$ une fonction holomorphe dans un voisinage du cercle unité contenant l'anneau $r_1 < |z| < r_2$ avec $r_1 < 1$ et $r_2 > 1$. Alors elle admet un développement de Laurent dans cet anneau qui est de plus normalement convergent dans tout sous anneau fermé.

Corollaire 1.72. Soit a une suite admettant une transformée en z soit $A(z)$:

1. La suite $a \in \ell^1(\mathbb{Z})$.
2. $A(\gamma_{-1})(t) = \hat{a}(t)$
3. $a_n = 0$ pour tout $n < 0$ (i.e. a est causal) si et seulement si $A(z)$ est holomorphe dans un voisinage du disque unité.
4. Si b est une suite admettant une transformée en z soit $B(z)$ alors $a * b$ admet une transformée en z donnée par le produit $A(z)B(z)$.

Proposition 1.73. Soit a une suite admettant une transformée en z soit $A(z)$ qui ne s'annule pas dans un voisinage du cercle unité. Alors a admet un inverse dans l'algèbre de convolution $\ell^1(\mathbb{Z})$ qui a pour transformée en z la fonction $1/A(z)$

Démonstration. Puisque $A(z)$ ne s'annule pas alors $1/A(z)$ est holomorphe dans ce voisinage et admet donc un développement de Laurent $\sum b_n z^n$ avec $b \in \ell^1(\mathbb{Z})$. En prenant $z = \gamma_{-1}$ on obtient donc $\hat{a}(t)\hat{b}(t) = 1$ pour tout $t \in \mathbb{T}$ et donc $a * b = \delta_0$. \square

La proposition suivante permet l'élimination du "bruit convolutif" dans certains cas.

Proposition 1.74. Soit b un "bruit" dans $\ell^1(\mathbb{Z})$, admettant une transformée en z soit $B(z)$ qui ne s'annule pas dans un voisinage du cercle unité et s un "signal" dans $\ell^2(\mathbb{Z})$ (resp. dans $\ell^1(\mathbb{Z})$). Alors l'équation $b * a = s$ a une unique solution dans $\ell^2(\mathbb{Z})$ (resp. dans $\ell^1(\mathbb{Z})$) soit $a = h * s$ où h est le filtre $\in \ell^1(\mathbb{Z})$, donné par les coefficients du développement de Laurent de $1/B(z)$. Si $B(z)$ ne s'annule pas sur un voisinage du disque unité alors $h(n) = 0$ pour $n < 0$ (h est causal).

Démonstration. Le filtre b admet un inverse h et on a nécessairement $h * b * a = a = h * s$. Il faut remarquer que si l'on n'impose pas de condition de sommabilité sur a la solution n'est plus du tout unique comme le montre l'exemple des équations de récurrence linéaires... \square

On applique souvent cette proposition à un filtre défini par un polynôme ou, ce qui revient au même, un filtre causal à support fini:

Définition 1.75. Si P est un polynôme de la forme $P(z) = \sum_{k=0}^p a_k z^k$ on lui associe le filtre \mathbf{p} défini par $\mathbf{p}(n) = a_n$ pour $0 \leq n \leq p$ et $\mathbf{p}(n) = 0$ sinon.

On vérifie immédiatement que la transformée en z du filtre \mathbf{p} est le polynôme $P(z)$ et que $\hat{\mathbf{p}}(t) = P(\gamma_{-1}(t))$.

Définition 1.76. (Filtres ARMA) On dira qu'un filtre h est ARMA s'il admet une transformée en z soit $H(z)$ de la forme $P(z)/Q(z)$ où P et Q sont des polynômes, Q n'ayant pas de racines de module 1.

Théorème 1.77. On a les propriétés suivantes:

1. Soit h un filtre ARMA. Alors pour des suites x et y dans $\ell^2(\mathbb{Z})$ (resp. dans $\ell^1(\mathbb{Z})$) la relation $y = h * x$ implique $q * y = p * x$.
2. Réciproquement si P et Q sont des polynômes, Q n'ayant pas de racine de module 1 alors pour un signal x donné, d'énergie finie (resp. stable), l'équation $\mathbf{q} * y = \mathbf{p} * x$ a une unique solution y dans $\ell^2(\mathbb{Z})$ (resp. dans $\ell^1(\mathbb{Z})$) donnée par $y = h * x$ où h est le filtre ARMA associé à la fraction $P(z)/Q(z)$.

Démonstration. Preuve évidente. . . □

L'importance des filtres ARMA provient de l'analyse de nombreux cas de modèles régressifs de la forme $\mathbf{q} * a = \mathbf{p} * s$ où a est le signal pur et s le signal observé.

Remarque 1.78. La formulation du théorème précédent ne peut guère être améliorée:

- Si Q a une racine de module 1, l'équation en y , $\mathbf{q} * y = x$ n'a en général pas de solution dans $\ell^2(\mathbb{Z})$ (resp. dans $\ell^1(\mathbb{Z})$).
- L'équation de récurrence $\mathbf{q} * y = x$ a en général une infinité de solutions, mais au plus une tend vers zéro à l'infini. . .

1.7 Annexe: Signaux déterministes

Il est maintenant temps de donner un petit lexique pour la traduction des résultats mathématiques précédents à la théorie du signal. . . Un *signal* ou un *filtre* sont des fonctions (a priori complexes), qui, selon les cas, sont définies sur \mathbb{R}, \mathbb{Z} ou \mathbb{T} . Ces deux quantités, équivalentes du point de vue du traitement mathématique, recouvrent en fait des notions physiques assez différentes: un signal est en principe une grandeur observable sur un intervalle de temps alors qu'un filtre désignera le plus souvent une fonction, soit inconnue de l'observateur (et sera alors considérée comme parasite) soit construite par l'observateur afin de "moduler" le signal observé.

Définition 1.79. On adoptera les dénominations suivantes:

- Un signal d'énergie finie est une fonction de carré sommable. (L'énergie est alors le carré de la norme $\| \cdot \|_2$ de cette fonction.)
- Un signal ou un filtre stable est une fonction sommable.
- Un signal analogique (resp. digital) est un signal à temps continu (resp. à temps discret).

- La transformée de Fourier d'un filtre stable est le gain de ce filtre.
- On dit qu'un filtre est causal s'il est nul sur \mathbb{R}^- ou sur les entiers négatifs (cas de \mathbb{Z}).
- La densité spectrale d'un signal d'énergie finie est le carré du module de sa transformée de Fourier.

On peut donc énoncer certains résultats précédents (proposition 1.67 et corollaire 1.69) sous la forme:

La fonction d'autocorrélation d'un signal d'énergie finie est la transformée de Fourier inverse de sa densité spectrale.

La densité spectrale du signal filtré d'un signal d'énergie finie f par un filtre stable est le produit de la densité spectrale de f par le carré du module du gain du filtre.

1.7.1 Echantillonnage et formule de Poisson

Définition 1.80. Signaux réguliers.

- On dit qu'un signal $f(t)$ est régulier s'il est continu et s'il existe $\epsilon > 0$ tel que $|t|^{1+\epsilon}f(t)$ soit une fonction bornée. La fonction $|f|^\alpha$ est alors sommable pour tout $\alpha \geq 1$.
- Pour une fonction mesurable f on désigne par f_c l'unique représentant continu de sa classe d'équivalence (lorsqu'il existe...).

Proposition 1.81. Soit f un signal sommable.

1. La série $\sum_{n \in \mathbb{Z}} f(x + 2nT)$ est absolument convergente pour presque tout x et définit une fonction périodique de période $2T$ sommable sur $[-T, +T]$.
2. Si de plus on suppose f régulier alors cette série converge pour tout x et uniformément sur tout compact vers une fonction continue.

Démonstration. (1) En utilisant la Proposition 1.18 il suffit de prouver que pour tout $k \in \mathbb{Z}$ la série $\sum_n \int_{(k-1)T}^{(k+1)T} |f(x + 2nT)| dx$ est convergente ce qui résulte d'un simple changement de variable et de l'intégrabilité de f .

(2) Si $|x| \leq D$, alors pour $|n| \geq D/T$ on a $|x+2nT| \geq |nT|$ et l'on en déduit $|f(x+2nT)| \leq K/|n|^{1+\epsilon}$. La série est donc normalement convergente sur tout compact \square

Définition 1.82. Soit $T > 0$ et $\tau = \frac{1}{2T}$.

- On désigne par H_T l'ensemble des éléments de $L^2(\mathbb{R}, dx)$ possédant une transformée de Fourier nulle hors de l'intervalle $[-T, T]$.
- On pose $\sigma_T(t) = \frac{1}{2T} \int_{-T}^T \gamma_t(u) du = \frac{\sin(2\pi t T)}{2\pi t T}$.

On peut remarquer que toute fonction de H_T possède un représentant continu et même analytique !

Théorème 1.83. On a les propriétés suivantes:

1. H_T est un sous espace fermé de $L^2(\mathbb{R}, dx)$.

2. La famille de fonctions $t \rightarrow \sqrt{2T}\sigma_T(t - n\tau)$ est une base orthonormée de H_T .
3. Pour $f \in H_T$ on a $\langle f, \sigma_T(\cdot - n\tau) \rangle = \tau f_c(n\tau)$. On en déduit donc que la suite $f_c(n\tau)$ est dans $\ell^2(\mathbb{Z})$.
4. Pour $f \in H_T$ on a

$$f_c(t) = \sum_{n \in \mathbb{Z}} f_c(n\tau) \sigma_T(t - n\tau)$$

La série converge pour tout t et dans $L^2(\mathbb{R}, dt)$.

5. Si la suite $f_c(n\tau)$ est dans $\ell^1(\mathbb{Z})$ (en particulier si f est régulière) alors on a convergence uniforme en t dans la série précédente.

Les deux dernières formules ci dessus constituent le **théorème d'échantillonnage** et la fréquence d'échantillonnage $\tau = 1/2T$ est appelée **fréquence de Nyquist**. On en déduit donc que pour un signal dont le spectre est essentiellement contenu dans un intervalle borné $[-A, A]$, il suffit de connaître un échantillonnage discret à une fréquence supérieure à $1/(2A)$ pour reconstituer ce signal à temps continu. *Démonstration.* (1) Soit F l'espace $L^2([-T, T], dx)$ identifié au sous espace de $L^2(\mathbb{R}, dx)$ des fonctions nulles hors de $[-T, T]$. Alors l'isométrie $\tilde{\mathcal{F}}$ envoie F sur H_T . Puisque F est fermé dans $L^2(\mathbb{R}, dx)$ il en est de même de H_T .

(2) La famille $\sqrt{\tau}\gamma_{-n\tau}\mathbf{1}_{[-T, T]}$ est une base orthonormée de F , il en est donc de même de la famille $\tilde{\mathcal{F}}(\sqrt{\tau}\gamma_{-n\tau}\mathbf{1}_{[-T, T]})(t) = \sqrt{2T}\sigma_T(t - n\tau)$ dans H_T .

(3) On a:

$$\langle f, \sigma_T(\cdot - n\tau) \rangle = \langle \mathcal{F}f, \tau\gamma_{-n\tau}\mathbf{1}_{[-T, T]} \rangle = \tau \int \mathcal{F}f(t)\gamma_{n\tau}(t) dt = \tau \tilde{\mathcal{I}}\mathcal{F}f(t) = \tau f_c(t)$$

Les deux dernières égalités étant une conséquence du fait que la fonction $\mathcal{F}f$ est en fait dans $L^1(\mathbb{R}, dx)$.

(4) D'après 1.32(conjuguée) on a pour tout t le développement convergent dans F :

$$\gamma_{-t}\mathbf{1}_{[-T, T]} = \sum_n \sigma_T(t - n\tau)\gamma_{-n\tau}\mathbf{1}_{[-T, T]}$$

Et donc, en utilisant le produit scalaire dans F , on a pour tout t :

$$\begin{aligned} f_c(t) &= \int \mathcal{F}f\gamma_t = \langle \mathcal{F}f, \gamma_{-t}\mathbf{1}_{[-T, T]} \rangle = \sum_n \sigma_T(t - n\tau) \langle \mathcal{F}f, \gamma_{-n\tau}\mathbf{1}_{[-T, T]} \rangle = \\ &= \sum_{n \in \mathbb{Z}} \sigma_T(t - n\tau) f_c(n\tau) \end{aligned}$$

Cette série converge dans $L^2(\mathbb{R}, dx)$ d'après 1.19 puisque la suite $f_c(n\tau)$ est dans $\ell^2(\mathbb{Z})$ et que la famille $\sigma_T(\cdot - n\tau)$ est orthogonale, de norme constante égale à τ

(5) Si $f_c(n\tau)$ est sommable la série est normalement convergente sur \mathbb{R} du fait de la majoration $\sup_t |\sigma_T(t - n\tau)| \leq 1$. \square

Théorème 1.84. Soit f un signal sommable ou de carré sommable, possédant une transformée de Fourier sommable (alors f a un représentant continu f_c).

1. La suite $\tau f_c(-n\tau)$ est la suite des coefficients de Fourier de la fonction sommable et périodique sur $[-T, T]$ définie par $\varphi(t) = \sum_{n \in \mathbb{Z}} \hat{f}(t + 2nT)$

2. Si φ est de carré sommable ou bien si la suite $f_c(n\tau)$ est dans $\ell^2(\mathbb{Z})$ alors on a :

$$\sum_{n \in \mathbb{Z}} \hat{f}(t + 2nT) = \tau \sum_{n \in \mathbb{Z}} f_c(n\tau) \gamma_{-n\tau}(t)$$

La dernière série converge dans $L^2([-T, T])$ et l'égalité est presque sûre par rapport à la variable t .

Démonstration. (1) Soit m un entier fixé. Puisque \hat{f} est sommable, la série définissant φ converge dans $L^1([-T, T], dx)$ et donc la série $\sum_n \gamma_{m\tau}(t) \hat{f}(t + 2nT)$ est intégrable terme à terme sur $[-T, T]$. On a donc :

$$\begin{aligned} \hat{\varphi}(-m) &= \tau \int_{-T}^T \varphi(t) \gamma_{m\tau}(t) dt = \tau \sum_{n \in \mathbb{Z}} \int_{-T}^T \gamma_{m\tau}(t) \hat{f}(t + 2nT) dt \\ &= \tau \sum_{n \in \mathbb{Z}} \int_{(2n-1)T}^{(2n+1)T} \gamma_{m\tau}(u) \hat{f}(u) du = \tau f_c(m\tau) \end{aligned}$$

(2) Si φ est de carré sommable sur $[-T, T]$ on peut en utilisant 1.32 écrire dans $L^2([-T, T], dx)$:

$$\varphi = \sum_n \hat{\varphi}(n) \gamma_{n\tau} = \tau \sum_n f_c(n\tau) \gamma_{-n\tau}$$

Si $f_c(n\tau)$ est de carré sommable alors d'après 1.19 on peut affirmer que la série du second membre converge dans $L^2([-T, T], dx)$ vers une fonction ψ . Or le calcul des coefficients de Fourier de ψ redonne la suite $f_c(-n\tau)$ et par conséquent $\varphi = \psi$. \square

La dernière formule de ce théorème peut prendre une forme beaucoup plus précise appelée "Formule de Poisson".

Corollaire 1.85. Soit f un signal régulier tel que \hat{f} soit aussi régulier. Alors :

$$\sum_{n \in \mathbb{Z}} \hat{f}(t + 2nT) = \tau \sum_{n \in \mathbb{Z}} f(n\tau) \gamma_{-n\tau}(t)$$

pour tout t . Les deux séries sont uniformément convergentes sur tout compact et définissent alors des fonctions continues de période $2T$. En particulier pour $t = 0$ (et en remplaçant $2T$ par T) on obtient la formule de Poisson :

$$\sum_{n \in \mathbb{Z}} \hat{f}(nT) = \frac{1}{T} \sum_{n \in \mathbb{Z}} f\left(\frac{n}{T}\right)$$

Démonstration. Puisque f est régulier alors la suite $f_c(n\tau) = f(n\tau)$ est sommable et donc la série du second membre converge uniformément sur \mathbb{R} . On a déjà vu que si \hat{f} est régulier alors la série du premier membre converge uniformément sur tout compact. Ces deux limites ponctuelles sont presque sûrement égales et elles sont en fait égales en tout point puisque ce sont des fonctions continues (comme limites uniformes de fonctions continues). \square

En remplaçant f par $\mathcal{T}_{-u}f$ on peut obtenir une formule plus complète :

Corollaire 1.86. Soit f un signal régulier tel que \hat{f} soit aussi régulier. Alors:

$$\sum_{n \in \mathbb{Z}} \hat{f}(t + 2nT) \gamma_u(t + 2nT) = \tau \sum_{n \in \mathbb{Z}} f(n\tau + u) \gamma_{-n\tau}(t)$$

pour tout t et tout u .

On peut se demander s'il n'existe pas d'autres procédés d'interpolation que celui décrit par le théorème d'échantillonnage. Soit donc f un signal tel que f et \hat{f} soient réguliers, φ une fonction continue, bornée de carré sommable. Alors on peut considérer la série

$$g(t) = \sum_{n \in \mathbb{Z}} f(n\tau) \varphi(t - n\tau)$$

qui converge dans $L^2(\mathbb{R}, dx)$ et uniformément en t vers la fonction continue g . D'après la formule de Poisson on a donc:

$$\hat{g}(t) = 2T \hat{\varphi}(t) \sum_{n \in \mathbb{Z}} \hat{f}(t + 2nT)$$

Si l'on veut pouvoir comparer g et f au sens de la norme L^2 (c'est à dire en fait \hat{f} et \hat{g}) il faut éviter le phénomène d'enroulement du spectre c'est à dire que l'on impose à \hat{f} de s'annuler hors de $[-F, F]$ avec $T \geq F$. La fonction périodique $\Psi(t) = \sum_{n \in \mathbb{Z}} \hat{f}(t + 2nT)$ vaut alors $\hat{f}(t)$ sur $[-T, +T]$ et pour que f et g soient voisines (dans $L^2(\mathbb{R}, dx)$), il faut que $2T \hat{\varphi}(t)$ soit proche de 1 sur $[-F, F]$ et proche de 0 sur le complémentaire de $[-T, T]$. Ceci conduit à une fréquence d'échantillonnage très élevée pour le cas de l'interpolation linéaire associée à la fonction $\varphi(t) = (1 - \frac{|t|}{\tau})^+$. On a en effet, dans ce cas, $2T \hat{\varphi}(t) = (\frac{\sin(\pi\tau t)}{\pi\tau t})^2$ et donc pour F donné, il faut choisir τ très petit pour espérer une approximation convenable.

1.7.2 Transformée de Fourier sur un ensemble fini

Soit $N \geq 1$ un entier fixé. On pose $\omega = \gamma_{-1}(\frac{1}{N})$.

Définition 1.87. Pour une suite finie de nombres complexes $(a_0, a_1, \dots, a_{N-1})$ on définit les suites périodiques (indexées par \mathbb{Z}) et de période N :

$$\mathcal{F}a(k) = \sum_{n=0}^{N-1} a_n \omega^{nk}, \quad \tilde{\mathcal{F}}a(k) = \sum_{n=0}^{N-1} a_n \omega^{-nk}$$

On remarque tout de suite, qu'en notant toujours par a le prolongement périodique à \mathbb{Z} de la suite finie $(a_0, a_1, \dots, a_{N-1})$ alors les sommes ci-dessus peuvent être calculées sur n'importe quel ensemble de N indices consécutifs. Nous identifieront donc une suite finie à son prolongement périodique.

Définition 1.88. Soit a et b deux suites périodiques de période N et J un ensemble de N indices consécutifs.

- Le produit scalaire est défini par $\langle a, b \rangle = \sum_{n \in J} a_n \bar{b}_n$
- La convolution est définie par $a * b(n) = \sum_{k \in J} a_k b_{n-k}$

Proposition 1.89. On a les propriétés suivantes:

1. $\langle \mathcal{F}a, b \rangle = \langle a, \tilde{\mathcal{F}}b \rangle$
2. $\mathcal{F}\tilde{\mathcal{F}} = \tilde{\mathcal{F}}\mathcal{F} = NI_d$
3. $\mathcal{F}(a * b) = \mathcal{F}(a)\mathcal{F}(b)$

La formule de Poisson permet de relier la transformée de Fourier finie à la transformée de Fourier usuelle:

Proposition 1.90. Soit T et τ vérifiant la relation $2NT\tau = 1$, f et \hat{f} réguliers. Si on définit les deux suites u et v , périodiques de période N par:

$$u(k) = \tau \sum_{n \in \mathbb{Z}} f(k\tau + nN\tau), \quad v(k) = \sum_{n \in \mathbb{Z}} \hat{f}(2kT + 2nNT)$$

alors on a les relations:

$$\mathcal{F}u = v, \quad \tilde{\mathcal{F}}v = Nu$$

Démonstration. Par application de la formule de Poisson 1.86 en $t = 0$, pour tout $\alpha > 0$ et tout u :

$$\alpha \sum_n f(u + n\alpha) = \sum_n \hat{f}(n/\alpha) \gamma_u(n/\alpha)$$

Si on fait le changement de variable $u = k\tau$, $\alpha = N\tau$ on a:

$$\begin{aligned} N\tau \sum_n f(k\tau + nN\tau) &= \sum_n \hat{f}(2Tn) \gamma_{k\tau}(2Tn) \\ u(k) &= \frac{1}{N} \sum_n \hat{f}(2Tn) \omega^{-nk} \\ \mathcal{F}u(p) &= \frac{1}{N} \sum_n \hat{f}(2Tn) \sum_{k=0}^{N-1} \omega^{(p-n)k} = \sum_r \hat{f}(2Tp + 2rnT) = v(p) \end{aligned}$$

La dernière formule est obtenue en constatant que la somme $\sum_{k=0}^{N-1} \omega^{(p-n)k}$ est nulle sauf pour n de la forme $n = p + rN$ avec r entier et dans ce cas elle vaut N .

La seconde formule se déduit de celle-ci grâce à la relation $\tilde{\mathcal{F}}\mathcal{F}u = Nu$. \square

Echantillonnage fini

On veut résoudre problème suivant: connaissant N valeurs d'un signal analogique f mesurées à des instants $T + k\Delta$, $k = 0, 1, \dots, N-1$ dans quelle mesure peut-on reconstituer la transformée de Fourier du signal original? L'on doit bien sûr supposer que l'intervalle de temps $[T, T + N\Delta]$ contient l'essentiel de ce signal c'est à dire qu'il est négligeable à l'extérieur de ce laps de temps. Malheureusement on ne peut pas avoir un signal à support compact dont la transformée de Fourier est elle même à support compact ce qui explique la formulation un peu "vague" de la proposition ci-dessous.

Proposition 1.91. Supposons f et \hat{f} réguliers, le signal f négligeable hors d'un intervalle du type $[T, T + N\Delta[$ et \hat{f} négligeable hors d'un intervalle du type $[F, F + 1/\Delta[$, Soit a le prolongement périodique de la suite finie $a_n = f(T + n\Delta)$, $n = 0, \dots, N-1$ et n_0 la partie entière de $F\Delta$. Alors la suite $\Delta\mathcal{F}a(k) \gamma_{-T}(\frac{k}{N\Delta} + \frac{n_0}{\Delta})$ est un échantillonnage "approché" de \hat{f} pour les N valeurs de k telles que les fréquences $(\frac{k}{N\Delta} + \frac{n_0}{\Delta})$ soient dans l'intervalle $[F, F + 1/\Delta]$

Démonstration. En appliquant la formule (1.86) on a, en posant $t = \frac{k}{N\Delta}$,

$$\begin{aligned} \sum_{n \in \mathbb{Z}} f(T + n\Delta) \gamma_{-n\Delta}(t) &= \frac{1}{\Delta} \sum_{n \in \mathbb{Z}} \hat{f}\left(t + \frac{n}{\Delta}\right) \gamma_T\left(t + \frac{n}{\Delta}\right) \\ \sum_{n \in \mathbb{Z}} f(T + n\Delta) \omega^{kn} &= \mathcal{F}a(k) = \frac{1}{\Delta} \sum_{n \in \mathbb{Z}} \hat{f}\left(\frac{k}{N\Delta} + \frac{n}{\Delta}\right) \gamma_T\left(\frac{k}{N\Delta} + \frac{n}{\Delta}\right) \end{aligned}$$

L'hypothèse faite sur le support de \hat{f} implique que la dernière somme écrite ne contient en fait qu'un seul terme d'où la conclusion. \square

Dans le cas d'intervalles symétriques on peut donner une formulation plus simple:

Proposition 1.92. *Supposons f et \hat{f} réguliers, $N = 2M$, le signal f négligeable hors de l'intervalle $[-M\Delta, M\Delta[$ et \hat{f} négligeable hors de l'intervalle $[-\frac{1}{2\Delta}, \frac{1}{2\Delta}[$. Soit a le prolongement périodique de la suite $a_n = f(n\Delta)$, $n = -M, \dots, M-1$. Alors la suite $\Delta \mathcal{F}a(k)$ pour $k = -M, \dots, M-1$ est un échantillonnage "approché" de \hat{f} aux fréquences $\frac{k}{N\Delta}$.*

Démonstration. On part de la formule: $\sum_n f(n\Delta) \gamma_{-n\Delta}(t) = \frac{1}{\Delta} \sum_n \hat{f}(t + n/\Delta)$ puis on pose $t = k/(N\Delta)$ et la preuve est alors identique à la précédente. \square

On en conclut que si f est négligeable hors d'un intervalle $[-A, A]$ et \hat{f} est négligeable hors d'un intervalle $[-B, B]$ alors le nombre d'observations de f doit être supérieur ou égal à $4AB$ pour avoir un échantillonnage correct de \hat{f} par utilisation de la transformée de Fourier finie.

Transformée de Fourier rapide

Dans le cas $N = 2M$ on peut décomposer le calcul de la transformée de Fourier sur N indices en deux calculs de transformée de Fourier à M indices ce qui fait passer d'un calcul de complexité N^2 à un calcul de complexité $N \ln(N)$ lorsque N est une puissance de 2. (En fait on peut appliquer ce procédé au cas $N = PQ$).

Proposition 1.93. *On définit les transformées de Fourier à M indices des sous-suites paires et impaires par:*

$$\mathcal{F}_p(k) = \sum_{n=0}^{M-1} a_{2n} \omega^{2nk}, \quad \mathcal{F}_i(k) = \sum_{n=0}^{M-1} a_{2n+1} \omega^{2nk}$$

Alors pour $k = 0, 1 \dots M-1$ on a les relations:

$$\mathcal{F}(k) = \mathcal{F}_p(k) + \omega^k \mathcal{F}_i(k), \quad \mathcal{F}(M+k) = \mathcal{F}_p(k) - \omega^k \mathcal{F}_i(k)$$

Il suffit alors d'écrire une procédure récursive pour implémenter l'algorithme.

1.7.3 Transformation de Laplace

Un outil très utilisé pour le calcul symbolique des filtres est la transformée de Laplace qui joue le rôle de la transformée en z dans le cas continu. On définit ci dessous la transformée de Laplace complexe, mais elle est le plus souvent utilisée dans le cas d'une variable z réelle.

Définition 1.94. Soit f une fonction mesurable. On définit alors sa transformée de Laplace

$$\mathcal{L}f(z) = \int e^{-zt} f(t) dt$$

du moins lorsque l'intégrale est absolument convergente c'est à dire pour les z tels que:

$$\Re(z) \in \Omega_f = \{x \in \mathbb{R} \mid \int e^{-xt} |f(t)| dt < \infty\}$$

Le seul résultat évoqué dans ce polycopié concerne le fait que la transformée de Laplace d'une fonction sommable à support compact est analytique. C'est pourquoi nous nous limiterons à la preuve de (2) dans le théorème suivant.

Théorème 1.95. Soit f une fonction mesurable.

1. Ω_f est un intervalle, $0 \in \Omega_f$ si et seulement si f est sommable.
2. Si f est sommable à support compact alors $\Omega_f = \mathbb{R}$ et $\mathcal{L}f$ est holomorphe dans le plan complexe.
3. La fonction $\mathcal{L}f$ est holomorphe dans le domaine $\{z \mid \Re(z) \in \overset{\circ}{\Omega}_f\}$. Dans ce domaine, ses dérivées successives sont données par les intégrales absolument convergentes:

$$\frac{d^n \mathcal{L}f(z)}{d^n(z)} = (-1)^n \int t^n \exp(-zt) f(t) dt$$

4. Soit f et g deux fonctions mesurables telles que $\Omega_f \cap \Omega_g$ contienne un ouvert sur lequel $\mathcal{L}f$ et $\mathcal{L}g$ coïncident. Alors on a $f = g$ presque partout.
5. Soit f et g deux fonctions mesurables telles que l'intégrale définissant $f * g(x)$ soit absolument convergente pour presque tout x . Alors $\Omega_f \cap \Omega_g \subset \Omega_{f*g}$ et sur cet ensemble on a $\mathcal{L}(f * g) = \mathcal{L}f \mathcal{L}g$.

Démonstration. Preuve de (2): le fait que $\Omega_f = \mathbb{R}$ est évident et l'on va montrer que $\mathcal{L}f$ est holomorphe dans le plan complexe en prouvant qu'elle est dérivable en tout point z_0 . D'après le théorème de dérivation sous le signe somme, il suffit de montrer que $|x \exp(-xz) f(x)|$ est majoré par une fonction intégrable fixe sur tout disque $|z| \leq R$. Le support de f étant contenu dans l'intervalle $|x| \leq K$ on a $|x \exp(-xz) f(x)| \leq K \exp(KR) |f(x)|$ □

Remarque 1.96. Cette proposition a des conséquences importantes:

- Une fonction sommable à support compact ne peut pas avoir une transformée de Fourier à support compact.
- Si f est sommable alors $\hat{f}(t) = \mathcal{L}f(-2i\pi t)$
- Si le point 0 est dans l'intérieur de Ω_f alors la transformée de Fourier de f est analytique et les fonctions $x^n f(x)$ sont sommables pour tout entier positif n .

Il n'existe malheureusement pas de critère simple (comme pour la transformée en z) permettant d'affirmer qu'une fonction holomorphe dans un "tube" est la transformée de Laplace d'une fonction sommable...

On utilise beaucoup la transformée de Laplace pour les fonctions à support dans \mathbb{R}^+ car cet ensemble est stable par convolution et pour une fonction f sommable on a $[0, \infty[\subset \Omega_f$.

Chapitre 2

Rappels de calcul des probabilités

2.1 Généralités

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Une application mesurable de cet espace dans \mathbb{R} (resp. \mathbb{R}^d) muni de sa tribu borélienne \mathcal{B} (resp. \mathcal{B}^d) est appelée variable aléatoire (v.a.) réelle, (resp. vecteur aléatoire, ou variable aléatoire d -dimensionnelle).

On définit l'intégrale $\int_{\Omega} X(\omega) d\mathbb{P}(\omega)$ comme on l'a vu au §1.1.2. Dans ce cas on utilise aussi la notation $\mathbb{E}(X)$ pour indiquer l'intégrale.

L'intégrale par rapport à une mesure de probabilité jouit de quelques propriétés supplémentaires. D'abord, en utilisant l'inégalité de Hölder et le fait que la fonction constante égale à 1 est dans tous les espaces L^p on obtient les inégalités $\|f\|_p \leq \|f\|_q$ si $q \geq p$, d'où les inclusions

$$L^{\infty}(\Omega, \mathbb{P}) \subset \dots \subset L^2(\Omega, \mathbb{P}) \subset L^1(\Omega, \mathbb{P}) .$$

Les théorèmes rappelés dans la première partie du cours (en particulier Lebesgue et Fubini) sont bien sûr toujours valables. Soit X un vecteur aléatoire. On définit la loi de X comme la probabilité \mathbb{P}_X sur \mathbb{R}^d muni de sa tribu borélienne \mathcal{B}^d , définie par $\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B))$ pour tout borélien B (on désigne souvent cette quantité par $\mathbb{P}(X \in B)$). On a alors pour toute fonction mesurable positive φ sur \mathbb{R}^d :

$$\mathbb{E}[\varphi(X)] = \int_{\mathbb{R}^d} \varphi(x) d\mathbb{P}_X(x)$$

Cette valeur est finie ou non. La fonction φ est donc \mathbb{P}_X intégrable si et seulement si $\varphi(X) \in L^1(\Omega, \mathbb{P})$. En particulier chaque composante de X est dans L^p si et seulement si $\|x\|^p$ est \mathbb{P}_X intégrable. Cette formule, qui ramène le calcul de $\mathbb{E}[\varphi(X)]$ à celui d'une intégrale sur \mathbb{R}^d , est d'un usage très fréquent.

Dans la suite, pour simplifier les notations, on écrira, pour une probabilité μ sur \mathbb{R}^d , $\mu(f)$ au lieu de $\int f d\mu$. On écrira donc aussi $\mathbb{P}_X(f)$ au lieu de $\int f \mathbb{P}_X$.

On dit que la loi \mathbb{P}_X admet une densité s'il existe une fonction positive $f \in L^1(\mathbb{R}^d, dx)$ telle que

- $\int_{\mathbb{R}^d} f(x) dx = 1$
- $\mathbb{P}(X_1 \in B_1, \dots, X_d \in B_d) = \int_{B_1 \times \dots \times B_d} f(x) dx$ pour toute famille (B_1, \dots, B_d) d'intervalles de \mathbb{R} . On en déduit que $\mathbb{E}(\varphi(X)) = \int_{\mathbb{R}^d} \varphi(x) d\mathbb{P}_X(x) = \int_{\mathbb{R}^d} f(x)\varphi(x) dx$

pour toute fonction φ mesurable positive ou mesurable bornée sur \mathbb{R}^d . En particulier, pour tout borélien B de \mathbb{R}^d , on a $\mathbb{P}(X \in B) = \int_B f(x) dx$.

On remarque qu'une densité n'est définie que presque partout relativement à la mesure de Lebesgue sur \mathbb{R}^d .

On est souvent amenés à considérer le problème suivant. Si $X = (X_1, \dots, X_d)$ est une v.a. de densité f_X et $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^m$, est-ce que $Y = \Phi(X)$ a une densité f_Y ? Comment la calculer? On rappelle ici deux situations intéressantes de ce type.

Si A est une matrice $d \times d$ inversible et $b \in \mathbb{R}^d$, alors $Y = AX + b$ a pour densité

$$f_Y(x) = \frac{1}{|\det A|} f_X(A^{-1}(x - b)) \quad (2.1)$$

De plus, si on désigne par Y une sous famille des coordonnées de X de taille $m \leq d$, alors Y admet une densité sur \mathbb{R}^m obtenue en intégrant la densité de X par rapport aux variables ne figurant pas dans Y . En particulier la v.a. X_1 a densité

$$f_{X_1}(t) = \int f_X(t, x_2, \dots, x_d) dx_2 \dots dx_d$$

2.2 Convergences

Définition 2.1. Soit X_n et X des v.a. (réelles ou vectorielles)

- La suite X_n converge presque sûrement (p.s.) vers X si $\mathbb{P}\{\omega; X_n(\omega) \rightarrow X(\omega)\} = 1$.
- La suite X_n converge en probabilité vers X si $\forall \varepsilon > 0$ on a $\mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0$.
- On dit que X_n converge en loi (ou étroitement) vers X si $\mathbb{E}(f(X_n)) \rightarrow \mathbb{E}(f(X))$ pour toute fonction f , continue et bornée. Il revient au même de dire que $\mathbb{P}_{X_n}(f) \rightarrow \mathbb{P}_X(f)$ pour toute fonction f , continue et bornée.

On vérifie immédiatement que $X_n = (X_{n,1}, \dots, X_{n,d})$ converge vers $X = (X_1, \dots, X_d)$ p.s. ou en probabilité ssi, pour $k = 1, \dots, d$, $X_{n,k}$ converge vers X_k dans le même sens. De plus ces deux types de convergence sont compatibles avec la structure d'espace vectoriel, c'est à dire que si deux suites convergent, il en est de même de toute combinaison linéaire et la limite est la combinaison linéaire des limites. La convergence en loi ne possède aucune de ces deux propriétés!!

Proposition 2.2. Soit X_n et X des v.a.

1. (Borel-Cantelli) Une condition suffisante de convergence presque sûre de X_n vers X est $\sum_n \mathbb{P}(|X_n - X| > \varepsilon) < \infty$, pour tout $\varepsilon > 0$.
2. Si $X_n \rightarrow X$ p.s., alors $X_n \rightarrow X$ en probabilité.
3. Si $X_n \rightarrow X$ en probabilité, alors il existe une sous suite X_{n_k} qui converge presque sûrement vers X .
4. Si $X_n \rightarrow X$ dans L^p , avec $1 \leq p < \infty$, alors $X_n \rightarrow X$ en probabilité.

(1): Il est facile de constater que la convergence presque sûre est équivalente à $\lim_n \downarrow \mathbb{P}(\cup_{k=n}^{\infty} \{|X_k - X| > \varepsilon\}) = 0$ pour tout $\varepsilon > 0$. Or

$$\mathbb{P}(\cup_{k=n}^{\infty} \{|X_k - X| > \varepsilon\}) \leq \sum_{k=n}^{\infty} \mathbb{P}(|X_k - X| > \varepsilon).$$

(2): Évident puisque $\{|X_n - X| > \varepsilon\} \subset \cup_{k=n}^{\infty} \{|X_k - X| > \varepsilon\}$.

(3): On définit la suite n_k par récurrence comme étant le plus petit entier $> n_{k-1}$ vérifiant $\mathbb{P}\{|X_{n_k} - X| > \frac{1}{k}\} \leq 2^{-k}$ et on remarque qu'une autre condition suffisante de convergence presque sûre est $\sum_n \mathbb{P}(|X_n - X| > \varepsilon_n) < \infty$ pour une suite $\varepsilon_n \rightarrow 0$.

(4): C'est une conséquence de l'inégalité de Bienaymé-Chebyshev:

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \frac{1}{\varepsilon^p} \|X - X_n\|_p^p$$

□

On étudie maintenant la convergence en loi.

Théorème 2.3. Soit μ_n et μ des probabilités sur \mathbb{R} telles que $\mu_n(\varphi) \rightarrow \mu(\varphi)$ pour toute fonction continue à support compact φ . Alors μ_n converge étroitement vers μ .

Démonstration. Soit f une fonction positive continue bornée et φ_l une suite de fonctions continues à support compact qui tend en croissant vers la fonction constante 1. Les fonctions $f\varphi_l$ sont continues à support compact et:

$$\lim_n \mu_n(f\varphi_l) = \mu(f\varphi_l), \quad \lim_n \mu_n(\varphi_l) = \mu(\varphi_l), \quad \lim_l \uparrow \mu(\varphi_l) = 1$$

Par conséquent des inégalités

$$|\mu(f) - \mu(f\varphi_l)| \leq \|f\|_{\infty}(1 - \mu(\varphi_l)), \quad |\mu_n(f) - \mu_n(f\varphi_l)| \leq \|f\|_{\infty}(1 - \mu_n(\varphi_l)),$$

on déduit:

$$|\mu_n(f) - \mu(f)| \leq |\mu_n(f\varphi_l) - \mu(f\varphi_l)| + \|f\|_{\infty}(2 - \mu(\varphi_l) - \mu_n(\varphi_l))$$

et on obtient le résultat en faisant tendre d'abord n puis l vers l'infini. □

Il faut bien remarquer que cet énoncé suppose que l'on sait à l'avance que μ est une probabilité (sinon l'énoncé est faux!).

Proposition 2.4. Si $X_n \rightarrow X$ en probabilité, alors $X_n \rightarrow X$ en loi.

Démonstration. Soit φ une fonction continue à support compact (donc uniformément continue) et $\varepsilon > 0$. Il existe $\alpha > 0$ tel que $|x - y| \leq \alpha$ implique $|\varphi(x) - \varphi(y)| \leq \varepsilon/2$. En utilisant la majoration $|\varphi| \leq K$ on obtient:

$$\begin{aligned} |\mathbb{E}(\varphi(X_n)) - \mathbb{E}(\varphi(X))| &\leq \mathbb{E}(|\varphi(X_n)) - \mathbb{E}(\varphi(X))|) = \\ &= \int_{|X - X_n| \leq \alpha} |\varphi(X) - \varphi(X_n)| d\mathbb{P} + \int_{|X - X_n| > \alpha} |\varphi(X) - \varphi(X_n)| d\mathbb{P} \\ &\leq \varepsilon/2 + 2K\mathbb{P}(|X - X_n| > \alpha) \end{aligned}$$

et, comme $X_n \rightarrow X$ en probabilité, $\mathbb{P}(|X - X_n| > \alpha) \rightarrow 0$. □

Proposition 2.5. Si X_n converge en probabilité vers une constante a et Y_n converge en loi vers Y alors $X_n + Y_n$ converge en loi vers $a + Y$.

Démonstration. En reprenant les notations de la preuve précédente on peut écrire:

$$|\mathbb{E}(\varphi(X_n + Y_n)) - \mathbb{E}(\varphi(a + Y))| \leq \mathbb{E}(|\varphi(X_n + Y_n) - \varphi(a + Y_n)|) + |\mathbb{E}(\varphi(a + Y_n)) - \mathbb{E}(\varphi(a + Y))|$$

La quantité $\mathbb{E}(\varphi(a + Y_n)) - \mathbb{E}(\varphi(a + Y))$ tend vers 0 par hypothèse, et pour le premier terme, on opère comme dans la démonstration précédente. \square

2.3 Indépendance

Soit $X = (X_1, X_2, \dots, X_n)$ une famille de v.a. réelles considérée comme une v.a. à valeurs dans \mathbb{R}^n . On peut définir sa loi \mathbb{P}_X par

$$\mathbb{P}_X(B_1 \times B_2 \times \dots \times B_n) = \mathbb{P}(\bigcap_{k=1}^n X_k^{-1}(B_k))$$

pour toute famille (B_1, \dots, B_n) de boréliens de \mathbb{R} .

Définition 2.6. On dit que les v.a. réelles X_n sont indépendantes si la loi \mathbb{P}_X est le produit des lois P_{X_k} , $k = 1 \dots n$, ou de façon équivalente:

$$\mathbb{P}\left(\bigcap_{k=1}^n X_k^{-1}(B_k)\right) = \prod_{k=1}^n \mathbb{P}(X_k^{-1}(B_k))$$

pour toute famille (B_1, \dots, B_n) de boréliens de \mathbb{R} ou encore

$$\mathbb{E}\left[\prod_{k=1}^n \varphi_k(X_k)\right] = \prod_{k=1}^n \mathbb{E}(\varphi_k(X_k))$$

pour toute famille $(\varphi_1, \dots, \varphi_n)$ de fonctions continues à support compact ou mesurables positives.

Proposition 2.7. Soit X_n (resp. Y_n) une suite de v.a. réelles qui converge en probabilité vers X (resp. Y). Si pour tout n les v.a. X_n et Y_n sont indépendantes il en est de même de X et Y

Démonstration. Soit φ et ψ des fonctions continues à support compact. En reprenant la preuve de 2.4 on voit que les suites $\varphi(X_n)$ et $\psi(Y_n)$ convergent dans $L^1(\Omega, \mathbb{P})$ vers $\varphi(X)$ et $\psi(Y)$ et l'on peut écrire:

$$\begin{aligned} & |\mathbb{E}[\varphi(X)\psi(Y)] - \mathbb{E}[\varphi(X_n)\psi(Y_n)]| = \\ & = |\mathbb{E}[(\varphi(X) - \varphi(X_n))\psi(Y)] + \mathbb{E}[\varphi(X_n)(\psi(Y) - \psi(Y_n))]| \leq \\ & \leq \|\psi\|_\infty \|\varphi(X) - \varphi(X_n)\|_1 + \|\varphi\|_\infty \|\psi(Y) - \psi(Y_n)\|_1 \end{aligned}$$

Il suffit alors d'utiliser la relation $\mathbb{E}[\varphi(X_n)\psi(Y_n)] = \mathbb{E}[\varphi(X_n)]\mathbb{E}[\psi(Y_n)]$. \square

De même on peut définir l'indépendance d'une famille $(X_k)_{k=1, \dots, n}$ de v.a. à valeurs dans \mathbb{R}^{m_k} par le fait que la loi de $X = (X_1, \dots, X_n)$, à valeurs dans $\mathbb{R}^{m_1 + \dots + m_n}$, est le produit des lois des X_k . On vérifie que, si X_k , $k = 1 \dots n$, sont des v.a. indépendantes à valeurs dans \mathbb{R}^{m_k} alors les v.a. $f_k(X_k)$ à valeurs dans \mathbb{R}^{s_k} le sont aussi pour toute famille de fonctions mesurables f_k définies sur \mathbb{R}^{m_k} et à valeurs dans \mathbb{R}^{s_k} .

Proposition 2.8. Soit (X_1, \dots, X_n) des v.a. indépendantes dont les lois possèdent des densités (f_1, \dots, f_n) . Alors $X = (X_1, \dots, X_n)$ possède une densité $f(x_1, \dots, x_n)$ donnée par $\prod_{k=1}^n f_k(x_k)$. Réciproquement si la variable aléatoire X possède une densité de cette forme alors les variables aléatoires (X_1, \dots, X_n) sont indépendantes.

Définition 2.9. Soit X et Y des v.a. réelles dans $L^2(\Omega, \mathbb{P})$. On définit

- $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$
- $\text{Var}(X) = \text{Cov}(X, X)$.

Proposition 2.10. 1. Soit X et Y des v.a. réelles indépendantes et intégrables. Alors le produit XY est intégrable et $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. Par conséquent la covariance de deux v.a. réelles indépendantes est nulle (la réciproque étant fausse).

2. Soit X_1, \dots, X_n et Y_1, \dots, Y_m des v.a. réelles dans $L^2(\Omega, \mathbb{P})$. Alors:

$$\text{Cov}\left(\sum_{k=1}^n \lambda_k X_k, \sum_{r=1}^m \mu_r Y_r\right) = \sum_{k=1}^n \sum_{r=1}^m \lambda_k \mu_r \text{Cov}(X_k, Y_r)$$

Si X_1, \dots, X_n sont indépendantes alors:

$$\text{Var}\left(\sum_{k=1}^n \lambda_k X_k\right) = \sum_{k=1}^n \lambda_k^2 \text{Var}(X_k)$$

Pour des v.a. réelles centrées et de carré sommable l'indépendance implique l'orthogonalité dans $L^2(\Omega, \mathbb{P})$ et donc en appliquant la proposition 1.19 on obtient:

Proposition 2.11. Soit $(X_n)_n$ une suite de v.a. réelles indépendantes dans $L^2(\Omega, \mathbb{P})$, centrées et telles que $\sum_n \text{Var}(X_n) < \infty$. Alors la série $\sum_n X_n$ converge dans $L^2(\Omega, \mathbb{P})$.

Un théorème très important en probabilités est la loi forte des grands nombres (ou théorème ergodique). Nous en donnons une preuve dans $L^2(\Omega, \mathbb{P})$ qui sera reprise dans le cadre des processus stationnaires. En fait ce théorème est valable dans $L^1(\Omega, \mathbb{P})$ mais sa démonstration est alors beaucoup plus difficile.

Théorème 2.12. Soit X_n une suite de v.a. réelles dans $L^2(\Omega, \mathbb{P})$, indépendantes et de même loi. Alors en notant m leur espérance commune on a:

$$M_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n) \rightarrow m, \quad p.s.$$

Démonstration. On peut supposer $m = 0$. L'inégalité de Bienaymé-Chebyshev:

$$\mathbb{P}(|M_n| > \varepsilon) \leq \frac{\text{Var}(X)}{n\varepsilon^2}$$

jointe au lemme de Borel-Cantelli montre que la sous suite M_{n^2} converge presque sûrement vers 0. Il reste à contrôler les termes M_k pour $n^2 \leq k \leq (n+1)^2$.

$$\begin{aligned} M_k - M_{n^2} &= \frac{X_1 + X_2 + \dots + X_k}{k} - \frac{X_1 + X_2 + \dots + X_{n^2}}{n^2} \\ &= \left(\frac{n^2}{k} - 1\right)M_{n^2} + \frac{X_{n^2+1} + \dots + X_k}{k} \\ |M_k - M_{n^2}| &\leq |M_{n^2}| + Z_n \end{aligned}$$

Avec

$$Z_n = \sup_{n^2 \leq k \leq (n+1)^2} \left| \frac{X_{n^2+1} + \dots + X_k}{k} \right| \leq \frac{|X_{n^2+1}| + \dots + |X_{(n+1)^2}|}{n^2}$$

On a

$$\mathbb{E}(Z_n^2) \leq \frac{1}{n^4} \sum_{i,j=n^2+1}^{(n+1)^2} \mathbb{E}(|X_i||X_j|) \leq \frac{1}{n^4} \sigma^2((n+1)^2 - n^2)^2 = \frac{1}{n^4} \sigma^2(2n+1)^2$$

Une nouvelle application de l'inégalité de Bienaymé-Chebyshev à Z_n montre que Z_n converge presque sûrement vers 0. Il ne reste plus qu'à écrire $|M_k| \leq |M_k - M_{n^2}| + |M_{n^2}| \leq 2|M_{n^2}| + Z_n$ pour conclure. \square

2.4 V.a. complexes

On peut sans difficultés étendre au cas complexe les notions ci-dessus. Soit X et Y des v.a. réelles. On dit que $Z = X + iY$ est dans L^p si X et Y sont dans L^p (ou de façon équivalente si $|Z| \in L^p$).

Définition 2.13. Soit $Z_1 = X_1 + iY_1$ et $Z_2 = X_2 + iY_2$ deux v.a. complexes.

1. On dit que Z_1 et Z_2 sont indépendantes si le couple (X_1, Y_1) est indépendant du couple (X_2, Y_2)
2. On définit

$$\text{Cov}(Z_1, Z_2) = \mathbb{E}[(Z_1 - \mathbb{E}(Z_1))(\overline{Z_2 - \mathbb{E}(Z_2)})] = \mathbb{E}(Z_1 \overline{Z_2}) - \mathbb{E}(Z_1)\mathbb{E}(\overline{Z_2})$$

On a alors un résultat analogue au cas réel:

Proposition 2.14. 1. Soit Z_1 et Z_2 des v.a. complexes indépendantes et intégrables. Alors le produit $Z_1 Z_2$ est intégrable et $\mathbb{E}(Z_1 \overline{Z_2}) = \mathbb{E}(Z_1)\mathbb{E}(\overline{Z_2})$. Par conséquent la covariance de deux v.a. complexes indépendantes est nulle (la réciproque étant fautive).

2. Soit $Z_1, \dots, Z_n, U_1, \dots, U_m$ des v.a. complexes dans $L^2(\Omega, \mathbb{P})$. Alors:

$$\text{Cov}\left(\sum_{k=1}^n \lambda_k Z_k, \sum_{r=1}^m \mu_r U_r\right) = \sum_{k=1}^n \sum_{r=1}^m \lambda_k \overline{\mu_r} \text{Cov}(Z_k, U_r)$$

Si Z_1, \dots, Z_n sont indépendantes alors:

$$\text{Var}\left(\sum_{k=1}^n \lambda_k Z_k\right) = \sum_{k=1}^n |\lambda_k|^2 \text{Var}(Z_k)$$

2.5 Fonctions caractéristiques

Définition 2.15. Par analogie avec le cas réel, pour des vecteurs t et x dans \mathbb{R}^d on pose $\gamma_t(x) = \exp(-2i\pi t'x)$ où t' désigne le transposé du vecteur colonne t .

Soit X une v.a. à valeurs dans \mathbb{R}^d et de loi μ . On appelle fonction caractéristique (ou transformée de Fourier de μ) la fonction définie pour $t \in \mathbb{R}^d$ par

$$\hat{\mu}(t) = \varphi_X(t) = \mathbb{E}(\gamma_{-t}(X)) = \int_{\mathbb{R}^d} \gamma_{-t}(x) d\mu(x)$$

Si la v.a. X a une densité f , alors $\varphi_X(t) = \int f(x)\gamma_{-t}(x) dx$. Donc la fonction caractéristique n'est rien d'autre que la transformée de Fourier de f . Il faut remarquer qu'en général, les traités de probabilité omettent le facteur 2π dans le calcul de la fonction caractéristique qui est alors définie par $\varphi_X(t) = \int_{\mathbb{R}^d} e^{it'x} d\mu(x)$ ce qui simplifie certaines formules, en particulier dans le cas gaussien. Pour des raisons de compatibilité avec le reste du polycopié, nous n'adopterons pas cette convention ici...

Si X est une v.a. d -dimensionnelle, et A une matrice $m \times d$ et $b \in \mathbb{R}^m$, alors il est immédiat d'exprimer la fonction caractéristique de Y à partir de celle de X :

$$\varphi_Y(t) = \mathbb{E}(e^{2i\pi t'(AX+b)}) = e^{2i\pi t'b} \mathbb{E}(e^{2i\pi t'AX}) = e^{2i\pi t'b} \varphi_X(A't), \quad t \in \mathbb{R}^m \quad (2.2)$$

Nous commençons par montrer que la fonction caractéristique définit la loi de X . La preuve est donnée pour $d = 1$ mais s'étend sans difficulté au cas général.

Proposition 2.16. L'application $\mu \rightarrow \hat{\mu}$ est injective.

Démonstration. Soit \mathcal{E} l'ensemble des transformées de Fourier de fonctions de $L^1(\mathbb{R}, dx)$. D'après le théorème de Stone-Weierstrass \mathcal{E} est dense dans \mathbf{C}_0 . Soit $f \in L^1(\mathbb{R}, dx)$; on peut écrire en utilisant Fubini

$$\mu(\hat{f}) = \int \left(\int e^{2i\pi tx} f(x) dx \right) d\mu(t) = \int f(x) \hat{\mu}(x) dx$$

Par conséquent si deux mesures bornées μ et ν ont la même transformée de Fourier elles coïncident sur \mathcal{E} . Or l'application $g \rightarrow \mu(g)$ est continue sur \mathbf{C}_0 . Donc μ et ν coïncident sur \mathbf{C}_0 \square

Proposition 2.17. Soit μ_n et μ des probabilités. Alors μ_n converge étroitement vers μ si et seulement si $\hat{\mu}_n(t)$ converge vers $\hat{\mu}(t)$ pour tout t .

Démonstration. En reprenant les notations de la proposition précédente on voit que $\mu_n(g) \rightarrow \mu(g)$ pour toute fonction $g \in \mathcal{E}$. On prouve sans peine que ceci implique la même convergence pour les fonctions de \mathbf{C}_0 d'où la convergence étroite d'après 2.3. \square

Proposition 2.18. Soit (X_1, X_2, \dots, X_n) des v.a. de somme S_n Alors:

1. Si ces v.a. sont indépendantes alors $\varphi_{S_n}(t) = \prod_{k=1}^n \varphi_{X_k}(t)$, pour tout t . La réciproque est fausse.
2. On a $\varphi_{(X_1, X_2, \dots, X_n)}(t_1, t_2, \dots, t_n) = \prod_{k=1}^n \varphi_{X_k}(t_k)$ pour tout (t_1, t_2, \dots, t_n) si et seulement si les v.a. X_1, \dots, X_n sont indépendantes.

Démonstration. (1): C'est une conséquence de Fubini puisque

$$\mathbb{E}(e^{2it\pi S_n}) = \mathbb{E}\left(\prod_{k=1}^n e^{2i\pi t X_k}\right) = \prod_{k=1}^n \mathbb{E}(e^{2i\pi t X_k})$$

(2): En utilisant Fubini, la loi produit des lois des X_k a pour fonction caractéristique le second membre de l'égalité. Le résultat est alors une conséquence de 2.16. \square

La Proposition 1.45 1. indique que l'on peut calculer l'intégrale de $x \rightarrow xf(x)$, simplement en dérivant sa transformée de Fourier en 0. Ceci à condition, bien sûr, que $xf(x)$ soit intégrable. De la même façon, les fonctions caractéristiques peuvent servir au calcul des moments d'un vecteur aléatoire.

Proposition 2.19. *Si le vecteur aléatoire X a tous ses moments d'ordre m finis, alors sa fonction caractéristique est m fois dérivable et*

$$\frac{\partial^{|\alpha|} \varphi_X}{\partial x^\alpha}(t) = (2i\pi)^{|\alpha|} \mathbb{E}(X^\alpha e^{2i\pi t' X}) \quad (2.3)$$

pour tout multi-indice $\alpha = (\alpha_1, \dots, \alpha_d)$, avec $|\alpha| = \alpha_1 + \dots + \alpha_d \leq m$.

En particulier, pour une v.a. réelle X intégrable, on a $2i\pi \mathbb{E}(X) = \varphi'_X(0)$ et si de plus, elle est de carré intégrable, alors $-4\pi^2 \mathbb{E}(X^2) = \varphi''_X(0)$.

Exemple 2.20. On considère une v.a. réelle de densité gaussienne centrée réduite, c'est-à-dire de densité $f_X(x) = (2\pi)^{-1/2} e^{-x^2/2}$. En reprenant le calcul fait dans 1.61 on obtient sa fonction caractéristique $\varphi_X(t) = e^{-2\pi^2 t^2}$. De même, le calcul fait dans 1.59 montre qu'une loi de Cauchy de densité $f_X(x) = \frac{1}{\pi(1+x^2)}$ a pour fonction caractéristique $\varphi_X(t) = e^{-2\pi|t|}$.

2.6 Processus Gaussiens

On note M' la transposée d'une matrice M .

Définition 2.21. *Matrices de covariance:*

- Pour un vecteur aléatoire réel ou complexe Z dans $L^2(\Omega, \mathbb{P})$ on définit le vecteur espérance $m_Z = \mathbb{E}(Z)$, et la matrice de covariance par:

$$\Sigma_Z = \mathbb{E}[(Z - m_Z)(\overline{Z' - m'_Z})]$$

Cette matrice est hermitienne positive. (Ici et dans la suite les vecteurs Z sont des vecteurs colonne).

- Pour deux vecteurs aléatoires réels ou complexes U et V dans $L^2(\Omega, \mathbb{P})$ on définit leur matrice de covariance "croisée" par

$$\Sigma_{(U,V)} = \mathbb{E}[(U - m_U)(\overline{V' - m'_V})]$$

On a évidemment $\Sigma_{(V,U)} = \overline{\Sigma'_{(U,V)}}$.

Remarque 2.22. Si L est la matrice d'une application linéaire de \mathbb{C}^n dans \mathbb{C}^p (ou de \mathbb{R}^n dans \mathbb{R}^p), le vecteur $Y = LZ$ a pour espérance $m_Y = Lm_Z$ et pour matrice de covariance $\Sigma_Y = L\Sigma_Z\overline{L'}$.

Définition 2.23. Variables gaussiennes réelles:

- On dit qu'une v.a. réelle X est gaussienne si sa fonction caractéristique est

$$\mathbb{E}[e^{2i\pi tX}] = e^{2i\pi tm - 2\pi^2\sigma^2 t^2} \quad (2.4)$$

pour $m, \sigma \in \mathbb{R}$. On dit alors que X suit une loi $\mathcal{N}(m, \sigma^2)$.

- On dit que le vecteur (X_1, \dots, X_n) est gaussien, si toute combinaison linéaire à coefficients réels des X_1, \dots, X_n est une v.a. gaussienne réelle.

Remarque 2.24. Il est utile de souligner une conséquence évidente de la définition 2.23: si X est une v.a. gaussienne m -dimensionnelle, A une matrice $d \times m$, $b \in \mathbb{R}^d$, alors la v.a. $Y = AX + b$ est une v.a. gaussienne (d -dimensionnelle, évidemment). En effet, toute combinaison linéaire des Y_1, \dots, Y_d peut s'écrire comme une combinaison linéaire des X_1, \dots, X_m plus un réel.

Remarque 2.25. Soit X une v.a. réelle gaussienne.

- Pour tout $p > 0$ la v.a. réelle $|X|^p$ est sommable, en particulier, $m = \mathbb{E}(X)$ et $\sigma^2 = \text{Var}(X)$. Il suffit en effet de remarquer que la fonction caractéristique (2.4) est indéfiniment dérivable et appliquer la proposition 2.19.
- Vu (2.2) et l'Exemple 2.20, la v.a. $Y = \sigma X + m$, où X suit une loi $\mathcal{N}(0, 1)$, suit une loi $\mathcal{N}(m, \sigma^2)$. Donc, grâce à (2.1) (avec la matrice 1×1 $A = \sigma^2$), si $\sigma \neq 0$ la loi de Y a pour densité $f_Y(x) = (2\pi\sigma^2)^{-1/2} e^{-(x-m)^2/(2\sigma^2)}$. Par contre, si $\sigma = 0$, c'est la loi d'une v.a. qui ne prend que la valeur m avec probabilité 1 (et qui n'a pas de densité).

Proposition 2.26. Un vecteur aléatoire réel (X_1, X_2, \dots, X_n) est gaussien si et seulement si sa fonction caractéristique est donnée par

$$\varphi_X(t) = \mathbb{E}[e^{2i\pi t'X}] = e^{2i\pi t'm_X - 2\pi^2 t'\Sigma_X t}, \quad t \in \mathbb{R}^n \quad (2.5)$$

Donc la donnée de la moyenne m_X et de la matrice de covariance Σ_X déterminent la loi de X , que l'on note $\mathcal{N}(m_X, \Sigma_X)$.

Démonstration. On suppose que (2.5) est valable. Alors, pour toute combinaison linéaire $Z = t'X$, on a

$$\phi_Z(\lambda) = \mathbb{E}[e^{2i\pi\lambda t'X}] = e^{2i\pi\lambda m - 2\pi^2\lambda^2\sigma^2}$$

où $m = \mathbb{E}(t'X) = t'm_X$ et $\sigma^2 = t'\Sigma_X t$. Z est donc gaussienne. Inversement, si X est gaussienne, alors

$$\varphi_X(t) = \mathbb{E}[e^{2i\pi t'X}] = \phi_Z(1) = e^{2i\pi m - 2\pi^2\sigma^2} = e^{2i\pi t'm_X - 2\pi^2 t'\Sigma_X t}$$

Comme toute combinaison linéaire de (X_1, \dots, X_n) s'écrit sous la forme $t'X$, la définition 2.23 est vérifiée. \square

Lemme 2.27. Soit $X = (U, V)$ un vecteur gaussien dans \mathbb{R}^n avec $U \in \mathbb{R}^p$ et $V \in \mathbb{R}^{n-p}$ alors U et V sont des variables aléatoires indépendantes si et seulement si $\Sigma_{U,V} = 0$.

Démonstration. Il suffit de constater que $\Sigma_{U,V} = 0$ est équivalent au fait que le produit des fonctions caractéristiques de U et V aux points t_1 et t_2 est égal à la fonction caractéristique de X au point (t_1, t_2) . \square

Exemple 2.28. Soit X_1, \dots, X_m des v.a. réelles $N(0, 1)$ indépendantes, alors le vecteur $X = (X_1, \dots, X_m)$ est gaussien. En effet la définition 2.23 est satisfaite, car la somme de v.a. gaussiennes indépendantes est toujours gaussienne. Évidemment ici $m_X = 0$ et $\Sigma_X = I$ (matrice identité), donc X suit une loi $\mathcal{N}(0, I)$. Cette v.a. a une densité, car, grâce à la proposition 2.8,

$$\begin{aligned} f_X(x_1, \dots, x_m) &= f_{X_1}(x_1) \dots f_{X_m}(x_m) = (2\pi)^{-m/2} e^{-x_1^2/2} \dots e^{-x_m^2/2} = \\ &= (2\pi)^{-m/2} e^{-\frac{1}{2}|x|^2} \end{aligned} \quad (2.6)$$

On a vu que les v.a. gaussiennes réelles ont une densité si et seulement si leur variance est non nulle. Que peut-on dire dans le cas multidimensionnel? L'outil de calcul de base est le résultat d'algèbre linéaire suivant.

Proposition 2.29. Soit Σ une matrice symétrique (resp. hermitienne) positive. Il existe alors une matrice A telle que $AA' = \Sigma$.

Démonstration. Si la matrice Σ est diagonale, les éléments Σ_{jj} sur la diagonale sont ≥ 0 , Σ étant positive. La matrice diagonale formée avec les racines carrées des Σ_{jj} fait l'affaire. Sinon on utilise la diagonalisation $\Sigma = KDK'$ avec K orthogonale (resp. unitaire) et si B est telle que $BB' = D$, alors $A = KBK'$ satisfait à la condition requise. \square

Si X est gaussienne $\mathcal{N}(0, I)$ et $AA' = \Sigma$, alors $Y = AX + m$ est une v.a. $\mathcal{N}(m, \Sigma)$. Si Σ est inversible on peut déduire la densité de Y , grâce aux relations (2.1) et (2.6):

$$\begin{aligned} f_Y(x) &= \frac{1}{|\det A|} f_X(A^{-1}(x - m)) = \frac{1}{(2\pi)^{m/2} |\det A|} \exp\left(-\frac{1}{2}|A^{-1}(x - m)|^2\right) = \\ &= \frac{1}{(2\pi)^{m/2} \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(x - m)' \Sigma^{-1} (x - m)\right) \end{aligned} \quad (2.7)$$

On remarque que l'on vient également de prouver que, pour tout $m \in \mathbb{R}^d$ et pour toute matrice Σ symétrique positive, il existe toujours un vecteur d -dimensionnel Y qui soit $\mathcal{N}(m, \Sigma)$, celui-ci n'ayant une densité que si Σ est inversible.

Le théorème suivant, très classique, est le *Théorème Central Limite* (TCL).

Théorème 2.30. Soit X_n une suite de v.a. à valeurs dans \mathbb{R}^d , indépendantes, de même loi, de carré intégrable, ayant pour moyenne m , et pour matrice de covariance Σ . On pose $S_n = X_1 + \dots, X_n$. Alors $W_n = \frac{1}{\sqrt{n}}(S_n - nm)$ converge en loi vers $N_d(0, \Sigma)$.

Démonstration. On peut supposer $m = 0$. Soit $Y \sim N_d(0, \Sigma)$. On a $\varphi_{W_n}(t) = (\varphi_{X_1}(\frac{t}{\sqrt{n}}))^n$. De plus, vu la proposition 2.19, $\frac{\partial}{\partial t_\Sigma} \varphi_{X_1}(0) = 0 = \frac{\partial}{\partial t_\Sigma} \varphi_Y(0)$, $\frac{\partial^2}{\partial t_j t_k} \varphi_{X_1}(0) = -4\pi^2 \Sigma_{j,k} = \frac{\partial^2}{\partial t_j t_k} \varphi_Y(0)$. On a donc

$$\phi_{X_1}(t) - \varphi_Y(t) = |t|^2 \varepsilon(t), \quad |\varepsilon(t)| \rightarrow_{t \rightarrow 0} 0.$$

Si $u, v \in \mathbb{C}$ avec $|u| \leq 1$, $|v| \leq 1$, on a $|u^n - v^n| \leq n|u - v|$ et donc, vu que $\varphi_Y(t) = \exp(-2\pi^2 t' \Sigma t) = (\varphi_Y(\frac{t}{\sqrt{n}}))^n$,

$$\begin{aligned} |\varphi_{W_n}(t) - \varphi_Y(t)| &= |(\varphi_{X_1}(\frac{t}{\sqrt{n}}))^n - (\varphi_Y(\frac{t}{\sqrt{n}}))^n| \\ &\leq n|\varphi_{X_1}(\frac{t}{\sqrt{n}}) - \varphi_Y(\frac{t}{\sqrt{n}})| \leq n \frac{|t|^2}{n} |\varepsilon(\frac{t}{\sqrt{n}})| \leq |t|^2 |\varepsilon(\frac{t}{\sqrt{n}})| \rightarrow_n 0 \end{aligned}$$

et W_n converge en loi vers $N_d(0, \Sigma)$ d'après la proposition 2.17 \square

Définition 2.31. Variables gaussiennes complexes:

- On dit qu'une v.a. complexe $Z = X + iY$ est gaussienne si X et Y sont des variables aléatoires gaussiennes réelles indépendantes et de même variance (attention: une variable gaussienne réelle ne peut donc pas être considérée comme gaussienne complexe de partie imaginaire nulle).
- On dit que le vecteur (Z_1, Z_2, \dots, Z_n) est gaussien, si toute combinaison linéaire à coefficients complexes des Z_k , $k = 1, \dots, n$ est une v.a. gaussienne complexe.

La définition 2.31 peut paraître excessivement exigeante. On aurait pu décider de dire qu'une v.a. complexe $Z = X + iY$ est gaussienne dès que le couple (X, Y) est gaussien dans \mathbb{R}^2 . Mais, dans ce cas, la loi de Z ne serait plus déterminée par son espérance et sa variance σ_Z^2 . En effet, en admettant que Z soit centrée, la variance de Z vaut

$$\sigma_Z^2 = \mathbb{E}(Z\bar{Z}) = \mathbb{E}(X^2 + Y^2)$$

La connaissance de σ_Z^2 est donc insuffisante à déterminer la loi de (X, Y) et donc de Z . Par contre si on rajoute l'hypothèse X et Y indépendantes et de même variance, on déduit immédiatement que (X, Y) est gaussienne bidimensionnelle de matrice de covariance $\begin{pmatrix} \frac{1}{2}\sigma_Z^2 & 0 \\ 0 & \frac{1}{2}\sigma_Z^2 \end{pmatrix}$.

Proposition 2.32. Un vecteur aléatoire complexe (Z_1, Z_2, \dots, Z_n) est gaussien si et seulement si, en posant $X = \Re(Z)$ et $Y = \Im(Z)$, on a:

1. Le vecteur (X, Y) de dimension $2n$ est gaussien réel.
2. $\Sigma_X = \Sigma_Y$ et $\Sigma_{(X,Y)} = -\Sigma_{(Y,X)}$.

En particulier $\Sigma_Z = 2(\Sigma_X + i\Sigma_{(Y,X)})$. Donc la donnée de m_Z et Σ_Z déterminent la loi de Z .

Démonstration. Soit $a = \alpha + i\beta$ un élément de \mathbb{C}^n :

$$\begin{aligned} a'Z &= (\alpha'X - \beta'Y) + i(\alpha'Y + \beta'X) = U + iV \\ \Sigma_U &= \alpha'\Sigma_X\alpha + \beta'\Sigma_Y\beta - \alpha'\Sigma_{X,Y}\beta - \beta'\Sigma'_{X,Y}\alpha \\ \Sigma_V &= \alpha'\Sigma_Y\alpha + \beta'\Sigma_X\beta + \alpha'\Sigma'_{X,Y}\beta + \beta'\Sigma_{X,Y}\alpha \\ \Sigma_{U,V} &= \alpha'\Sigma_{X,Y}\alpha + \alpha'\Sigma_X\beta - \beta'\Sigma_Y\alpha - \beta'\Sigma'_{X,Y}\beta \end{aligned}$$

On suppose Z gaussien. Alors $V = \alpha'Y + \beta'X$ est gaussien réel pour tout choix de α et β et donc (X, Y) est gaussien dans \mathbb{R}^{2n} . En faisant $\beta = 0$ la relation $\text{Var}(U) = \text{Var}(V)$

entraîne $\Sigma_X = \Sigma_Y$ et l'indépendance de U et V implique $\Sigma_{U,V} = \alpha' \Sigma_{X,Y} \alpha = 0$ pour tout α et donc $\Sigma_{X,Y}$ antisymétrique. On a donc prouvé 1. et 2.

On suppose 1. et 2. vérifiés. D'après 1. le couple (U, V) est gaussien dans \mathbb{R}^2 . Les relations ci-dessus et 2. entraînent $\Sigma_U = \Sigma_V$ et $\Sigma_{U,V} = 0$ d'où l'on déduit que $a'Z$ est une variable gaussienne complexe.

On a $\Sigma_Z = \Sigma_X + \Sigma_Y - i\Sigma_{(X,Y)} + i\Sigma_{(Y,X)} = 2(\Sigma_X + i\Sigma_{(Y,X)})$ □

Définition 2.33. *Un processus réel (resp complexe) X_t est dit gaussien si tout système fini extrait est un vecteur aléatoire gaussien réel (resp. complexe).*

Il est facile de constater que l'on peut associer à un processus gaussien réel, un processus gaussien complexe ayant la même fonction de covariance (en considérant deux copies indépendantes)

Proposition 2.34. *Soit X_t un processus gaussien réel (resp complexe). Alors la fermeture dans $L^2(\Omega, \mathbb{P})$ des combinaisons linéaires finies à coefficients réels (resp complexes) d'éléments de la famille X_t est formée de variables aléatoires gaussiennes réelles (resp complexes).*

Démonstration. Le cas réel. On vérifie que si une suite X_n de v.a. réelles converge dans $L^2(\Omega, \mathbb{P})$ vers X alors $\mathbb{E}(X_n) \rightarrow \mathbb{E}(X)$ et $\text{Var}(X_n) \rightarrow \text{Var}(X)$. La convergence en moyenne quadratique implique la convergence étroite et donc la convergence des fonctions caractéristiques ce qui termine la preuve.

Le cas complexe. Si une suite gaussienne complexe $Z_n = X_n + iY_n$ converge dans $L^2(\Omega, \mathbb{P})$ vers $Z = X + iY$ alors $X_n \rightarrow X$ et $Y_n \rightarrow Y$ dans L^2 . Donc X et Y sont gaussiennes et de même variance. Elles sont aussi indépendantes, grâce à la proposition 2.7. □

Remarque 2.35. Si $(X_t)_t$ est un processus gaussien réel, on indique $H(X)$ la fermeture dans L^2 des combinaisons, à coefficients complexes des X_t . On a alors que $\Re(H(X))$ et $\Im(H(X))$ sont formés de variables gaussiennes. Mais $H(X)$ n'est pas un espace gaussien complexe.

La proposition suivante montre que la seule considération des propriétés du second ordre d'un processus stationnaire est tout à fait insuffisante pour caractériser celui-ci puisqu'il existe toujours un processus gaussien stationnaire avec les mêmes propriétés du second ordre .

Proposition 2.36. *Soit c une fonction de type positif sur T . Si c est réelle (resp. complexe), c est la fonction de covariance d'un p.s.c. gaussien réel (resp. complexe).*

Démonstration. La matrice $M_n(i, j) = c(t_i - t_j)$ est définie positive. On construit, pour tout n , la loi de probabilité gaussienne μ_n sur $\prod_{k=-n}^n \mathbb{R}$ associée à la matrice de covariance M_n et on utilise le théorème de limite projective des probabilités. □

En fait un processus gaussien stationnaire au sens défini plus haut vérifie une propriété de stationnarité beaucoup plus forte: La loi de tout système fini extrait de ce processus est invariante par translation des indices (puisque'elle ne dépend que de la matrice de covariance).

2.7 Le Modèle linéaire

On dispose d'un appareil de mesure que l'on cherche à étalonner. A priori, le signal x reçu est de type vectoriel, que on l'écrit sous la forme d'un vecteur ligne à p composantes. On suppose que la mesure y de ce signal, effectuée par cet appareil, s'écrit sous la forme $y = x\theta + \sigma b$, où θ est un paramètre inconnu dans \mathbb{R}^p et b un "bruit aléatoire" que l'on suppose suivre une loi normale centrée réduite. On se propose d'étalonner cet appareil (c'est à dire d'estimer θ et σ) en envoyant une suite de $n > p$ signaux connus et en notant les n mesures faites, en supposant que les bruits successifs sont indépendants. On note X la matrice ($n \times p$) avec $p \leq n$ dont les vecteurs lignes sont les n signaux envoyés, Y le vecteur colonne des mesures et B le vecteur des bruits. Dans toute la suite, on suppose que la matrice X est de rang p . Le modèle s'écrit donc

$$Y = X\theta + \sigma B$$

Définition 2.37. On pose $\hat{\theta} = (X'X)^{-1}X'Y$ et $\hat{\sigma}^2 = \frac{1}{n-p}\|Y - X\hat{\theta}\|^2$.

Théorème 2.38. (Gauss-Markov) Les estimateurs $\hat{\theta}$ et $\hat{\sigma}^2$ des paramètres θ et σ^2 possèdent les propriétés suivantes:

- Ils sont sans biais et indépendants.
- Ce sont des estimateurs de variance minimum dans la classe des estimateurs sans biais du couple (θ, σ^2) .
- $\hat{\theta}$ suit une loi gaussienne d'espérance θ et de matrice de covariance $\sigma^2(X'X)^{-1}$.
- $\frac{n-p}{\sigma^2}\hat{\sigma}^2$ suit une loi de χ^2 à $n-p$ degrés de liberté.

Pour calculer des intervalles ou des régions de confiance ainsi que pour effectuer des tests on utilise le résultat suivant:

Proposition 2.39. Sous les hypothèses ci dessus on a:

- Soit d_k le terme diagonal d'ordre k de la matrice $(X'X)^{-1}$. Pour $1 \leq k \leq p$ la v.a. $\frac{\hat{\theta}_k - \theta_k}{\sqrt{d_k \hat{\sigma}^2}}$ suit une loi de Student à $n-p$ degrés de liberté.
- La v.a. $\frac{(n-p)\|X\hat{\theta} - X\theta\|^2}{p\|Y - X\hat{\theta}\|^2}$ suit une loi de Fisher à $(p, n-p)$ degrés de liberté.

Chapitre 3

Théorie spectrale des processus du second ordre

3.1 Processus du second ordre

Soit T un ensemble d'indices égal à \mathbb{R} ou \mathbb{Z} . Dans la suite on appelle processus du second ordre une famille de v.a. réelles ou complexes $\{X_t ; t \in T\}$ appartenant à $L^2(\Omega, \mathbb{P})$. Dans toute la suite les espaces fonctionnels considérés sont des espaces vectoriels de fonctions à valeurs complexes sur le corps des complexes (même lorsque les processus sont réels). Dans le cas $T = \mathbb{R}$ on supposera que ce processus est mesurable c'est à dire que l'application $(\omega, t) \rightarrow X_t(\omega)$ est mesurable (pour le produit de la tribu sur Ω par la tribu borélienne de \mathbb{R}).

Définition 3.1. Pour $T = \mathbb{R}$ ou \mathbb{Z}

- On note par $H(X)$ (resp $H_t(X)$) le sous espace de Hilbert de $L^2(\Omega, \mathbb{P})$ égal à la fermeture dans $L^2(\Omega, \mathbb{P})$ de l'espace des combinaisons linéaires finies à coefficients complexes d'éléments de X (resp d'éléments d'indices $\leq t$).

Dans le cas $T = \mathbb{Z}$

- On appelle bruit blanc une suite orthogonale centrée de $L^2(\Omega, \mathbb{P})$ de variance constante non nulle. Lorsque cette variance est égale à 1 on dit que c'est un bruit blanc normalisé.
- On dit que le processus X est purement aléatoire si $\cap_n H_n(X) = \{0\}$.
- On dit que le processus X est déterministe si $H_n(X) = H(X)$, pour tout $n \in \mathbb{Z}$.
- On appelle processus d'innovation la suite $I_n = X_n - \text{Proj}\{X_n | H_{n-1}(X)\}$

Proposition 3.2. Soit X_n un processus du second ordre. Alors on peut décomposer X_n sous la forme $X_n = X'_n + X''_n$ de telle façon que la suite X'_n soit purement aléatoire, la suite X''_n soit déterministe et que $H(X) = H(X') \oplus H(X'')$.

Démonstration. Soit $F = \bigcap H_n(X)$ et E l'orthogonal de F dans $H(X)$. On pose $X'_n = \text{Proj}\{X_n|F\}$ et $X''_n = \text{Proj}\{X_n|E\}$. Il reste à prouver que X'_n est déterministe et X''_n purement aléatoire. Or $H_n(X') = \text{Proj}\{H_n(X)|F\} = F$ et de même:

$$\bigcap H_n(X'') = \bigcap \text{Proj}\{H_n(X)|E\} = \text{Proj}\{\bigcap H_n(X)|E\} = \{0\}. \quad \square$$

Proposition 3.3. *Le processus d'innovation I_n est une suite orthogonale et on a l'inclusion $H_n(I) \subset H_n(X)$. On a égalité si et seulement si le processus X est purement aléatoire.*

Démonstration. Soit $m > n$ alors $I_m \in (H_{m-1})^\perp \subset (H_n)^\perp$ et de plus $I_n \in H_n$ d'où l'orthogonalité. La relation $I_n \in H_n$ implique $H_n(I) \subset H_n(X)$. Ensuite si on note par $e(v_1, v_2, \dots, v_k)$ l'espace vectoriel engendré par les vecteurs (v_1, v_2, \dots, v_k) on voit que pour $p \geq 0$

$$H_n(X) = e(I_n) \oplus H_{n-1}(X) = \dots = e(I_n, I_{n-1}, \dots, I_{n-p}) \oplus H_{n-p-1}(X)$$

Si $Y \in H_n(X)$, alors en projetant Y sur les deux sous espaces $e(I_n, I_{n-1}, \dots, I_{n-p})$ et $H_{n-p-1}(X)$ on a $Y = A_p + B_p$. De plus, $A_p \rightarrow \text{Proj}\{Y|H_n(I)\}$ et en reprenant la notation $F = \bigcap H_n(X)$, on a $B_p \rightarrow \text{Proj}\{Y|F\}$. On en déduit la relation $H_n(X) = H_n(I) \oplus F$. \square

Définition 3.4. *Soit X un processus du second ordre. On pose*

$$m(t) = \mathbb{E}(X_t), \quad K(s, t) = \mathbb{E}((X_s - m(s))\overline{(X_t - m(t))})$$

- On dit que X est un processus stationnaire centré (en abrégé p.s.c.) si $m(t) = 0$ et si $K(s, t)$ ne dépend que de la différence $s - t$. On définit alors la fonction de covariance c par $c(t) = K(s + t, s) = \mathbb{E}(X_{t+s}\overline{X_s}) = \mathbb{E}(X_t\overline{X_0}) = \mathbb{E}(X_0\overline{X_{-t}})$.
- Si $T = \mathbb{R}$ on dit que le processus X est continu en moyenne quadratique (en abrégé continu en m.q.) si l'application $t \rightarrow X_t$ de \mathbb{R} dans $L^2(\Omega, \mathbb{P})$ est continue.

Proposition 3.5. *Soit X_n un p.s.c. L'application $\mathcal{T}(X_n) = X_{n-1}$ se prolonge par linéarité en un opérateur unitaire de $H(X)$.*

Démonstration. On adopte les notations de 1.11 avec $H_1 = H_2 = H(X)$ et $F_1 = F_2$ le sous espace vectoriel formé des combinaisons linéaires finies des X_n . On pose alors $\mathcal{T}(\sum_{k=1}^n \lambda_k X_k) = \sum_{k=1}^n \lambda_k X_{k-1}$ et l'on a:

$$\begin{aligned} \left\| \mathcal{T}\left(\sum_{k=1}^n \lambda_k X_k\right) \right\|_2^2 &= \sum_{k=1}^n \sum_{r=1}^n \lambda_k \bar{\lambda}_r \mathbb{E}(X_{k-1} \overline{X_{r-1}}) = \\ &= \sum_{k=1}^n \sum_{r=1}^n \lambda_k \bar{\lambda}_r \mathbb{E}(X_k \overline{X_r}) = \left\| \sum_{k=1}^n \lambda_k X_k \right\|_2^2 \end{aligned}$$

Cette relation montre en particulier, que les images par \mathcal{T} de deux combinaisons linéaires des X_n correspondant à la même fonction de F_1 sont égales. L'opérateur \mathcal{T} est donc bien défini sur F_1 et est une isométrie surjective sur F_2 . \square

Si B_n est un bruit blanc, il est égal à son innovation I_n . Il s'en suit que $H_n(B) = H_n(I)$ et que donc B_n est purement aléatoire. En fait on a le résultat suivant:

Proposition 3.6. Soit X_n un p.s.c. Alors le processus d'innovation est un bruit blanc. De plus X_n est purement aléatoire si et seulement si il existe $a \in \ell^2(\mathbb{Z})$ et un bruit blanc B_n tel que $X_n = \sum_{k=0}^{\infty} a_k B_{n-k}$ (et dans ce cas, on peut choisir $B_n = I_n$).

Démonstration. Le fait que le processus d'innovation soit un bruit blanc est une conséquence de l'existence de l'isométrie \mathcal{T} . Si X_n satisfait à la formule ci dessus on a $H_n(X) \subset H_n(B)$ et donc B étant purement aléatoire il en est de même de X . Réciproquement, si X_n est purement aléatoire alors $H_n(X) = H_n(I)$. On en déduit pour tout n le développement $X_n = \sum_{k=0}^{\infty} \langle X_n, I_{n-k} \rangle I_{n-k}$. En utilisant l'isométrie \mathcal{T} on voit que le produit scalaire $\langle X_n, I_{n-k} \rangle$ est indépendant de n . \square

Proposition 3.7. La fonction de covariance d'un p.s.c. vérifie:

1. $c(0) \geq 0$, $c(-t) = \overline{c(t)}$, $|c(t)| \leq c(0)$
2. c est une fonction de type positif c'est à dire que pour tout entier $n \geq 1$ et tout système (t_1, t_2, \dots, t_n) la matrice carrée d'ordre n définie par $M(i, j) = c(t_i - t_j)$ est hermitienne et de type positif.
3. Dans le cas $T = \mathbb{R}$ on a:
 - (a) Si le processus est continu en m.q. alors c est une fonction continue.
 - (b) Si c est continue à l'origine alors le processus est continu en m.q. .

Démonstration. (1) On a $c(t) = \mathbb{E}(X_t \overline{X_0}) = \mathbb{E}(X_0 \overline{X_{-t}})$ ce qui prouve les deux premières inégalités, la troisième étant une conséquence directe de l'inégalité de Schwarz.

(2) La matrice $M(i, j) = c(t_i - t_j)$ est la matrice de covariance du vecteur aléatoire $(X_{t_1}, \dots, X_{t_n})$. Elle est donc hermitienne et de type positif, comme toute matrice de covariance.

(3) On a la relation $c(t+h) - c(t) = \langle (X_{t+h} - X_t), X_0 \rangle$ et en utilisant l'inégalité de Schwarz:

$$|c(t+h) - c(t)| \leq \|X_0\|_2 \|X_{t+h} - X_t\|_2$$

ce qui prouve (a). Ensuite

$$\|X_{t+h} - X_t\|_2^2 = \|X_{t+h}\|_2^2 + \|X_t\|_2^2 - 2\Re\langle X_{t+h}, X_t \rangle = 2(c(0) - \Re[c(h)])$$

ce qui prouve (b). \square

3.2 Théorèmes de Herglotz et Bochner

On rappelle la notation $\boxed{\gamma_x(y) = \gamma_y(x) = e^{-2i\pi xy}}$

Définition 3.8. (Transformées de Fourier des mesures bornées sur \mathbb{T} et \mathbb{R})

- Si μ est une mesure sur \mathbb{T} on pose $\hat{\mu}(n) = \int_{\mathbb{T}} \gamma_{-n}(x) d\mu(x)$
- Si μ est une mesure bornée sur \mathbb{R} on pose $\hat{\mu}(t) = \int_{\mathbb{R}} \gamma_{-t}(x) d\mu(x)$

On peut remarquer que ces définitions sont compatibles avec le calcul des fonctions caractéristiques et les définitions du premier chapitre si la mesure μ est une probabilité, dans le sens que $\phi_{\mu}(t) = \hat{\mu}(\frac{t}{2\pi})$.

Proposition 3.9. *On a les propriétés suivantes:*

1. L'application $\mu \rightarrow \hat{\mu}$ est injective.
2. La transformée de Fourier d'une mesure positive (finie dans le cas de \mathbb{R}) est une fonction de type positif (uniformément continue dans le cas de \mathbb{R}).
3. La suite de probabilités μ_n converge étroitement vers la probabilité μ si et seulement si $\hat{\mu}_n(t)$ converge vers $\hat{\mu}(t)$ pour tout $t \in T$.

Démonstration. Le cas du Tore

(1) L'application $f \rightarrow \mu(f)$ est continue sur $C(\mathbb{T})$. Par conséquent si pour deux mesures μ et ν les formes linéaires $\mu(f)$ et $\nu(f)$ coïncident sur les exponentielles complexes elles sont égales sur $C(\mathbb{T})$ puisque les combinaisons linéaires d'exponentielles sont denses dans $C(\mathbb{T})$.

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

$$\sum_{r=1}^n \sum_{l=1}^n z_r \bar{z}_l \hat{\mu}(r-l) = \int \sum_{r=1}^n \sum_{l=1}^n z_r \bar{z}_l \gamma_{-r} \gamma_l d\mu = \int \left(\left| \sum_{r=1}^n z_r \gamma_{-r} \right|^2 \right) d\mu \geq 0$$

(3) La partie "seulement si" est évidente. Soit $f \in C(\mathbb{T})$ et $\epsilon > 0$. Il existe une combinaison linéaire d'exponentielle g telle que la norme de $f - g$ dans $C(\mathbb{T})$ soit inférieure à $\epsilon/4$. D'où:

$$|\mu_n(f) - \mu(f)| \leq |\mu_n(f) - \mu_n(g)| + |\mu_n(g) - \mu(g)| + |\mu(g) - \mu(f)| \leq \epsilon/2 + |\mu_n(g) - \mu(g)|$$

et il suffit d'appliquer la convergence de $\mu_n(g)$ vers $\mu(g)$.

Dans le cas de \mathbb{R} seule la continuité uniforme de $\hat{\mu}(t)$ n'a pas été prouvée. Elle résulte de:

$$|\hat{\mu}(t+h) - \hat{\mu}(t)| \leq \int |\gamma_h(x) - 1| d\mu(x)$$

et d'une application du théorème de Lebesgue. □

On voit maintenant que certaines propriétés des fonctions de covariance, en réalité ce sont des propriétés de toute fonction de type positif.

Proposition 3.10. *Soit c une fonction de type positif sur T alors :*

1. $c(0) \geq 0$, $c(-t) = \overline{c(t)}$, $|c(t)| \leq c(0)$
2. Dans le cas $T = \mathbb{R}$, si c est continue à l'origine alors c est uniformément continue.

Démonstration. (1) On écrit la propriété de type positif

- En un seul point : on obtient $c(0) \geq 0$
- En deux points $t_1 = 0$, $t_2 = t$ d'abord avec $z_1 = z_2 = 1$: on obtient $2c(0) + c(t) + c(-t) \geq 0$, d'où $\Im m(c(-t)) = -\Im m(c(t))$. Ensuite avec $z_1 = z_2 = i$: on obtient $2c(0) + ic(t) - ic(-t) \geq 0$, d'où $\Re e(c(-t)) = \Re e(c(t))$.
- Le déterminant $\begin{vmatrix} c(0) & c(t) \\ c(-t) & c(0) \end{vmatrix}$ est positif d'où $|c(t)| \leq c(0)$.

(2) On écrit que le déterminant

$$\begin{aligned} & \begin{vmatrix} c(0) & c(t) & c(t+h) \\ c(-t) & c(0) & c(h) \\ c(-t-h) & c(-h) & c(0) \end{vmatrix} = \\ & = c(0)(c(0)^2 - |c(h)|^2) - c(t)(c(-t)c(0) - c(h)c(-t-h)) + \\ & \quad + c(t+h)(c(-t)(c(-t)c(-h) - c(-t-h))) = \\ & = c(0)(c(0)^2 - |c(h)|^2) - c(0)(|c(t+h)|^2 + |c(t)|^2) + 2\Re\{c(t)c(h)c(-t-h)\} \end{aligned}$$

est positif d'où, si on rajoute et on soustrait $2c(0)\Re\{c(t)c(t+h)\}$, on trouve

$$c(0)|c(t+h) - c(t)|^2 \leq c(0)(c(0)^2 - |c(h)|^2) + 2\Re\{\overline{c(t+h)}c(t)(c(h) - c(0))\}$$

or $|\Re\{\overline{c(t+h)}c(t)(c(h) - c(0))\}| \leq c(0)^2|c(h) - c(0)|$ et donc lorsque $h \rightarrow 0$ la différence $c(t+h) - c(t)$ tend vers 0 uniformément en t . \square

Nous allons utiliser le théorème suivant de Paul Lévy. On pourra remarquer que 3. de la Proposition 3.9 en est un cas particulier.

Proposition 3.11. (Paul Lévy) Soit $(\mu_n)_n$ une suite de mesures (positives), de masse finie sur \mathbb{R} (resp. sur \mathbb{T}) telle que pour tout $t \in \mathbb{R}$ (resp. $t \in \mathbb{Z}$) on ait $\lim_n \widehat{\mu}_n(t) = \varphi(t)$, la fonction $\varphi(t)$ étant continue à l'origine dans le cas de \mathbb{R} . Alors $\varphi(t)$ est la transformée de Fourier d'une probabilité μ sur \mathbb{R} (resp. sur \mathbb{T}) (et donc la suite μ_n converge étroitement vers μ).

Théorème 3.12. (Herglotz) Soit c une fonction de type positif sur \mathbb{Z} alors il existe une unique mesure (positive) sur \mathbb{T} soit μ telle que:

$$c(n) = \int_{\mathbb{T}} \gamma_n(t) d\mu(t) = \widehat{\mu}(-n), \quad \forall n \in \mathbb{Z}$$

Si c est sommable, alors la mesure μ a une densité f par rapport à la mesure de Lebesgue sur \mathbb{T} donnée par la série uniformément convergente

$$f(t) = \sum_{\mathbb{Z}} c(n)\gamma_{-n}(t) = \widehat{c}(t) \quad (3.1)$$

Démonstration. Pour un entier positif N on pose

$$g_N(t) = \frac{1}{N} \sum_{n=1}^N \sum_{m=1}^N \gamma_n(-t)\gamma_m(t)c(n-m) = \sum_{k=-N+1}^{N-1} \left(1 - \frac{|k|}{N}\right) c(k)\gamma_{-k}(t)$$

Puisque c est de type positif et $\gamma_n(-t) = \overline{\gamma_n(t)}$, la fonction $g_N(t)$ est positive et si l'on note μ_N la mesure positive de densité $g_N(t)$ par rapport à la mesure de Lebesgue sur \mathbb{T} on a

$$\widehat{\mu}_N(p) = \int_{\mathbb{T}} g_N(t)\gamma_{-p}(t) dt = \sum_{k=-N+1}^{N-1} c(k)\langle \gamma_{-k}, \gamma_p \rangle = \left(1 - \frac{|p|}{N}\right) c(-p)$$

pour $|p| < N$ et $\widehat{\mu}_N(p) = 0$ pour $|p| \geq N$. Évidemment on a $\lim_N \widehat{\mu}_N(-p) = c(p)$. Par le théorème de Paul Lévy la suite $(\widehat{\mu}_N)_N$ converge étroitement vers une mesure μ , et par conséquent $\widehat{\mu}(p) = c(-p)$ pour tout p . Si $c \in \ell^1(\mathbb{Z})$ alors la série dans 3.1 converge dans $L^1(\mathbb{T}, dx)$ et on intègre terme à terme pour vérifier que $\widehat{f}(p) = c(-p) = \widehat{\mu}(p)$ \square

Théorème 3.13. (Bochner) Soit c une fonction de type positif sur \mathbb{R} et continue à l'origine. Alors il existe une unique mesure (positive, finie) μ sur \mathbb{R} telle que:

$$c(t) = \int_{\mathbb{R}} \gamma_t(x) d\mu(x) = \widehat{\mu}(-t), \quad \forall t \in \mathbb{R}$$

Si c est sommable, alors la mesure μ a une densité f par rapport à la mesure de Lebesgue sur \mathbb{R} donnée par

$$f(t) = \int_{\mathbb{R}} c(x)\gamma_{-t}(x) dx = \widehat{c}(t) \quad (3.2)$$

Démonstration. Pour T réel positif on définit la fonction positive continue et à support compact $h_T(t) = (1 - \frac{|t|}{T})^+$. Pour un entier positif N on pose

$$g_N(t) = \frac{1}{N} \int_0^N \int_0^N \gamma_x(-t)\gamma_y(t)c(x-y) dx dy = \int_{\mathbb{R}} h_N(u)c(u)\gamma_{-t}(u) du$$

Comme pour le théorème de Herglotz, on veut calculer sa transformée de Fourier et montrer qu'elle converge vers une fonction limite et appliquer le théorème de Paul Lévy. À la différence de la démonstration du théorème de Herglotz on ne peut pas appliquer le théorème de Fubini, car la fonction $t \rightarrow \gamma_{-t}(u)$ n'est pas intégrable sur \mathbb{R} . On calcule alors

$$\int_{\mathbb{R}} g_N(t)h_T(t)\gamma_u(t) dt = \int_{\mathbb{R}} h_N(v)c(v)K_T(u-v) dv$$

où K_T désigne le noyau de Féjer (définition 1.47. La fonction $v \rightarrow h_N(v)c(v)$ est continue et intégrable et donc, d'après les propriétés du noyau de Féjer (lemme 1.49),

$$\lim_{T \rightarrow \infty} \int_{\mathbb{R}} g_N(t)h_T(t)\gamma_u(t) dt = \lim_{T \rightarrow \infty} \int_{\mathbb{R}} h_N(v)c(v)K_T(u-v) dv = c(u)h_N(u)$$

En faisant $u = 0$ le théorème de convergence monotone implique

$$\int_{\mathbb{R}} g_N(t) dt = \lim_{T \rightarrow \infty} \int_{\mathbb{R}} g_N(t)h_T(t) dt = c(0)$$

et donc g_N est intégrable et son intégrale ne dépend pas de n (et vaut donc $c(0)$). Par conséquent les mesures positives μ_N de densité g_N par rapport à la mesure de Lebesgue sur \mathbb{R} ont toutes pour masse totale $c(0)$. Le théorème de Lebesgue donne maintenant

$$\widehat{\mu}_N(u) = \widehat{g}_N(u) = \lim_{T \rightarrow \infty} \int_{\mathbb{R}} g_N(t)h_T(t)\gamma_{-u}(t) dt = c(-u)h_N(-u),$$

d'où

$$\lim_{N \rightarrow \infty} \widehat{\mu}_N(u) = c(-u)$$

La fonction $u \rightarrow c(-u)$ étant continue à l'origine, le théorème de Paul Lévy permet d'affirmer qu'il existe une mesure positive μ de masse totale $c(0)$ telle que $c(-u) = \widehat{\mu}(u)$. Si c est sommable, alors les fonctions g_N sont uniformément bornées par $\|c\|_1$ et convergent ponctuellement vers f . Par conséquent si φ est une fonction continue à support compact on peut appliquer le théorème de Lebesgue, d'où

$$\mu(\varphi) = \lim_{N \rightarrow \infty} \mu_N(\varphi) = \lim \int_{\mathbb{R}} g_N(t)\varphi(t) dt = \int_{\mathbb{R}} f(t)\varphi(t) dx$$

et la preuve est complète. \square

Définition 3.14. Soit X un p.s.c. (continu en m.q. si $T = \mathbb{R}$) de fonction de covariance c_X . On dira alors que la mesure μ_X donnée par les théorèmes d'Herglotz ou de Bochner est la mesure spectrale de X .

On remarquera que l'on a $\hat{\mu}_X(t) = c_X(-t)$ et non $c_X(t)$... En fait on a choisi cette convention de telle sorte que dans les cas usuels où la fonction de covariance est sommable, on obtienne la densité spectrale par une transformation de Fourier.

Par ailleurs, si le processus $(X_t)_t$ est réel, la fonction de covariance c est réelle et donc symétrique. On a donc aussi, dans ce cas, $\hat{\mu}_X(t) = c_X(t)$. En général si on définit $\tilde{\mu}_X(A) = \mu_X(-A)$, on vérifie tout de suite que $\widehat{\tilde{\mu}_X}(t) = \hat{\mu}_X(-t)$ et donc $\widehat{\tilde{\mu}_X}(t) = c(t)$.

Ceux qui connaissent la théorie spectrale des opérateurs unitaires se convaincront facilement que dans le cas du tore, μ_X est la mesure spectrale de l'opérateur unitaire \mathcal{T} calculée sur un vecteur X_k (quel que soit $k \in \mathbb{Z}$) puisque:

$$c_X(n) = \int_{\mathbb{T}} \gamma_n(t) d\mu_X(t) = \langle \mathcal{T}^n X_k, X_k \rangle, \quad \forall n, k \in \mathbb{Z}$$

Exemple 3.15. Quelle est la mesure spectrale d'un bruit blanc de variance σ^2 ? Dans ce cas on a $c(0) = \sigma^2$, $c(k) = 0$ pour tout $k \neq 0$. La relation (3.2), donne $f(t) = \sigma^2$. Donc la mesure spectrale a une densité constante égale à σ^2 .

On a la "réciproque" suivante qui est une conséquence immédiate de 2.36 :

Remarque 3.16. Soit μ une mesure positive finie sur \mathbb{T} (resp sur \mathbb{R}) alors il existe au moins un p.s.c. de mesure spectrale μ . (On peut même le choisir gaussien).

Exemple 3.17. (P.s.c. à spectre discret) Soit $\mathbb{T} =]-\frac{1}{2}, \frac{1}{2}]$ et $-\frac{1}{2} < \lambda_1 < \dots < \lambda_m \leq \frac{1}{2}$. Soit $A(1), \dots, A(m)$ une famille de v.a. complexes, centrées, de carré intégrable, deux à deux orthogonales, avec $\text{Var}(A_j) = \sigma_j^2$. On pose

$$X_n = \sum_{j=1}^m A(j) \gamma_{\lambda_j}(n) = \sum_{j=1}^m A(j) e^{-2\pi i n \lambda_j} \quad (3.3)$$

Montrons que $(X_n)_{n \in \mathbb{Z}}$ est un p.s.c. En effet $X_n \in L^2$ et

$$\mathbb{E}[X_{k+r} \overline{X_r}] = \sum_{j, \ell=1}^m \mathbb{E}[A(j) \overline{A(\ell)}] \gamma_{\lambda_j}(k+r) \overline{\gamma_{\lambda_\ell}(r)} = \sum_{j=1}^m \sigma_j^2 \gamma_{\lambda_j}(k) := c_X(k)$$

Cette dernière quantité ne dépendant pas de r , $(X_n)_{n \in \mathbb{Z}}$ est un p.s.c. Quelle est sa mesure spectrale? Si l'on pose $\mu = \sum_{j=1}^m \sigma_j^2 \delta_{\lambda_j}$ (où δ_{λ_j} est la masse de Dirac en λ_j), alors

$$c_X(k) = \sum_{j=1}^m \sigma_j^2 \gamma_{\lambda_j}(k) = \int_{\mathbb{T}} \gamma_t(k) d\mu(t) \quad (3.4)$$

Donc μ est la mesure spectrale de X soit μ_X .

En général le processus $(X_n)_{n \in \mathbb{Z}}$ défini par (3.3) est complexe. Si on veut construire un processus réel, on supposera alors que:

$$\lambda_{m+1-j} = -\lambda_j, \quad A(-j) = \overline{A(j)}$$

On peut remarquer que la condition $A(-j) = \overline{A(j)}$ n'est pas incompatible avec l'orthogonalité de $A(-j)$ et $A(j)$ car si l'on pose $A(j) = C(j) + iD(j)$, cette condition est équivalente à $\mathbb{E}(C(j))^2 = \mathbb{E}(D(j))^2$ et $\mathbb{E}(C(j)D(j)) = 0$ (et alors $\sigma_j^2 = \sigma_{-j}^2 = \mathbb{E}(C(j))^2 + \mathbb{E}(D(j))^2$). On impose donc que la mesure spectrale μ_X soit symétrique. De plus, les relations précédentes impliquent que si 0 est l'un des valeurs λ_j , alors la v.a. $A(0)$ est réelle. Il est immédiat de vérifier que, dans ces conditions, le processus $(X_n)_{n \in \mathbb{Z}}$ est réel. En supposant, par exemple, m pair soit $m = 2r$, et en numérotant les λ et les v.a. A sous la forme $\pm\lambda_j$; $A_{\pm j}$; $j = 1, \dots, r$ on obtient

$$X_n = 2 \sum_{j=1}^r \{C(j) \cos(2\pi n \lambda_j) - D(j) \sin(2\pi n \lambda_j)\} \quad (3.5)$$

Si on reprend (3.4), la fonction de covariance de ce processus vaut donc

$$c_X(k) = 2 \sum_{j=1}^r \sigma_j^2 \cos(2\pi k \lambda_j)$$

On voit donc que la fonction de covariance c_X apparaît comme une superposition de cosinus de fréquences différentes, chacune avec un poids σ_j^2 . Les calculs effectués autour de ce modèle simple nous permettent de faire quelques considérations qui peuvent éclairer la signification intuitive des objets qui ont été introduits dans les derniers paragraphes. Puisque la fonction $t \rightarrow \cos(2\pi t \lambda_j)$ a une période égale à $t_j = \lambda_j^{-1}$, les grandes valeurs de σ_j^2 correspondent à la présence dans la fonction de covariance c_X de composantes importantes ayant une période t_j . Plus précisément, supposons que σ_ℓ^2 , soit significativement plus grand que les autres σ_j^2 . Si on pose $\sigma^2 = 2 \sum_{j=1}^r \sigma_j^2$, le quotient σ_ℓ^2 / σ^2 sera proche de 1 pour $j = \ell$ et proche de 0 sinon. En se rappelant que $\sigma^2 = c_X(0) = \text{Var}(X_n)$ pour tout n , le coefficient de corrélation entre X_n et X_{n+k} vaut

$$\rho(k) = \frac{c_X(k)}{c_X(0)} = \sum_{j=1}^r \frac{\sigma_j^2}{\sigma^2} \cos(2\pi k \lambda_j) \sim \cos(2\pi k \lambda_\ell)$$

Cette quantité qui oscille entre ± 1 indique donc une forte corrélation périodique (positive ou négative) entre X_n et X_{n+k}

De même, en considérant la représentation 3.5, le processus $(X_n)_{n \in \mathbb{Z}}$ devrait donc présenter une périodicité de période λ_ℓ^{-1} . De façon plus générale, si la mesure spectrale μ_X a une densité f_X par rapport à la mesure de Lebesgue, la relation

$$c_X(k) = \int_{\mathbb{T}} \gamma_t(k) f_X(t) dt$$

peut être vue comme une décomposition de la fonction c_X comme une superposition des exponentielles complexes γ_t , chacune avec un poids " $f_X(t) dt$ ". Les valeurs $t \neq 0$ des fréquences où f_X a un maximum indiqueront donc la présence de périodicités de l'ordre de $\frac{1}{t}$ du processus $(X_n)_{n \in \mathbb{Z}}$. Donc la recherche de périodicités dans le comportement d'un p.s.c. se fait de façon naturelle avec l'étude de la mesure spectrale.

Exemple 3.18. (P.s.c. à temps continu)

Soit U et V des variables aléatoires indépendantes, U suivant une loi uniforme sur $[-1, +1]$, V suivant une loi μ sur \mathbb{R} . Pour $t \in \mathbb{R}$ on pose $X_t = U \gamma_t(V)$. Alors X_t est

centré puisque $\mathbb{E}(X_t) = \mathbb{E}(U)\mathbb{E}(\gamma_t(V))$ et $\mathbb{E}(U) = 0$. D'autre part:

$$\mathbb{E}(X_{t+h}\bar{X}_h) = \mathbb{E}(U^2)\mathbb{E}(\gamma_{t+h}(V)\gamma_{-h}(V)) = \frac{1}{3}\mathbb{E}(\gamma_t(V))$$

On en déduit que X_t est stationnaire et que $c_X(t) = \frac{1}{3} \int \gamma_t(x) d\mu(x)$ et par conséquent la mesure spectrale μ_x est égale à $\mu/3$.

On peut aussi considérer le processus X_t défini par la formule $X_t = \exp(-2i\pi(tV + U))$, où U et V sont des variables aléatoires indépendantes, U suivant une loi uniforme sur $[0, 1]$ et V suivant une loi μ . Alors X_t est centré puisque $\mathbb{E}(X_t) = \mathbb{E}(\gamma_1(U))\mathbb{E}(\gamma_t(V))$ et $\mathbb{E}(\gamma_1(U)) = \int_0^1 \gamma_1(u) du = 0$. D'autre part:

$$\mathbb{E}(X_{t+h}\bar{X}_h) = \mathbb{E}(|\gamma_1(U)|^2)\mathbb{E}(\gamma_{t+h}(V)\gamma_{-h}(V)) = \mathbb{E}(\gamma_t(V))$$

On en déduit que X_t est stationnaire et que $c_X(t) = \int \gamma_t(x) d\mu(x)$ et par conséquent la mesure spectrale μ_x est égale à μ .

3.3 Filtrage des processus

On rappelle les dénominations introduites dans la définition 1.79:

- On dit qu'une suite $a = a_n, n \in \mathbb{Z}$ est un filtre stable si $a \in \ell^1(\mathbb{Z})$. On dit que cette suite est un filtre causal si $a_n = 0$ pour $n < 0$.
- On dit qu'une fonction $a(x), x \in \mathbb{R}$ est un filtre stable si $a \in L^1(\mathbb{R}, dx)$. On dit que cette fonction est un filtre causal si $a(t) = 0$ pour $t < 0$.

Proposition 3.19. *Soit X un p.s.c. et a un filtre stable.*

1. Si $T = \mathbb{Z}$, la série $Y = \sum_{n \in \mathbb{Z}} a_n X_n$ converge dans $L^2(\Omega, \mathbb{P})$ et \mathbb{P} -p.s. De plus $\mathbb{E}(Y) = 0$, $\|Y\|_2^2 = \sum_m \sum_n a_n \bar{a}_m c_X(n - m)$ et $Y \in H(X)$.
2. Si $T = \mathbb{R}$, l'intégrale $Y = \int_{t \in \mathbb{R}} a(t) X_t dt$ converge absolument \mathbb{P} -p.s. De plus $\mathbb{E}(Y) = 0$, $\|Y\|_2^2 = \int \int a(t) \bar{a}(u) c_X(t - u) dt du$ et $Y \in H(X)$.
3. Dans tous les cas on a $\|Y\|_2 \leq \|a\|_1 \sqrt{c(0)}$

Démonstration. (1) La convergence dans $L^2(\Omega, \mathbb{P})$ et \mathbb{P} -p.s. est une conséquence immédiate de la proposition 1.19 puisque $\|X_n\|_2^2 = c(0)$. D'autre part Y est la limite de la suite $Y_n = \sum_{k=-n}^n a_k X_k$ qui appartient à $H(X)$. Par continuité (puis bicontinuité) du produit scalaire:

$$\begin{aligned} \mathbb{E}(Y) &= \langle Y, 1 \rangle = \sum_n a_n \langle X_n, 1 \rangle = 0 \\ \|Y_n\|_2^2 &= \left\langle \sum_n a_n X_n, \sum_m a_m X_m \right\rangle = \sum_n \sum_m a_n \bar{a}_m \langle X_n, X_m \rangle \end{aligned}$$

(2) La fonction $(t, \omega) \rightarrow |a(t)X_t(\omega)|$ est $dt \otimes d\mathbb{P}$ intégrable puisque

$$\mathbb{E} \left[\int |a(t)X_t| dt \right] = \int |a(t)| \mathbb{E}(|X_t|) dt \leq \sqrt{c(0)} \|a\|_1$$

où on a utilisé l'inégalité $\|X_t\|_1 \leq \|X_t\|_2 = \sqrt{c(0)}$. On en déduit par Fubini que l'intégrale $\int |a(t)X_t| dt$ est \mathbb{P} intégrable donc \mathbb{P} p.s. finie. Ensuite

$$\mathbb{E}(Y\bar{Y}) = \mathbb{E} \left[\int a(t)X_t dt \int \overline{a(u)X_u} du \right]$$

et on vérifie que $(t, u, \omega) \rightarrow |a(t)||a(u)||X_t||X_u|$ est $dt \otimes du \otimes d\mathbb{P}$ intégrable en utilisant la majoration $\mathbb{E}(|X_t||X_u|) \leq c(0)$. En utilisant Fubini on obtient:

$$\|Y\|_2^2 = \int \int a(t)\overline{a(u)}\mathbb{E}(X_t\bar{X}_u) dt du$$

Intuitivement le fait $Y \in H(X)$ est assez évident, puisque Y "s'écrit comme une fonction des X_t ". Pour le prouver rigoureusement il suffit de montrer que si $Z \in L^2(\Omega, \mathbb{P})$ est orthogonal à tous les X_t alors Z est orthogonal à Y . Or $\langle Y, Z \rangle = \mathbb{E}(Y\bar{Z}) = \mathbb{E}(\int a(t)X_t\bar{Z} dt)$. Le théorème de Fubini donne alors $\langle Y, Z \rangle = \int a(t)\langle X_t, Z \rangle dt = 0$. En choisissant $Z = 1$ on a $\mathbb{E}(Y) = 0$. \square

Remarque 3.20. Dans le cas où X est un bruit blanc, on conserve la convergence de la série $\sum_{n \in \mathbb{Z}} a_n X_n$ si l'on a seulement $a \in \ell^2(\mathbb{Z})$, mais on n'a plus nécessairement la convergence p.s.

Le théorème suivant montre comment peut-on utiliser les filtres pour obtenir de nouveaux p.s.c. à partir d'anciens.

Théorème 3.21. Soit X un p.s.c. et a un filtre stable (on rappelle la définition $\check{a}(k) = \overline{a(-k)}$).

1. Si $T = \mathbb{Z}$, le processus $Y_n = \sum_{k \in \mathbb{Z}} a_k X_{n-k} = (a * X)_n$ est un p.s.c. de covariance $c_Y(n) = a * \check{a} * c_X(n)$. Si a est causal alors $H_n(Y) \subset H_n(X)$.
2. Si $T = \mathbb{R}$, le processus $Y_t = \int_{\mathbb{R}} a(s)X_{t-s} ds = (a * X)_t$ est un p.s.c. de covariance $c_Y(t) = a * \check{a} * c_X(t)$. Le processus Y_t est toujours continu en m.q. et si a est causal alors $H_t(Y) \subset H_t(X)$.

Démonstration. Puisque $Y_n = \sum_{r \in \mathbb{Z}} a_{n-r} X_r$, grâce à la proposition 3.19 on sait déjà que le processus Y est centré et de carré sommable. Pour montrer qu'il est stationnaire, on reprend les calculs faits ci-dessus en remplaçant $\mathbb{E}(Y_t\bar{Y}_t)$ par $\mathbb{E}(Y_{t+h}\bar{Y}_h)$. Dans le cas discret si a est causal $Y_n = \sum_{k=0}^{\infty} a_k X_{n-k}$ et il est immédiat que $Y_n \in H_n(X)$. Dans le cas continu on prouve que si Z est orthogonal à tous les X_s , $s \leq t$ alors Z est orthogonal à Y_t . Pour prouver que Y est continu en m.q. on remarque que c_Y est la convolution de $a * \check{a}$ qui est dans $L^1(\mathbb{T}, dx)$ par c_X qui est une fonction bornée, le résultat est donc une fonction uniformément continue (théorème 1.62). On sait (proposition 3.7) que la continuité de la fonction de covariance est équivalente à la continuité en m.q. di processus. \square

Remarque 3.22. On aura très souvent affaire à des p.s.c. de la forme

$$X_n = \sum_{k \in \mathbb{Z}} a_k B_{n-k} = (a * X)_n$$

où a est un filtre stable et $(B_n)_n$ est un bruit blanc de variance σ^2 . Le théorème permet tout de suite de calculer la fonction de covariance d'un tel p.s.c.:

$$c_X(k) = \sigma^2 a * \check{a}(k) = \sigma^2 \sum_{n \in \mathbb{Z}} a(n)\check{a}(k-n) = \sigma^2 \sum_{n \in \mathbb{Z}} a(n)\overline{a(n-k)}$$

Si de plus a est causal, alors pour $k \geq 0$ on a $c_X(k) = \sigma^2 \sum_{n \in \mathbb{N}} a(n+k)\overline{a(n)}$

Remarque 3.23. La convolution de deux filtres stables u et v est un filtre stable. S'ils sont de type causal, il en est de même de leur convolution. On a les relations d'associativité et commutativité : $(u * v) * X = u * (v * X) = v * (u * X) = (v * u) * X$

3.4 Représentation spectrale des p.s.c.

Théorème 3.24. Soit X un p.s.c. (continu en m.q. si $T = \mathbb{R}$) défini sur l'ensemble d'indices T et de fonction de covariance c_X . Il existe une unique isométrie notée \mathcal{W}_X entre $L^2(\mu_X)$ et $H(X)$ qui vérifie:

$$X_t = \mathcal{W}_X(\gamma_t), \quad \forall t \in T$$

De plus si a est un filtre stable et $Y = a * X$ alors:

$$Y_t = \mathcal{W}_X(\hat{a}\gamma_t), \quad \mu_Y = |\hat{a}|^2 \mu_X$$

Démonstration. La preuve est la même dans les deux cas, on se contentera donc de l'écrire dans le cas de \mathbb{R} .

(1) En reprenant les notations de la proposition 1.11 on pose $H_1 = L^2(\mathbb{R}, \mu_X)$, $H_2 = H(X)$, et F_1 (resp F_2) désigne l'ensemble des combinaisons linéaires des exponentielles complexes (resp. des X_t). On pose $\mathcal{W}_X(\gamma_t) = X_t$ que l'on "prolonge" par linéarité à F_1 . On a alors

$$\begin{aligned} \left\| \sum_{k=1}^n \lambda_k \gamma_{t_k} \right\|_{H_1}^2 &= \sum_{k=1}^n \sum_{r=1}^n \lambda_k \bar{\lambda}_r \int \gamma_{t_k - t_r} d\mu_X = \sum_{k=1}^n \sum_{r=1}^n \lambda_k \bar{\lambda}_r c_X(t_k - t_r) \\ \left\| \sum_{k=1}^n \lambda_k X_{t_k} \right\|_{H_2}^2 &= \sum_{k=1}^n \sum_{r=1}^n \lambda_k \bar{\lambda}_r \mathbb{E}(X_{t_k} \bar{X}_{t_r}) = \sum_{k=1}^n \sum_{r=1}^n \lambda_k \bar{\lambda}_r c_X(t_k - t_r) \end{aligned}$$

Ceci prouve que si une fonction dans F_1 possède deux développements alors ceux-ci ont même image par \mathcal{W}_X donc \mathcal{W}_X est bien définie sur F_1 et c'est une isométrie surjective sur F_2 d'où le résultat.

(2) Pour montrer que $Y_t = \mathcal{W}_X(\hat{a}\gamma_t)$ il suffit de prouver que pour tout X_s on a

$$\langle Y_t, X_s \rangle_{H_2} = \langle \mathcal{W}_X(\hat{a}\gamma_t), X_s \rangle_{H_2}$$

Puisque $X_s = \mathcal{W}_X(\gamma_s)$ et \mathcal{W}_X est une isométrie, ceci est équivalent à:

$$\langle Y_t, X_s \rangle_{H_2} = \langle \mathcal{W}_X(\hat{a}\gamma_t), \mathcal{W}_X(\gamma_s) \rangle_{H_2} = \langle \hat{a}\gamma_t, \gamma_s \rangle_{H_1}$$

car alors la v.a. $Y_t - \mathcal{W}_X(\hat{a}\gamma_t)$, qui appartient à $H(X)$ serait orthogonale à tout élément de $H(X)$ et donc nulle. Or on a tout de suite $\langle \hat{a}\gamma_t, \gamma_s \rangle_{H_1} = \int \hat{a}\gamma_{t-s} d\mu_X$ alors que, grâce au théorème de Fubini,

$$\begin{aligned} \langle Y_t, X_s \rangle_{H_2} &= \mathbb{E} \left[\int a(u) X_{t-u} du \bar{X}_s \right] = \int a(u) \mathbb{E}[X_{t-u} \bar{X}_s] du = \\ &= \int a(u) c_X(t-s-u) du = \int a(u) du \int \gamma_{t-s-u}(v) d\mu_X(v) = \\ &= \int \gamma_{t-s}(v) d\mu_X(v) \int a(u) \gamma_{-u}(v) du = \int \hat{a}(v) \gamma_{t-s}(v) d\mu_X(v) \end{aligned}$$

Pour le calcul de la mesure spectrale de Y on écrit:

$$c_Y(t) = \mathbb{E}(Y_{t+h}\bar{Y}_h) = \langle \mathcal{W}_X(\hat{a}\gamma_{t+h}), \mathcal{W}_X(\hat{a}\gamma_h) \rangle_{H_2} = \langle \hat{a}\gamma_{t+h}, \hat{a}\gamma_h \rangle_{H_1} = \int |\hat{a}|^2 \gamma_t d\mu_X$$

La mesure spectrale μ_Y admet donc la densité $|\hat{a}|^2$ par rapport à μ_X \square

On a, comme dans le cas déterministe (voir 1.83), un théorème d'échantillonnage:

Théorème 3.25. *Soit $(X_t)_{t \in \mathbb{R}}$ un p.s.c. continu en m.q. dont la mesure spectrale est à support dans l'intervalle $[-T, T]$ et possède une densité bornée par rapport à la mesure de Lebesgue. Alors en posant $\sigma_T(t) = (2\pi Tt)^{-1} \sin(2\pi Tt)$ on a:*

$$X_t = \sum_{n \in \mathbb{Z}} \sigma_T(t - \frac{n}{2T}) X_{\frac{n}{2T}},$$

la convergence ayant lieu dans $L^2(\Omega, \mathbb{P})$ pour tout $t \in \mathbb{R}$. On en déduit une formule d'échantillonnage de la covariance:

$$c_X(t) = \sum_{n \in \mathbb{Z}} \sigma_T(t - \frac{n}{2T}) c_X(\frac{n}{2T})$$

Démonstration. D'après la proposition 1.32, en posant $\tau = 1/(2T)$ et:

$$a_n(t) = \tau \int_{-T}^T \gamma_t(x) \gamma_{-n\tau}(x) dx = \sigma_T(t - n\tau)$$

on peut écrire l'égalité $\gamma_t = \sum a_n(t) \gamma_{n\tau}$, la série étant convergente dans $L^2([-T, T], dx)$. Or μ_X ayant une densité bornée cette série converge aussi dans $L^2([-T, T], \mu_X)$ et en appliquant l'isométrie \mathcal{W}_X on a le résultat demandé. Pour la covariance on écrit que

$$c_X(t) = \int \gamma_t d\mu_X = \sum a_n(t) \int \gamma_{n\tau} d\mu_X = \sum a_n(t) c_X(n\tau)$$

\square

On peut remarquer que la formule d'échantillonnage de la covariance pourrait être directement obtenue par le théorème d'échantillonnage déterministe.

On a vu que, si a est un filtre stable, la relation $Y_t = \mathcal{W}_X(\hat{a}\gamma_t)$ définit un nouveau p.s.c. $(Y_t)_t$. On remarque néanmoins que cette expression garde un sens si l'on remplace \hat{a} par une quelconque fonction $\varphi \in L^2(\mu_X)$. On appelle ces processus des processus filtrés généralisés. Pour ceux-ci il n'est pas possible d'obtenir des expressions en termes d'intégrales ou de séries, comme dans la Proposition 3.19. Plus précisément

Définition 3.26. *Soit X un p.s.c. (continu en m.q. si $T = \mathbb{R}$) et $\varphi \in L^2(T, \mu_X)$. On définit le processus filtré généralisé Y par la formule $Y_t = \mathcal{W}_X(\varphi\gamma_t)$. Par analogie, on appellera φ le gain du filtre généralisé.*

Proposition 3.27. *Soit Y un processus filtré généralisé de X par un filtre de gain φ . Alors Y est un p.s.c. (continu en m.q. si $T = \mathbb{R}$) de mesure spectrale $d\mu_Y = |\varphi|^2 d\mu_X$ et pour $g \in L^2(T, \mu_Y)$ on a:*

$$\mathcal{W}_Y(g) = \mathcal{W}_X(g\varphi)$$

Démonstration. D'abord il est évident que Y_t est centrée pour tout t . D'après la formule de Parseval $\langle Y_{t+s}, Y_s \rangle = \langle \mathcal{W}_X(\varphi\gamma_{t+s}), \mathcal{W}_X(\varphi\gamma_s) \rangle = \int |\varphi|^2 \gamma_t d\mu_X$ ce qui prouve à la fois que c'est un p.s.c. et que $d\mu_Y = |\varphi|^2 d\mu_X$. Si $g \in L^2(T, \mu_Y)$ alors $g\varphi \in L^2(T, \mu_X)$ et $\|g\|_{L^2(\mu_Y)} = \|\varphi g\|_{L^2(\mu_X)}$. Donc, si $g_n \rightarrow g$, $g_n\varphi \rightarrow g\varphi$ dans $L^2(T, \mu_X)$. Il suffit donc de vérifier que $\mathcal{W}_Y(g) = \mathcal{W}_X(g\varphi)$ pour $g = \gamma_t$, $t \in T$, ce qui est évident. \square

On retrouve le fait que, dans le cas de $T = \mathbb{Z}$ et d'un bruit blanc X , on peut définir $a * X$ pour $a \in \ell^2(\mathbb{Z})$ mais la série n'est pas nécessairement p.s. convergente (cf Remarque 3.23). Le filtrage généralisé permet aussi d'inverser la convolution dans certains cas, en se passant de la notion de "transformée en z " que l'on verra plus loin, mais, comme on l'a déjà fait remarquer, cette notion reste très "théorique"...

Proposition 3.28. *Soit Y un filtré généralisé de X par un filtre de gain $\varphi \in L^2(T, \mu_X)$. Alors X s'écrit comme un filtré généralisé de Y si et seulement si $\varphi \neq 0$ μ_X -p.p. Dans ce cas, X est un filtré généralisé de Y par la fonction $\psi = \frac{1}{\varphi} \mathbf{1}_{\{\varphi \neq 0\}}$.*

Démonstration. On voit tout de suite que $\psi \in L^2(T, \mu_Y)$. Si $A = \{\varphi = 0\}$ et $\mu_X(A) > 0$, alors la v.a. $\mathcal{W}_X(1_A)$ est non nulle et appartient à $H(X)$. Si elle appartenait aussi à $H(Y)$, il existerait $g \in L^2(T, \mu_Y)$ telle que $g\varphi = 1_A$, ce qui n'est pas possible. Par contre si $\mu_X(A) = 0$, soit Z le processus filtré généralisé de Y de gain ψ . Alors Z est le processus filtré généralisé de X par $\psi\varphi$. Il suffit de remarquer que cette fonction est égale à 1 μ_X -p.p. \square

Chapitre 4

L'intégrale stochastique

4.1 Processus à accroissements orthogonaux et intégrale stochastique

On introduit dans cette section l'intégrale stochastique de fonctions par rapport à un processus. Celle ci est parfois appelée "petite" intégrale stochastique car l'on n'intègre que des fonctions et non des processus, comme dans la "vraie" intégrale stochastique.

Définition 4.1. *Un processus de carré sommable $\{Z_t; t \in \mathbb{R}\}$ est dit à accroissements orthogonaux (en abrégé p.a.o.) si pour tout système $(t_1 < t_2 < \dots < t_n)$ les variables aléatoires $(Z_{t_1} - Z_{t_2}), \dots, (Z_{t_n} - Z_{t_{n-1}})$ sont orthogonales dans $L^2(\Omega, \mathbb{P})$. On dit que le processus Z est continu en moyenne quadratique (en abrégé continu en m.q.) si l'application $t \mapsto Z_t$ de \mathbb{R}^+ dans $L^2(\Omega, \mathbb{P})$ est continue.*

Dans la plupart des cas, les processus considérés sont réels, définis seulement pour $t \in \mathbb{R}^+$ et sont nuls pour $t = 0$. Pour un tel processus $\{X_t; t \geq 0\}$ on définit $H(X)$ comme la fermeture dans $L^2(\Omega, \mathbb{P})$ des combinaisons linéaires à coefficients réels des variables aléatoires $\{X_t; t \geq 0\}$. On peut associer une "pseudo mesure spectrale" à un p.a.o (mesure structurelle) qui n'est pas nécessairement finie. Cette mesure structurelle est alors définie sur les réels positifs.

Théorème 4.2. *Soit $\{Z_t; t \in \mathbb{R}^+\}$ un p.a.o. continu à droite en m.q. et tel que $Z_0 = 0$.*

1. *La formule $\mu_Z([0, t]) = \|Z_t\|_2^2$ définit une mesure de Radon positive sur \mathbb{R}^+ appelée mesure structurelle de Z , vérifiant $\mu_Z(\{0\}) = 0$ et $\mu_Z(]t, t+h]) = \|(Z_{t+h} - Z_t)\|_2^2$.*
2. *Il existe une unique isométrie linéaire et surjective de $L^2(\mathbb{R}^+, \mu_Z)$ sur $H(Z)$, notée $\int \varphi(t) dZ_t$, prolongeant l'application $]s, t] \mapsto Z_t - Z_s$. En particulier on a la relation de Parseval:*

$$\int \varphi_1(t)\varphi_2(t) d\mu_Z(t) = \mathbb{E}\left(\left(\int \varphi_1(t) dZ_t\right)\left(\int \varphi_2(t) dZ_t\right)\right)$$

3. *Si le processus Z_t est centré alors $\mathbb{E}\{\int \varphi(t) dZ_t\} = 0$*
4. *L'intégrale stochastique $\int \varphi(t) dZ_t$ est l'unique élément Y de $H(Z)$ vérifiant $\mathbb{E}(Y Z_t) = \int_{[0, t]} \varphi(x) d\mu_Z(x)$ pour tout $t \in \mathbb{R}^+$.*

Démonstration. (1) On a

$$\mu_Z([0, t+h]) = \|Z_{t+h}\|_2^2 = \|(Z_{t+h} - Z_t) + (Z_t - Z_0)\|_2^2 = \|(Z_{t+h} - Z_t)\|_2^2 + \|Z_t\|_2^2$$

ce qui prouve que $t \rightarrow \mu_Z([0, t])$ est croissante et continue à droite. Par conséquent μ est une mesure positive sur \mathbb{R}^+ . De plus, $\mu_Z(\{0\}) = \lim_{t \downarrow 0} \mu_Z([0, t]) = \lim_{t \downarrow 0} \|Z_t - Z_0\|_2^2 = 0$.

(2)

On pose $H_1 = L^2(\mathbb{R}^+, \mu_Z)$, $H_2 = H(Z)$, et F_1 (resp F_2) désigne l'ensemble des combinaisons linéaires des indicatrices d'ensembles de type $]s, t]$ (resp. des Z_t). Alors F_1 (resp F_2) est dense dans H_1 (resp. H_2). On définit alors l'application Ψ de F_1 dans F_2 par:

$$\Psi\left(\sum_{k=1}^n \lambda_k \mathbf{1}_{\{]s_k, t_k]\}}\right) = \sum_{k=1}^n \lambda_k (Z_{t_k} - Z_{s_k})$$

On a

$$\begin{aligned} \left\| \sum_{k=1}^n \lambda_k \mathbf{1}_{\{]s_k, t_k]\}} \right\|_{H_1}^2 &= \sum_{k=1}^n \sum_{r=1}^n \lambda_k \lambda_r \mu_Z(]s_k, t_k] \cap]s_r, t_r]) \\ \left\| \sum_{k=1}^n \lambda_k (Z_{t_k} - Z_{s_k}) \right\|_{H_2}^2 &= \sum_{k=1}^n \sum_{r=1}^n \lambda_k \lambda_r \mathbb{E}((Z_{t_k} - Z_{s_k})(Z_{t_r} - Z_{s_r})) \end{aligned}$$

Il ne reste plus qu'à vérifier que $\mu_Z(]s_k, t_k] \cap]s_r, t_r]) = \mathbb{E}((Z_{t_k} - Z_{s_k})(Z_{t_r} - Z_{s_r}))$, ce qui résulte de l'orthogonalité des accroissements de Z . Cette relation d'isométrie montre que si une fonction dans F_1 possède deux développements distincts, alors ceux-ci ont même image par Ψ , donc Ψ est bien définie sur F_1 et c'est une isométrie surjective sur F_2 d'où le résultat, par le théorème d'extension des isométries.

(3)

Si Z_t est centré alors $H(Z)$ est formé de v.a. centrées car tout élément de $H_2 = H(Z)$ est limite dans L^2 et donc aussi dans L^1 d'une suite d'éléments de F_2 qui sont manifestement centrés.

(4)

Soit Y et U dans $H(Z)$ alors $Y = U$ si et seulement si $\langle Y, Z_t \rangle = \langle U, Z_t \rangle$ pour tout $t \geq 0$. Or en prenant les produits scalaires dans $L^2(\Omega, \mathbb{P})$ puis $L^2(\mathbb{R}^+, \mu_Z)$ on a:

$$\left\langle \int \varphi(s) dZ_s, Z_t \right\rangle = \left\langle \varphi, \mathbf{1}_{[0, t]} \right\rangle = \int_0^t \varphi(x) d\mu_Z(x)$$

□

Définition 4.3. Pour $t \geq 0$ et $\varphi \in L_{loc}^2(\mathbb{R}^+, \mu_Z)$, on posera

$$\int_0^t \varphi(u) dZ_u = \int \mathbf{1}_{\{]0, t]\}}(u) \varphi(u) dZ_u$$

On peut construire d'autres p.a.o à partir d'un p.a.o. En effet:

Proposition 4.4. Soit $\{Z_t; t \in \mathbb{R}^+\}$ un p.a.o. continu à droite en m.q., de mesure structurelle μ_Z et $\varphi \in L_{loc}^2(\mathbb{R}^+, \mu_Z)$. On pose $Y_t = \int_0^t \varphi(u) dZ_u$. Alors le processus Y_t est un p.a.o., continu à droite en m.q., de mesure structurelle $|\varphi|^2 \mu_Z$ et pour toute fonction $f \in L^2(\mathbb{R}^+, \mu_Y)$ on a:

$$\int f(t) dY_t = \int f(t) \varphi(t) dZ_t$$

Démonstration. Pour $s > t$ on a $Y_t - Y_s = \int \mathbf{1}_{\llbracket s, t \rrbracket}(u) \varphi(u) dZ_u$ et le fait que Y_t soit un p.a.o résulte de la relation de Parseval. Pour la continuité à droite on remarque que $\|Y_{t+h} - Y_t\|_2^2 = \int \mathbf{1}_{\llbracket t, t+h \rrbracket}(u) \varphi^2(u) d\mu_Z(u)$ et il suffit d'utiliser le théorème de Lebesgue. La même égalité définit la mesure spectrale de Y . Enfin la relation $\int f(t) dY_t = \int f(t) \varphi(t) dZ_t$ est vraie pour toute fonction f qui est une combinaison linéaire d'indicatrices et on passe à la limite dans $L^2(\mathbb{R}^+, \mu_Y)$. \square

4.2 L'intégrale stochastique et les processus stationnaires

Soit Z_t , $t \in \mathbb{R}$, un p.a.o. continu en m.q. à valeurs complexes. On peut définir sa mesure structurelle μ_Z sur \mathbb{R} par la formule $\mu_Z(\llbracket s, t \rrbracket) = \mathbb{E}(|Z_t - Z_s|^2)$ ainsi que l'intégrale stochastique associée en recopiant la construction de la section précédente à condition de considérer que $H(Z)$ est alors la fermeture dans $L^2(\Omega, \mathbb{P})$ des combinaisons linéaires à coefficients complexes des variables aléatoires $\{Z_t; t \in \mathbb{R}\}$.

Soit X un p.s.c. continu en m.q. La proposition suivante montre qu'on peut lui associer un p.a.o Z , continu en m.q. La mesure structurelle de Z est alors égale à la mesure spectrale de X , elle est donc finie.

Proposition 4.5. *Soit $\{X_t; t \in \mathbb{R}\}$ un p.s.c continu en m.q. et soit \mathcal{W}_X l'isomorphisme entre $L^2(\mu_X)$ et $H(X)$ défini en 3.24. En posant $\mathcal{W}_X(B) = \mathcal{W}_X(\mathbf{1}_B)$, la formule $Z_t = \mathcal{W}_X(\llbracket -\infty, t \rrbracket)$ définit un p.a.o centré, continu à droite en m.q. et l'on a $\|Z_t - Z_s\|_2^2 = \mu_X(\llbracket s, t \rrbracket) = \mu_Z(\llbracket s, t \rrbracket)$. Dans la suite, on notera Z^X ce p.a.o déduit du p.s.c X .*

Démonstration. Soit $(t_1 < t_2 < \dots < t_n)$ alors $Z_{t_k} - Z_{t_{k-1}} = \mathcal{W}_X(\llbracket t_{k-1}, t_k \rrbracket)$ pour $k = 2, \dots, n$ et l'orthogonalité dans $L^2(\Omega, \mathbb{P})$ est une conséquence de l'orthogonalité des indicatrices d'intervalles $\llbracket t_{k-1}, t_k \rrbracket$ dans $L^2(\mathbb{R}, \mu_X)$. De plus $\|Z_{t+h} - Z_t\|_2^2 = \mu_X(\llbracket t, t+h \rrbracket)$ et la continuité à droite de $t \rightarrow Z_t$ est une conséquence du théorème de Lebesgue. \square

Réciproquement, on peut construire un p.s.c. à partir d'un p.a.o:

Proposition 4.6. *Soit $\{Z_t; t \in \mathbb{R}^+\}$ un p.a.o. centré, continu à droite en m.q. et de mesure structurelle μ_Z finie. Alors la formule $X_t = \int \gamma_t(u) dZ_u$ définit un p.s.c. de mesure spectrale μ_Z .*

Démonstration. Puisque μ_Z est finie alors les exponentielles complexes sont de carré intégrable. La formule de Parseval permet d'écrire:

$$\mathbb{E}(X_{t+s} \bar{X}_s) = \int \gamma_{t+s}(u) \bar{\gamma}_s(u) d\mu_Z(u) = \int \gamma_t(u) d\mu_Z(u)$$

ce qui prouve que X_t est stationnaire et que sa mesure spectrale est égale à μ_Z . \square

On peut combiner les deux résultats précédents pour arriver à la représentation dite de Cramer-Khintchine, qui permet d'écrire un p.s.c. comme une "sorte" de transformée de Fourier:

Proposition 4.7. *Soit $\{X_t; t \in \mathbb{R}\}$ un p.s.c continu en m.q. et soit Z_t^X le p.a.o associé par la proposition 4.5. Alors pour tout $f, g \in L^2(\mu_X) \equiv L^2(\mu_Z)$ on a:*

1. $\int f(x) dZ_x^X = \mathcal{W}_X(f)$
2. En particulier, $X_t = \int \gamma_t(x) dZ_x^X$, (représentation de Cramer-Khintchine)

Démonstration.

(1) Pour $f(x) = \mathbf{1}_{\llbracket u, v \rrbracket}(x)$ on a

$$\int f(x) dZ_x^X = Z_v^X - Z_u^X = \mathcal{W}_X(\llbracket -\infty, v \rrbracket) - \mathcal{W}_X(\llbracket -\infty, u \rrbracket) \geq \mathcal{W}_X(\llbracket u, v \rrbracket) = \mathcal{W}_X(f)$$

La même relation subsiste si f est une combinaison linéaire d'intervalles semi-ouverts. Par conséquent les deux isométries "intégrale de Wiener" et \mathcal{W}_X coïncident sur un sous espace dense de $L^2(\mu_X) \equiv L^2(\mu_Z)$ et donc partout.

(2) Ceci résulte de la formule précédente, en prenant $f(x) = \gamma_t(x)$. \square

Ceci prouve que, dans le cas d'un p.a.o. déduit d'un p.s.c., les isométries définies par l'intégrale de Wiener et \mathcal{W}_X sont identiques. C'est pourquoi on remplace souvent cette isométrie \mathcal{W}_X par la notation intégrale correspondante.

Corollaire 4.8. Soit $\{X_t; t \in \mathbb{R}\}$ un p.s.c continu en m.q. et soit $h \in L^1(\mathbb{R}, dx)$ un filtre stable. alors le p.s.c. $Y = h * X$ admet la représentation

$$Y_t = \int \gamma_t(x) \widehat{h}(x) dZ_x^X.$$

De même, pour $\varphi \in L^2(\mathbb{R}, \mu_X)$, on peut considérer le processus filtré généralisé de X soit $Y_t = \int \varphi \gamma_t dZ^X$. Alors pour $g \in L^2(\mathbb{R}, \mu_Y)$ on a $\int g dZ^Y = \int g \varphi dZ^X$. De plus, on peut écrire $X_t = \int \psi \gamma_t dZ^Y$ si et seulement si $\varphi \neq 0$ μ_X p.p et alors $\psi = \frac{1}{\varphi} \mathbf{1}_{\{\varphi \neq 0\}}$.

Démonstration. On sait d'après 3.24 que $Y_t = \mathcal{W}_X(\gamma_t \widehat{h})$. Il suffit ensuite de traduire les propositions 3.27 et 3.28. \square

Les énoncés précédents traitaient les cas des processus stationnaires indexés par \mathbb{R} . Les mêmes arguments permettent de produire des résultats similaires pour un processus stationnaire $(X_n)_{n \in \mathbb{Z}}$. Dans la suite on va représenter le tore \mathbb{T} avec l'intervalle $]0, 1[$.

Proposition 4.9. Soit $(Z_t)_{t \in \mathbb{T}}$ un p.a.o. centré, continu à droite en m.q. et de mesure structurelle μ_Z finie. Alors la formule $X_k = \int_{\mathbb{T}} \gamma_k(u) dZ_u$ définit un p.s.c. de mesure spectrale μ_Z .

Proposition 4.10. Soit $(X_n)_{n \in \mathbb{Z}}$ un p.s.c. continu en m.q. et \mathcal{W}_X l'isomorphisme entre $L^2(\mu_X)$ et $H(X)$ défini dans le Théorème 3.24.

1. La formule $Z_t = \mathcal{W}_X(\mathbf{1}_{]0, t]})$ définit un p.a.o. centré, continu à droite en m.q. et on a $\|Z_t - Z_s\|_2^2 = \mu_X(\llbracket s, t \rrbracket) = \mu_Z(\llbracket s, t \rrbracket)$. On notera Z^X ce p.a.o. associé au p.s.c. X .
2. Pour toute $f \in L^2(\mu_X) \equiv L^2(\mu_Z)$ on a $\int_{\mathbb{T}} f(u) dZ_u^X = \mathcal{W}_X(f)$. En particulier (représentation de Cramer-Khintchine)

$$X_n = \int_{\mathbb{T}} \gamma_n(u) dZ_u^X \tag{4.1}$$

Dans le prochain paragraphe on va introduire des exemples classiques de p.a.o. Mais d'abord on va traiter un exemple particulièrement simple, ce qui permettra, entre autre, de comprendre certains aspects intuitifs des théorèmes de représentation que l'on vient de voir.

Exemple 4.11. On reprend ici l'exemple de p.s.c. étudié en 3.17 et l'on veut déterminer le p.a.o. Z^X associé à X . Si on pose, pour $-\frac{1}{2} < t \leq \frac{1}{2}$

$$Z_t = \sum_{\lambda_j \leq t} A(j)$$

alors Z est un p.a.o.: si $s < u < t$, on a $Z_t - Z_u$ =somme des $A(j)$ tels que $u < \lambda_j \leq t$ alors que $Z_u - Z_s$ =somme des $A(\ell)$ tels que $s < \lambda_\ell \leq u$, et comme les v.a. $A(j)$ sont deux à deux orthogonales, il en est de même pour $Z_t - Z_u$ et $Z_u - Z_s$. Le processus $(Z_t)_t$ a un comportement particulièrement simple: il est constant entre deux valeurs λ_j et λ_{j+1} et en λ_j il fait un saut "d'amplitude" $A(j)$. La trajectoire $t \rightarrow Z_t(\omega)$ est donc constante dans les intervalles $[\lambda_j, \lambda_{j+1}[$ et dans cette intervalle prend la valeur $A(1) + \dots + A(j)$. Il est clair aussi que

$$\int_{\mathbb{T}} f(t) dZ_t = \sum_{j=1}^n A(j) f(\lambda_j)$$

d'où, en remplaçant f par γ_k ,

$$\int_{\mathbb{T}} \gamma_k(t) dZ_t = \sum_{j=1}^n A(j) \gamma_k(\lambda_j) = \sum_{j=1}^n A(j) \gamma_{\lambda_j}(k) = X_k \quad (4.2)$$

et donc $Z = Z^X$.

La représentation (4.2) a une interprétation intuitive simple: le processus $(X_n)_{n \in \mathbb{Z}}$ apparaît comme une superposition des exponentielles complexes γ_{λ_j} avec les poids aléatoires $A(j)$, qui ont une variance égale à $\sigma_j^2 = \mu_X(\lambda_j)$. En général la représentation de Cramèr-Khintchine (4.1) peut être vue comme une décomposition de $(X_n)_{n \in \mathbb{Z}}$ dans une "somme" des exponentielles complexes γ_u , chacune avec un poids aléatoire de variance $d\mu_X(u)$, ces poids étant non corrélés entre eux.

4.3 Les exemples fondamentaux

Les deux exemples les plus importants de p.a.o. sont le mouvement brownien et le processus de Poisson.

On dit qu'un processus $(Z_t)_{t \in T}$ est un *processus à accroissements indépendants* (p.a.i.) si et seulement si pour tout système $(t_1 < t_2 < \dots < t_n)$ les v.a. $(Z_{t_1} - Z_{t_2}), \dots, (Z_{t_n} - Z_{t_{n-1}})$ sont indépendantes. Il est clair qu'un p.a.i. est un p.a.o. si de plus il est centré (ou du moins si ses accroissements sont centrés) et de carré intégrable.

Les deux exemples les plus importants de p.a.o. (en fait des processus à accroissements indépendants (p.a.i.)) sont le mouvement brownien et le processus de Poisson.

4.3.1 Le mouvement brownien.

Définition 4.12. On appelle *mouvement brownien* un processus réel $B_t, t \geq 0$, vérifiant:

- $B(0) = 0$ et l'application $t \rightarrow B_t(\omega)$ est continue pour presque tout ω
- Le processus B_t est un processus à accroissements indépendants tel que pour $h > 0$, la loi de $(B_{t+h} - B_t)$ soit une loi normale centrée de variance h

Théorème 4.13. Soit B_t un mouvement brownien.

1. B_t est un p.a.o. continu en m.q., de fonction de covariance $K(s, t) = s \wedge t$ et de mesure structurelle la mesure de Lebesgue sur \mathbb{R}^+ .
2. Soit $\varphi \in L^2_{loc}(\mathbb{R}^+, dx)$ et V_0 une variable aléatoire gaussienne, indépendante du processus B_t . On pose $V_t = V_0 + \int_0^t \varphi(u) dB_u$. Alors V_t est un processus gaussien et pour $t \geq 0$ on a $\text{Cov}(V_{t+s}, V_s) = \text{Var}(V_0) + \int_0^s \varphi^2(x) dx$

Démonstration. (1)

B_t est un p.a.i. centré, donc un p.a.o. De plus $\|B_{t+h} - B_t\|^2 = |h|$ d'où la continuité en m.q. et le fait que la mesure structurelle est égale à la mesure de Lebesgue. Le processus étant centré on a par exemple pour $s > t$:

$$K(s, t) = \mathbb{E}(B_s B_t) = \langle B_s - B_t, B_t \rangle + \langle B_t, B_t \rangle = t$$

(2)

Soit $t_1 < t_2 < \dots < t_n$. Il faut prouver que $X = \sum_{k=1}^n \lambda_k V_{t_k}$ suit une loi gaussienne. Or on peut écrire en posant $t_0 = 0$

$$X = a_0 V_0 + \sum_{k=0}^n a_k \int_{t_k}^{t_{k+1}} \varphi(s) dB_s = a_0 V_0 + \sum_{k=0}^n a_k W_k$$

et il suffit de vérifier que les variables aléatoires figurant dans cette expression sont gaussiennes et indépendantes. Or V_0 est gaussienne et les autres sont dans $H(B)$ qui est formé de variables gaussiennes. D'après la proposition 2.7, V_0 est indépendant de toute v.a. dans $H(B)$ et donc des W_k . Il reste ensuite à vérifier que les variables aléatoires W_k sont indépendantes. C'est évident si l'on remplace la fonction φ par une combinaison linéaire d'indicatrices d'ensembles et on utilise à nouveau 2.7

Pour le calcul de la covariance (qui ne suppose pas que V_0 soit gaussienne) on écrit:

$$\mathbb{E}(V_{t+s} V_s) = \mathbb{E} \left(\left\{ (V_0 - \mathbb{E}(V_0)) + \int_0^{t+s} \varphi(u) dB_u \right\} \left\{ (V_0 - \mathbb{E}(V_0)) + \int_0^s \varphi(v) dB_v \right\} \right)$$

Or si V_0 est indépendante de B_t on a $\mathbb{E}((V_0 - \mathbb{E}(V_0))Z) = 0$ pour tout élément $Z \in H(B)$. On en déduit la formule proposée par application de la relation de Parseval. \square

L'intégrale stochastique $\int \varphi(t) dB_t$ est souvent appelée intégrale stochastique par rapport à un "bruit blanc" car la mesure structurelle de B est la mesure de Lebesgue sur \mathbb{R}^+ alors que la mesure spectrale d'un bruit blanc normalisé est la mesure de Lebesgue sur $[0, 1[$.

On peut obtenir une formule d'intégration par parties pour un p.a.o. construit à partir du Brownien comme indiqué dans la proposition 4.4:

Proposition 4.14. Soit $\{B_t; t \in \mathbb{R}^+\}$ un mouvement brownien, $\varphi(x) \in L^2_{loc}(\mathbb{R}^+, dx)$ et $Y_t = \int_0^t \varphi(u) dB_u$. Pour $f \in C^1(\mathbb{R}^+)$, on a la formule d'intégration par parties:

$$\int_a^b f(u) dY_u + \int_a^b f'(u) Y_u du = f(b)Y_b - f(a)Y_a$$

Démonstration. Il suffit de prouver la formule pour $a = 0$, et on pose $b = t$. Tout revient à prouver que le terme $W_t = f(t)Y_t - \int_0^t f'(u)Y_u du$ est bien égal à l'intégrale stochastique $\int_0^t f(u) dY_u = \int_0^t f(u)\varphi(u) dB_u$. En reprenant la caractérisation donnée dans 4.2, et en utilisant la relation de Parseval, ceci est équivalent à l'égalité

$$\mathbb{E}(B_s W_t) = \mathbb{E}(B_s \int_0^t f(u)\varphi(u) dB_u) = \int_0^{t \wedge s} f(u)\varphi(u) du$$

pour tout $s \geq 0$. Or

$$\mathbb{E}(B_s W_t) = f(t)\mathbb{E}(B_s Y_t) - \int_0^t f'(u)\mathbb{E}(B_s Y_u) du$$

en justifiant l'application du théorème de Fubini dans le dernier terme écrit. Par application de la relation de Parseval on obtient $\mathbb{E}(B_s Y_t) = \int_0^{t \wedge s} \varphi(u) du$ et par conséquent tout revient à vérifier que:

$$f(t) \int_0^{t \wedge s} \varphi(u) du - \int_0^t f'(u) \left(\int_0^{u \wedge s} \varphi(v) dv \right) du = \int_0^{t \wedge s} f(u)\varphi(u) du$$

Ceci résulte d'une simple application de la formule d'intégration par parties usuelle, en distinguant les cas $s \geq t$ et $s \leq t$. \square

On peut par exemple déduire de ce théorème que pour f de classe C^1 , la variable aléatoire (gaussienne) $\int_0^t f'(u)B_u du$ a pour variance $\int_0^t (f(1) - f(u))^2 du$.

On peut généraliser la proposition 4.14:

Proposition 4.15. Soit $\{B_t; t \in \mathbb{R}^+\}$ un mouvement brownien, $\varphi(x) \in L_{loc}^2(\mathbb{R}^+, dx)$, $H(t)$ un processus stochastique à trajectoires continues et $Y_t = Y_0 + \int_0^t H(u) du + \int_0^t \varphi(u) dB_u$. Pour $f \in C^1(\mathbb{R}^+)$ on a la formule d'intégration par parties:

$$\int_a^b f(u) dY_u + \int_a^b f'(u)Y_u du = f(b)Y_b - f(a)Y_a$$

où dY_u signifie $H(u)du + \varphi(u) dB_u$.

Démonstration. Il suffit d'appliquer la formule d'intégration par parties usuelle à la fonction $h(t) = Y_0 + \int_0^t H(u) du$ et la formule précédente au processus $W_t = \int_0^t \varphi(u) dB_u$. \square

Un exemple classique d'utilisation de l'intégrale stochastique est la résolution de l'équation de Langevin $dV_t + bV_t dt = \sigma dB_t$. Ce genre d'équation perturbée se rencontre fréquemment en physique.

Théorème 4.16. Pour V_0 donné, l'équation de Langevin possède une unique solution appelée processus d'Ornstein Uhlenbeck donnée par:

$$V_t = e^{-bt} \left(V_0 + \sigma \int_0^t e^{bs} dB_s \right)$$

Si V_0 est une variable aléatoire indépendante du mouvement brownien B_t et de variance v on a:

1. V_t est un processus du second ordre d'espérance $\mathbb{E}(V_t) = e^{-bt}\mathbb{E}(V_0)$ et de covariance

$$K(s+t, s) = \left(\frac{\sigma^2}{2b}\right) \exp(-tb) + \left(v - \frac{\sigma^2}{2b}\right) \exp(-2bs - bt)$$

En particulier, si V_0 est centré, et si $v = \sigma^2/(2b)$, (ce qui implique $b > 0$), le processus V_t est stationnaire centré et sa mesure spectrale est une loi de Cauchy.

2. Si V_0 est gaussienne alors V_t est un processus gaussien.

Démonstration. En fait l'équation de Langevin doit s'interpréter comme une équation intégrale:

$$V_t = V_0 - \int_0^t bV_u du + \sigma B_t$$

$V(t)$ est donc un processus continu, et en appliquant la proposition 4.15 avec $f(t) = e^{bt}$ on obtient:

$$e^{bt}V_t - V_0 = \int_0^t e^{bu}(-bV(u) du + \sigma dB_u) + \int_0^t be^{bu}V_u du = \sigma \int_0^t e^{bu} dB_u$$

On trouve la formule annoncée en revenant à la variable V .

(1)

Il suffit de reprendre les calculs de 4.13. On a alors pour $t \geq 0$:

$$\text{Cov}(V_{t+s}, V_s) = e^{-bt-2bs} \left(v + \sigma^2 \int_0^s e^{2bu} du \right)$$

d'où la formule proposée. Si $v = \sigma^2/2b$ alors $b > 0$ et $c_V(t) = \frac{\sigma^2}{2b} \exp(-|t|b)$ qui est bien la transformée de Fourier d'une loi de Cauchy (à un facteur multiplicatif près...).

(2)

C'est une conséquence directe de 4.13. □

On peut aussi voir l'équation de Langevin comme un calcul "d'intérêts composés" de taux $\tau = -b > 0$ et pour lequel le capital est perturbé par un bruit blanc σdB_t . On prend un capital initial déterministe V_0 et soit $W_t = (V_t - V_0 e^{t\tau}) / (V_0 e^{t\tau}) = (\sigma \int_0^t e^{-s\tau} dB_s) / V_0$ la fluctuation relative autour de la valeur non perturbée. Lorsque $t \rightarrow \infty$, la variance de W_t converge vers $\sigma^2 / (2\tau V_0^2)$. On considère aussi des modèles du genre $dV_t = (-b + \sigma dB_t)V_t$ avec taux d'intérêt perturbé, mais la solution nécessite alors l'introduction de la "vraie" intégrale stochastique qui définit l'intégrale des processus et non plus seulement des fonctions.

On remarquera que la notation dB_s ne représente pas du tout une intégration par rapport à une mesure (aléatoire) sur \mathbb{R}^+ . En effet, on peut montrer que les trajectoires du mouvement brownien sont presque sûrement à variation totale infinie sur tout intervalle et ne sauraient donc être considérées comme la fonction de répartition d'une mesure. Ceci est en contraste avec le cas du processus de Poisson ci dessous, pour lequel la notation dN_s désigne bel et bien une mesure (aléatoire, ponctuelle).

On peut définir le p.a.i. centré B_t pour $t \in \mathbb{R}$ de la façon suivante: on considère, pour $t \in \mathbb{R}^+$, deux mouvements browniens indépendants $B_t^{(1)}$ et $B_t^{(2)}$, et on pose $B_t = B_t^{(1)}$ si $t \geq 0$ et $B_t = B_{-t}^{(2)}$ si $t \leq 0$. Si l'on revient aux considérations ci dessus, on voit que

le p.a.o B_t ne peut être déduit d'un p.s.c puisque sa mesure structurelle est la mesure de Lebesgue sur \mathbb{R} et n'est donc pas finie. Par contre, la transformée de Fourier, au sens des distributions de la mesure de Lebesgue sur \mathbb{R} est égale à la masse de Dirac en 0 soit δ_0 . On peut alors se demander s'il serait possible de généraliser la notion de p.s.c de telle sorte que cette mesure δ_0 puisse apparaître comme une "distribution" de covariance. Cette approche a été développée par de nombreux auteurs et consiste à considérer l'application $t \rightarrow X_t(\omega)$ non plus comme une fonction (aléatoire) mais comme une distribution (aléatoire). Ceci permet alors de donner un sens mathématique à des filtres du genre "dérivation" ou "valeur principale" qui ne peuvent s'interpréter comme des convolutions par des fonctions. Dans le cas de processus sur \mathbb{R}^+ , ceci permet aussi d'utiliser toute la puissance du calcul symbolique dans $\mathcal{D}(\mathbb{R}^+)$.

4.3.2 Le processus de Poisson.

Définition 4.17. Soit $Y_n; n \geq 1$ une suite de variables aléatoires indépendantes de loi exponentielle de paramètre λ . Pour $n \geq 1$, on pose $T_n = Y_1 + Y_2 + \dots + Y_n$. Le processus de Poisson N_t est alors défini par

$$N_t = \sum_{n \geq 1} \mathbf{1}_{[0,t]}(T_n)$$

Pour une fonction f mesurable positive sur \mathbb{R}^+ on pose $N(f) = \sum_{n \geq 1} f(T_n)$ et si f est l'indicatrice d'un borélien A on écrira $N(A)$ plutôt que $N(\mathbf{1}_{\{A\}})$.

Si on interprète les instants T_n comme les instants où un certain événement se produit, la v.a. N_t compte le nombre d'événements qui ont eu lieu dans l'intervalle $[0, t]$. La proposition suivante est prouvée dans nombre de cours de probabilité:

Proposition 4.18. Le processus N_t est un p.a.i. (non centré) et la loi de $N_{t+h} - N_h$ est une loi de Poisson de paramètre λt . La fonction de covariance est donnée par $K(s, t) = \lambda(s \wedge t)$. Plus généralement si A_1, A_2, \dots, A_r sont des boréliens bornés disjoints, les variables aléatoires $N(A_1), N(A_2), \dots, N(A_r)$ sont indépendantes et suivent des lois de Poisson de paramètres respectifs $\lambda|A_1|, \lambda|A_2|, \dots, \lambda|A_r|$ où $|A|$ désigne la mesure de Lebesgue de A .

La quantité $N(f)$ est aussi quelquefois notée $\int f(t) dN_t$, le terme dN_t faisant référence à la mesure ponctuelle $\sum_{n \geq 1} \delta_{T_n}$ qui met une masse unité à chaque saut de la trajectoire N_t considérée comme une fonction de répartition.

Proposition 4.19. Soit N_t un processus de Poisson de paramètre λ .

1. Soit f une fonction positive de $L^1(\mathbb{R}^+, dt)$ alors

$$\mathbb{E}(N(f)) = \lambda \int_0^{+\infty} f(t) dt$$

2. Soit f et g des fonctions positives dans $L^2(\mathbb{R}^+, dt)$, alors

$$\text{Cov}(N(f), N(g)) = \lambda \int_0^{+\infty} f(t)g(t) dt$$

Démonstration. On établit la première formule pour $f = \mathbf{1}_{\{A\}}$ avec A de mesure finie et on passe aux limites croissantes de combinaisons linéaires finies. Pour la seconde, on raisonne de même avec $f = \mathbf{1}_{\{A\}}$ et $g = \mathbf{1}_{\{B\}}$ et on vérifie qu'il suffit d'établir cette formule dans les cas A et B disjoints et $A = B$. \square

Si l'on veut utiliser la construction de l'intégrale par rapport à un p.a.o., on doit considérer le processus centré $\widetilde{N}_t = N_t - \lambda t$ (sinon la propriété de p.a.i. n'implique pas celle de p.a.o. . .) qui a pour mesure structurelle λdt sur \mathbb{R}^+ et l'on a:

Proposition 4.20. Soit φ une fonction positive de $L^2(\mathbb{R}^+, dt)$:

$$N(\varphi) = \int \varphi(t) d\widetilde{N}_t + \lambda \int \varphi(t) dt$$

Supposons par exemple que des photons arrivent sur une plaque aux instants T_n . Un photon arrivé à l'instant u induit une luminosité à l'instant $t > u$ donnée par $g(t - u)$, g étant une fonction positive, décroissante, nulle sur \mathbb{R}^- . Alors la luminosité totale à l'instant t est:

$$L_t = \sum_{n \leq N_t} g(t - T_n) = N(\mathbf{1}_{\{[0,t]\}}(u)g(t - u)) = \int_0^t g(t - u) d\widetilde{N}_u + \lambda \int_0^t g(t - u) du$$

L'intégrale par rapport à \widetilde{N}_u correspond à la partie purement aléatoire de L_t ("shot noise") tandis que l'intégrale de Lebesgue est appelée "luminosité continue". En utilisant 4.19, on voit que pour t et h positifs:

$$\mathbb{E}(L_t) = \lambda \int_0^t g(t - u) du, \quad \text{Cov}(L_t, L_{t+h}) = \lambda \int_0^t g(t - u)g(t + h - u) du$$

En choisissant par exemple $g(t) = \exp(-t\alpha)$ avec $\alpha > 0$, il est facile de constater que le processus $L_t - \mathbb{E}(L_t)$ devient "asymptotiquement stationnaire".

On peut remarquer que les fonctions de covariance du processus de Poisson et du mouvement brownien sont identiques, ce qui indique encore une fois que les propriétés du second ordre d'un processus sont tout à fait insuffisantes pour caractériser son comportement. . .

Chapitre 5

Processus stationnaires aléatoires

Dans ce chapitre on ne considère que des processus indexés par \mathbb{Z} et principalement des p.s.c. qui satisfont à une relation du type

$$X_n - \sum_{k=1}^p \alpha_k X_{n-k} = B_n + \sum_{j=1}^q \beta_j B_{n-j} \quad (5.1)$$

où $(B_n)_{n \in \mathbb{Z}}$ est un bruit blanc de variance σ^2 . C'est une famille intéressante, puisque d'un coté il est facile de résoudre des problèmes tels que la prédiction, c'est-à-dire de trouver la meilleure prévision de X_{n+m} , $m > 0$, sachant X_n, X_{n-1}, \dots . D'autre part, puisque un tel modèle ne dépend que d'un petit nombre de paramètres (les α_k , le β_j , σ^2 ainsi que p et q) il est relativement simple de modéliser une série de données provenant d'applications concrètes avec un tel processus. Il s'agira alors d'estimer d'abord les paramètres ci-dessus. Ceci est le sujet du chapitre prochain, alors que dans celui-ci on va étudier de plus près les p.s.c. du type (5.1).

5.1 Inversion des filtres

On utilise ici les résultats sur la transformée en z établis dans la section 1.6. L'inversion de l'opération de filtrage constitue leur application principale dans le contexte des p.s.c.

Théorème 5.1. *Soit Y un p.s.c. et h un filtre admettant une transformée en z soit $H(z)$, holomorphe et non nulle dans un anneau contenant le cercle unité. Soit $\sum_{n \in \mathbb{Z}} a_n z^n$ le développement de Laurent de $1/H(z)$. Alors*

- *Le filtre $a = (a_n)_{n \in \mathbb{Z}}$ est stable.*
- *L'équation en X , $h * X = Y$ a une unique solution stationnaire centrée donnée par $X = a * Y$.*
- *Si H est holomorphe et non nulle dans un voisinage du disque unité alors le filtre a est causal c'est à dire que $a_n = 0$ pour $n < 0$. On en déduit que $H_n(X) = H_n(Y)$ pour tout n .*

Démonstration. On sait que le filtre h admet un inverse a dans $\ell^1(\mathbb{Z})$ et donc $a * h = \delta_0$. Si h est causal alors $H_n(Y) \subset H_n(X)$ et si a est causal $H_n(X) \subset H_n(Y)$.

□

On va s'intéresser particulièrement aux filtres dont la transformée en z est un polynôme. Le polynôme $P(z) = \sum_{k=0}^p \alpha_k z^k$ est la transformée en z du filtre \mathbf{p} défini par $\mathbf{p}(k) = \alpha_k$ pour $0 \leq k \leq p$ et $\mathbf{p}(k) = 0$ sinon. Il est évidemment causal et stable. On a alors $\hat{\mathbf{p}}(t) = P(\gamma_{-1}(t))$ et on peut considérer le processus filtré $(\mathbf{p} * X)_n = \sum_{k=0}^p \alpha_k X_{n-k}$.

Corollaire 5.2. *Soit Y un p.s.c. et P un polynôme sans racines de module 1. Alors l'équation $\mathbf{p} * X = Y$ a une unique solution X stationnaire centrée donnée par $X = a * Y$ où le filtre stable a est donné par le développement de Laurent de $1/P(z)$. Si P a toutes ses racines de module > 1 alors le filtre a est causal et $H_n(X) = H_n(Y)$ pour tout n .*

On verra qu'il est très important que le filtre inverse $(a_n)_{n \in \mathbb{Z}}$ soit causal. Donc la condition "toutes les racines de P sont de module > 1 " apparaîtra de façon très naturelle dans la suite.

5.2 Prédiction

Définition 5.3. *Soit $p \geq 1$. On va noter $\hat{X}_{n+p,n}$ la projection orthogonale de X_{n+p} sur $H_n(X)$.*

On a vu que $\hat{X}_{n+p,n}$ est la meilleure approximation (dans $L^2(\Omega, \mathbb{P})$) de X_{n+p} par un élément de $H_n(X)$. Donc $\hat{X}_{n+p,n}$ est la *prédiction* de X_{n+p} en fonction des X_k , $k \leq n$. On appelle *erreur de prédiction* la quantité $\|X_{n+p} - \hat{X}_{n+p,n}\|_2$.

En utilisant l'opérateur unitaire \mathcal{T} il revient au même de calculer la projection de X_p sur $H_0(X)$. Afin de pouvoir calculer cette projection orthogonale il est donc très important de pouvoir exhiber une base orthonormée de $H_n(X)$. En fait d'après la décomposition orthogonale $X_n = X'_n + X''_n$ obtenue dans 3.2 on obtient

$$\hat{X}_{n+p,n} = \text{Proj}\{X'_{n+p} | H_n(X')\} + \text{Proj}\{X''_{n+p} | H_n(X'')\} = \text{Proj}\{X'_{n+p} | H_n(X')\} + X''_{n+p}$$

et le problème est ramené au cas d'un processus purement aléatoire. Or dans ce cas on sait que $H_n(X')$ est égal à $H_n(I')$ où le processus d'innovation $(I'_n)_n$ est un bruit blanc et donc la famille $(I'_k; k \leq n)$ est, à un facteur constant près, une base orthonormée de $H'_n(X)$. Ceci explique pourquoi est-il important, pour un p.s.c. donné, de calculer le processus d'innovation. Nous allons montrer que le calcul est très simple dans les cas des processus ARMA, que l'on verra dans les paragraphes prochains.

5.3 Processus AR, MA et ARMA

5.3.1 Processus AR

Définition 5.4. *On dit que un p.s.c. X est $AR(p)$ (autorégressif d'ordre p) s'il existe un polynôme P de degré p avec $P(0) \neq 0$ et un bruit blanc normalisé B tel que $\mathbf{p} * X = B$.*

Remarquons d'abord que, pour que il existe un p.s.c. AR associé à un polynôme P dans le sens de la définition 5.4, il est nécessaire que P n'ait pas de racines de module 1, comme il suit des deux énoncés suivants. Si cette condition est satisfaite, d'après le théorème 5.1 un tel processus est unique pour (P, B) donnés.

Proposition 5.5. *Soit P et Q des polynômes sans racines communes de module 1. Si la mesure sur \mathbb{T} de densité $|Q(\gamma_{-1})|^2 / |P(\gamma_{-1})|^2$ est finie, alors P n'a pas de racines de module 1.*

Démonstration. Supposons que P ait une racine de module 1, soit $z_0 = \gamma_{-1}(t_0)$. Alors $P(z) = (z - z_0)^r R(z)$. Pour t voisin de t_0 on aura donc $|P(\gamma_{-1}(t))|^2 \sim |t - t_0|^{2r}$ et la fraction $|Q(\gamma_{-1})|^2/|P(\gamma_{-1})|^2$ n'est donc pas intégrable sur \mathbb{T} . \square

Corollaire 5.6. *Soit Y un bruit blanc. S'il existe une solution stationnaire centrée à l'équation $\mathfrak{p} * X = Y$ alors P n'a pas de racines de module 1.*

Démonstration. On a $\mu_Y = |P(\gamma_{-1})|^{-2} \mu_X$ or μ_Y est la mesure de Lebesgue sur le tore. En écrivant que la mesure μ_X est de masse totale finie on obtient le résultat. \square

On remarque aussi que l'on peut toujours remplir la condition $P(0) \neq 0$ en décalant le processus B . Le théorème 5.1 permet d'obtenir un processus AR comme filtré d'un bruit blanc, par le filtre $(a_n)_n$, où les a_n sont les coefficients du développement de Laurent de $1/P(z)$. Mais si P a des racines de module < 1 , $1/P(z)$ n'est pas holomorphe dans un voisinage du disque unité et ce filtre ne sera pas causal. Puisque cette dernière propriété est très importante pour le calcul de la prédiction, c'est pourquoi le résultat suivant est de particulière importance. Il affirme que l'on peut toujours se ramener à une situation où le polynôme P a toutes ses racines de module > 1 .

Proposition 5.7. *Soit P un polynôme sans racines de module 1 et tel que $P(0) \neq 0$. Alors il existe un polynôme \tilde{P} et un réel $\sigma > 0$ uniques tels que:*

- \tilde{P} a toutes ses racines de module > 1 , $\tilde{P}(0) = 1$.
- $\sigma|P(z)| = |\tilde{P}(z)|$ sur le cercle unité.

On dira alors que le couple (σ, \tilde{P}) est le couple canonique associé à P . Les polynômes P et \tilde{P} ont même degré.

Démonstration. Soit ξ une racine de P . Pour $|z| = 1$ on a la relation $|z - \xi| = |1 - \xi\bar{z}| = |\bar{\xi}z - 1| = |\bar{\xi}||z - 1/\bar{\xi}|$ (on multiplie par $z^{-1} = \bar{z}$). Donc le module du quotient des deux polynômes $z - \xi$ et $z - 1/\bar{\xi}$ est constant sur le cercle unité. Si ξ est une racine de P de module < 1 , on remplace alors le terme $(z - \xi)$ dans la factorisation de P par $(z - 1/\bar{\xi})$. Le polynôme ainsi obtenu a alors toutes ses racines de module > 1 et il suffit de le diviser par son terme constant (non nul) pour obtenir \tilde{P} .

Il reste à vérifier l'unicité de \tilde{P} . Pour ceci soit A et B des polynômes dont les racines sont de module > 1 , ayant le même coefficient d'ordre 0 et tels que $|A(z)| = c|B(z)|$ sur le cercle unité. On note ξ_k , $k = 1 \dots n$ les racines de A et η_r , $r = 1 \dots m$ les racines de B . On a alors, pour $|z| = 1$,

$$\prod_{k=1}^n (z - \xi_k)(\bar{z} - \bar{\xi}_k) = c^2 \prod_{r=1}^m (z - \eta_r)(\bar{z} - \bar{\eta}_r)$$

On multiplie à droite et gauche par z^{n+m} et on se rappelle que, pour $|z| = 1$, $z\bar{z} = 1$. Donc

$$z^m \prod_{k=1}^n (z - \xi_k)(1 - \bar{\xi}_k z) = c^2 z^n \prod_{r=1}^m (z - \eta_r)(1 - \bar{\eta}_r z)$$

Les deux derniers polynômes, puisqu'ils coïncident sur le cercle unité, sont identiques. On en déduit que l'ensemble des racines $E_A = \{\xi_k, 1/\bar{\xi}_k\}$ $k = 1 \dots n$ du premier doit être identique à $E_B = \{\eta_r, 1/\bar{\eta}_r\}$, $r = 1 \dots m$. Puisque les ξ_k et les η_r sont de module

> 1 (et donc les $1/\bar{\xi}_k$ et $1/\bar{\eta}_r$, sont de module < 1), il faut que $n = m$ et que $\{\xi_k, k = 1, \dots, n\} = \{\eta_k, k = 1, \dots, n\}$. Les polynômes A et B ayant les mêmes racines, ils sont proportionnels. Puisque ils ont le même coefficient d'ordre 0, ils sont égaux \square

Théorème 5.8. Soit X un processus AR, (σ, \tilde{P}) le couple canonique associé à P et a le filtre stable associé au développement de Taylor de $1/\tilde{P}$, qui est donc causal. Alors:

1. μ_X a une densité donnée par $1/|P(\gamma_{-1})|^2 = \sigma^2/|\tilde{P}(\gamma_{-1})|^2$.
2. Si on pose $U = \tilde{\mathfrak{p}} * X$, le processus U est un bruit blanc de variance σ^2 .
3. $X = a * U$ et donc $H_n(X) = H_n(U)$ pour tout n . Il s'en suit que U est le processus d'innovation de X et que X est purement aléatoire.

Démonstration. (1): C'est une conséquence de $d\mu_X = \frac{1}{|P(\gamma_{-1})|^2} d\mu_B$.

(2): On a $\mu_U = |\tilde{P}(\gamma_{-1})|^2 d\mu_X = \mu_B \frac{|\tilde{P}(\gamma_{-1})|^2}{|P(\gamma_{-1})|^2} = \sigma^2 d\mu_B$

(3) Évident \square

La relation $U = \tilde{\mathfrak{p}} * X$ s'appelle *la relation ARMA canonique*. Le polynôme canonique intervient de façon décisive dans le calcul de la prédiction.

Proposition 5.9. Soit X un processus $AR(p)$ dont le polynôme canonique est de la forme $\tilde{P}(z) = 1 - \sum_{k=1}^p \alpha_k z^k$. Alors la prédiction $\hat{X}_{n+1,n}$ de X_{n+1} en fonction de $H_n(X)$ est donnée par:

$$\hat{X}_{n+1,n} = \sum_{k=1}^p \alpha_k X_{n+1-k}$$

et l'erreur de prédiction est égale à σ .

Démonstration. La relation canonique $U = \tilde{\mathfrak{p}} * X$ s'écrit $U_{n+1} = X_{n+1} - \sum_{k=1}^p \alpha_k X_{n+1-k}$ à l'ordre $n+1$. Puisque U_{n+1} est orthogonal à $H_n(U) = H_n(X)$ et $\sum_{k=1}^p \alpha_k X_{n+1-k} \in H_n(X)$ on a immédiatement le résultat. \square

Remarque 5.10. Comme $\hat{X}_{n+1,n}$ est combinaison linéaire de X_n, \dots, X_{n-p} , on en déduit que, pour un processus $AR(p)$, les projections de X_{n+1} sur $H_n(X)$ et sur l'espace engendré par (X_n, \dots, X_{n-p}) coïncident.

Remarque 5.11. Si $s > 1$, pour trouver $\hat{X}_{n+s,n}$ en fonction des $X_k, k \leq n$ on effectue plusieurs projections successives. Remarquons néanmoins que le calcul de la prédiction est très facile si on se contente d'exprimer $\hat{X}_{n+s,n}$ en fonction du bruit blanc $(U_n)_n$. En effet, utilisant le filtre a (qui est l'inverse de \tilde{p}), on a

$$X_{n+s} = \sum_{k=0}^{\infty} a_k U_{n+s-k} = \underbrace{\sum_{k=0}^{s-1} a_k U_{n+s-k}}_{\text{orthogonal à } H_n(X)} + \underbrace{\sum_{k=s}^{\infty} a_k U_{n+s-k}}_{\text{appartenant à } H_n(X)}$$

Donc $\hat{X}_{n+s,n} = \sum_{k=s}^{\infty} a_k U_{n+s-k}$ Ceci permet de faire tout de suite le calcul de l'erreur, car on a

$$\|X_{n+s} - \hat{X}_{n+s,n}\|_2^2 = \left\| \sum_{k=0}^{s-1} a_k U_{n+s-k} \right\|_2^2 = \sigma^2 \sum_{k=0}^{s-1} |a_k|^2$$

et donc l'erreur vaut

$$\sigma \left(\sum_{k=0}^{s-1} |a_k|^2 \right)^{1/2}$$

□

On peut caractériser un processus AR par sa densité spectrale:

Proposition 5.12. *Soit X un p.s.c. de densité spectrale $1/|P(\gamma_{-1})|^2$ où P est un polynôme tel que $P(0) \neq 0$. Alors X est un processus AR.*

Démonstration. On définit V par $V = \mathfrak{p} * X$. Alors V est un bruit blanc et donc X est AR. □

Définition 5.13. *Soit X un p.s.c. On définit la corrélation partielle d'ordre $s \geq 1$, soit $r_X(s)$, comme le coefficient de X_{n-s} dans la projection orthogonale de X_n sur l'espace engendré par $(X_{n-1}, \dots, X_{n-s})$.*

Ceci conduit à une autre caractérisation des processus AR.

Théorème 5.14. *Soit X un p.s.c. Alors X est AR(p) si et seulement si $r_X(p) \neq 0$ et $r_X(s) = 0$ pour $s > p$.*

Démonstration. Soit X un processus AR(p) de polynôme canonique $\tilde{P}(z) = 1 - \sum_{k=1}^p \alpha_k z^k$. Il vérifie donc l'équation canonique $\sum_{k=1}^p \alpha_k X_{n-k} + U_n = X_n$ où $(U_n)_{n \in \mathbb{Z}}$ est un bruit blanc. Puisque U_n est orthogonal à $H_{n-1}(X)$ et donc à $H_{n-s}(X)$ pour tout s , la projection canonique de X_n sur l'espace engendré par $(X_{n-1}, \dots, X_{n-s})$ est, pour $s \geq p$, égale à $\sum_{k=1}^p \alpha_k X_{n-k}$. Pour $s = p$ on trouve $r_X(p) = \alpha_p \neq 0$, alors que $r_X(s) = 0$ pour $s > p$. Réciproquement, si $r_X(s) = 0$ pour $s > p$ on a alors pour tout $m \geq 0$:

$$\text{Proj}(X_n | (X_{n-1}, X_{n-2}, \dots, X_{n-p-m-1})) = \text{Proj}(X_n | (X_{n-1}, X_{n-2}, \dots, X_{n-p-m}))$$

D'après la Proposition 1.10 ceci implique que la projection de X_n sur H_{n-1} est égale à celle sur $(X_{n-1}, \dots, X_{n-p})$ et donc à une combinaison linéaire de la forme $\sum_{k=1}^p \alpha_k X_{n-k}$ avec $\alpha_p \neq 0$. On en déduit que l'innovation de $(X_n)_n$ (qui est un bruit blanc) est de la forme $X_n - \sum_{k=1}^p \alpha_k X_{n-k}$ et donc que X est AR(p).

5.3.2 Processus MA

Définition 5.15. *On dit que un p.s.c. X est MA(q) (moyenne mobile d'ordre q) s'il existe un polynôme Q , sans racine de module 1, de degré q , avec $Q(0) \neq 0$ et un bruit blanc normalisé B tel que $X = \mathfrak{q} * B$.*

En anglais "moyenne mobile" se dit *moving average*, ce qui explique l'abrégié MA.

Remarque 5.16. Un processus du type $X = \mathfrak{q} * B$ est toujours un p.s.c. À la différence du cas AR, la condition "sans racine de module 1" n'est pas nécessaire pour que le processus X défini par la relation ci-dessus soit un p.s.c. On l'impose tout de même parce qu'elle devient très vite nécessaire. Par exemple pour exprimer le bruit blanc $(B_n)_n$ comme filtré de $(X_n)_n$, car sinon, en général, $H(X) \subsetneq H(B)$.

Le théorème 5.1 permet d'obtenir le bruit blanc B comme filtré de X . Mais si Q a des racines de module < 1 ce filtre ne sera pas causal, on n'aura pas la relation $H_n(X) = H_n(B)$. C'est pourquoi, comme dans le cas AR , on va chercher une autre représentation de X .

On définit le couple canonique comme dans la proposition 5.7 mais on choisit la convention $|Q(z)| = \sigma|\tilde{Q}(z)|$ sur le cercle unité. La preuve du théorème suivant est immédiate.

Théorème 5.17. *Soit X un processus MA , (σ, \tilde{Q}) le couple canonique associé à Q et b le filtre causal et stable associé au développement de Taylor de $1/\tilde{Q}$. Alors:*

1. μ_X a une densité donnée par $|Q(\gamma_{-1})|^2$.
2. Si on pose $U = b * X$, le processus U est un bruit blanc de variance σ^2 .
3. $X = \tilde{\mathbf{q}} * U$ (équation canonique) et donc $H_n(X) = H_n(U)$ pour tout n . Il s'en suit que le processus U est l'innovation du processus X .

On peut caractériser un processus MA par sa densité spectrale:

Proposition 5.18. *Soit X un p.s.c. de densité spectrale $|Q(\gamma_{-1})|^2$ où Q est un polynôme n'ayant pas de racines de module 1 et tel que $Q(0) \neq 0$. Alors X est un processus MA .*

Démonstration. Soit b le filtre associé au développement de Laurent de $1/Q(z)$. On définit le processus V par $V = b * X$. Alors V est un bruit blanc et $X = \mathbf{q} * V$. Donc est MA . \square

Un aspect typique des processus MA est le fait que leur fonction de covariance est à support fini: on a en effet $c_X(k) = 0$ pour tout $k, |k| > q$. Ceci peut être vu avec la formule de la covariance: $c_X(k) = \mathbf{q} * \tilde{\mathbf{q}}(k) = \sum_{n \geq k} \mathbf{q}(n)\mathbf{q}(n-k)$ où, de façon plus intuitive, en remarquant que, si on note $\tilde{Q}(z) = 1 + \beta_1 z + \dots + \beta_q z^q$ (et $\beta_0 = 1$),

$$c_X(k) = \mathbb{E} \left[\sum_{j=0}^q \beta_j U_{n+k-j} \sum_{h=0}^q \tilde{\beta}_h \bar{U}_{n-h} \right] = \sum_{j,h=0}^q \beta_j \tilde{\beta}_h \mathbb{E}[U_{n+k-j} \bar{U}_{n-h}]$$

Or, si $k > q$, les indices $n+k-j$ sont tous $> n$. Donc $\mathbb{E}[U_{n+k-j} \bar{U}_{n-h}] = 0$ pour tout h, j . Par contre si $k = q$ alors $\mathbb{E}[U_{n+k-j} \bar{U}_{n-h}] = 1$ pour $j = q$ et $h = 0$, alors que $\mathbb{E}[U_{n+k-j} \bar{U}_{n-h}] = 0$ pour toute autre valeur des indices. Il s'ensuit $c_X(q) = \beta_q \neq 0$. Cette propriété caractérise, à quelque chose près, les processus MA , comme le montre le résultat suivant.

Théorème 5.19. *Soit X un p.s.c. Alors X est $MA(q)$ si et seulement si:*

1. Sa covariance vérifie $c_X(q) \neq 0$ et $c_X(n) = 0$ pour $n > q$.
2. Sa mesure spectrale μ_X possède une densité strictement positive sur \mathbb{T} .

Démonstration. Si X est $MA(q)$, on vient de voir (1) et (2) est une conséquence du théorème 5.17 (1) et du fait que Q n'a pas de racine de module 1.

Pour la réciproque, d'après la proposition 5.18, il suffit de montrer que la densité de μ_X soit $f = \sum_{k=-q}^q c_X(k)\gamma_{-k}$ peut s'écrire sous la forme $f = |Q(\gamma_{-1})|^2$ pour un polynôme Q de degré q (qui n'aura pas de racines de module 1 d'après la condition $f > 0$). Pour ceci, pour $z \neq 0$, on considère la fonction $h(z) = \sum_{k=-q}^q c_X(k)z^k$ et le polynôme $H(z) = z^q h(z)$. On remarque alors que:

- $H(0) = c_X(-q) = \overline{c_X(q)} \neq 0$ et H est de degré $2q$.
- Si z est racine de H , il en est de même de $1/\bar{z}$ car $H(1/\bar{z}) = \overline{H(z)}/\bar{z}^{2q}$.
- H n'a pas de racine de module 1 puisqu'alors f s'annulerait en un point.

Il s'en suit que si z_1, z_2, \dots, z_q sont les racines de H de module > 1 , on peut écrire

$$H(z) = K \prod_{k=1}^q (1 - z/z_k)(1 - z\bar{z}_k)$$

En posant $Q(z) = \sqrt{|K|} \prod_{k=1}^q (1 - z/z_k)$ on a pour $|z| = 1$, $|h(z)| = |H(z)| = |Q(z)|^2$. \square

On verra dans le prochain paragraphe la solution du problème de prédiction, mais notons néanmoins que, si $s > q$,

$$\hat{X}_{n+s,n} = 0$$

En effet, si $s > q$, on a vu que $0 = c_X(s) = \mathbb{E}[X_{n+s}\bar{X}_n]$. Donc X_{n+s} est orthogonal à X_n et, par le même argument, est orthogonal à X_k pour tout $k \leq n$. Donc X_{n+s} est orthogonal à $H_n(X)$.

5.3.3 Processus ARMA

Définition 5.20. On dit que un p.s.c. X est ARMA s'il existe des polynômes P et Q avec $P(0) \neq 0$, $Q(0) \neq 0$, Q n'ayant pas de racines de module 1, et un bruit blanc normalisé B tel que $\mathfrak{p} * X = \mathfrak{q} * B$.

- Si un tel processus existait, alors il aurait une densité spectrale $|Q(\gamma_{-1})/P(\gamma_{-1})|^2$. Donc, d'après la proposition 5.5, le polynôme P n'a certainement pas de racines de module 1. Par contre, si cette condition est satisfaite, le théorème 5.1 implique qu'un tel processus existe et est unique, pour toute donnée du triplet (P, Q, B) .
- On peut toujours remplir les conditions $P(0) \neq 0$, $Q(0) \neq 0$ en décalant les processus B et X .
- Soit P et Q des polynômes vérifiant $P(0) \neq 0$, $Q(0) \neq 0$ et sans racines de module 1. Si X est un processus AR associé au polynôme P , alors $Y = \mathfrak{q} * X$ est un processus ARMA associé aux polynômes P et Q . En effet si $(B_n)_n$ est un bruit blanc tel que $\mathfrak{p} * X = B$, alors $\mathfrak{p} * Y = \mathfrak{p} * \mathfrak{q} * X = \mathfrak{q} * \mathfrak{p} * X = \mathfrak{q} * B$, le produit de convolution étant commutatif.

Proposition 5.21. Soient P et Q des polynômes avec $P(0) \neq 0$, $Q(0) \neq 0$, et n'ayant pas de racines de module 1. Il existe alors des polynômes \tilde{P} et \tilde{Q} et un réel $\sigma > 0$ uniques tels que:

1. \tilde{P} et \tilde{Q} n'ont pas de racines de module ≤ 1 et $\tilde{P}(0) = \tilde{Q}(0) = 1$.
2. \tilde{P} et \tilde{Q} n'ont pas de racines communes.
3. $\sigma \frac{|P(z)|}{|Q(z)|} = \frac{|\tilde{P}(z)|}{|\tilde{Q}(z)|}$ sur le cercle unité.

Un tel triplet $(\sigma, \tilde{P}, \tilde{Q})$ sera appelé triplet canonique associé à (P, Q) et si p et q sont les degrés respectifs de \tilde{P} et \tilde{Q} , on dira que X est ARMA(p, q).

Démonstration. On fabrique \tilde{P} et \tilde{Q} comme dans 5.7 puis on élimine leurs racines communes éventuelles. Il reste à montrer l'unicité. Soit alors (A_1, A_2) et (B_1, B_2) deux couples satisfaisant aux propriétés de l'énoncé et de degrés respectifs p_1, p_2 et q_1, q_2 . En faisant le même calcul que dans la preuve de 5.7 on obtient que A_1 et B_1 d'une part et A_2, B_2 d'autre part ont les mêmes racines et sont donc proportionnels. \square

Définition 5.22. On définit

le filtre constructeur a associé au développement de Taylor de $\frac{\tilde{Q}}{\tilde{P}}$
le filtre blanchisseur b associé au développement de Taylor de $\frac{\tilde{P}}{\tilde{Q}}$

On remarque que les deux filtres a et b sont de type causal et l'un l'inverse de l'autre. Dans le cas d'un processus AR le filtre blanchisseur b a une transformée en z qui est un polynôme. Dans le cas MA c'est le filtre constructeur a qui a une transformée en z qui est un polynôme.

Théorème 5.23. Soit X un processus ARMA et $(\sigma, \tilde{P}, \tilde{Q})$ le triplet canonique associé à (P, Q) . Soient a et b les filtres constructeur et filtre blanchisseur respectivement. Alors:

1. μ_X a une densité donnée par $|Q(\gamma_{-1})|^2/|P(\gamma_{-1})|^2$.
2. Soit U défini par $U = b * X$, alors U est un bruit blanc de variance σ^2 tel que $X = a * U$. Donc $H_n(X) = H_n(U)$ pour tout n . Il s'en suit que le processus U est l'innovation du processus X et que X est purement aléatoire.
3. $\tilde{p} * X = \tilde{q} * U$ (équation ARMA canonique).

Démonstration. (1) est évident. On montre facilement que la mesure spectrale de U est $d\mu_U(t) = \sigma^2 dt$ et donc U est un bruit blanc. Les autres calculs sont faciles en tenant compte du fait que $a * b = \delta_0$, $\tilde{p} * a = \tilde{q}$, $\tilde{q} * b = \tilde{p}$ et que a et b sont de type causal. \square

Ce théorème a une application directe à la prédiction qui est une généralisation de 5.9.

Proposition 5.24. Soit X un processus ARMA. Alors la prédiction $X_{n+1,n}$ est donnée par:

$$\hat{X}_{n+1,n} = - \sum_{k=1}^{\infty} b_k X_{n+1-k}$$

et l'erreur de prédiction est égale à σ (ici $(b_n)_n$ est toujours le filtre blanchisseur).

Démonstration. La relation $U = b * X$ s'écrit à l'ordre $n+1$:

$$U_{n+1} = X_{n+1} + \sum_{k=1}^{\infty} b_k X_{n+1-k}$$

Puisque U_{n+1} est orthogonal à $H_n(X)$ et $\sum_{k=1}^{\infty} b_k X_{n+1-k} \in H_n(X)$ on a immédiatement le résultat. \square

Exemple 5.25. On considère le modèle AR $X_n = U_n + \sum_{k=1}^p \alpha_k X_{n-k}$. Dans ce cas simple le filtre blanchisseur est $b_n = \alpha_n$ (avec la convention $b_k = 0$ sauf si $0 \leq k \leq p$). On retrouve donc la formule de la proposition 5.9.

Exemple 5.26. On considère le modèle $MA(1)$

$$X_n = U_n - \beta U_{n-1}$$

où $(U_n)_n$ est un bruit blanc de variance 1. On suppose $|\beta| < 1$, ce qui garantit que le polynôme $Q(z) = 1 - \beta z$ n'a que des racines (une seule, en fait) de module > 1 . Alors

$$\frac{1}{Q(z)} = \sum_{k=0}^{\infty} \beta^k z^k$$

Donc le filtre blanchisseur est $b_n = (\beta)^n$ pour $n \geq 0$ et $b_n = 0$ pour $n < 0$ et

$$\hat{X}_{n+1,n} = - \sum_{k=1}^{\infty} \beta^k X_{n+1-k}$$

L'erreur de prédiction est égal à 1. Si par contre $|\beta| > 1$, alors le polynôme Q n'est pas canonique et, pour calculer $\hat{X}_{n+1,n}$ il faut d'abord trouver la forme canonique. Si on reprend l'argument de la preuve de la proposition 5.7, on a, pour $|z|^2 = z\bar{z} = 1$

$$|1 - \beta z| = |\bar{z} - \beta| = |z - \bar{\beta}| = |\beta| |1 - z/\bar{\beta}|$$

Le polynôme $\tilde{Q}(z) = 1 - z/\bar{\beta}$ a pour racine, $\bar{\beta}$, qui est de module > 1 et $\sigma|\tilde{Q}(z)| = |Q(z)|$ sur $|z| = 1$, avec $\sigma = |\beta|$. Si on note \tilde{b} le filtre inverse de \tilde{q} , par le théorème 5.23, le p.s.c. $V = \tilde{b} * X$ est un bruit blanc de variance σ^2 et on a la relation ARMA canonique $X = \tilde{q} * V$. On peut maintenant répéter l'argument précédent avec β remplacé par $1/\bar{\beta}$, ce qui donne

$$\hat{X}_{n+1,n} = - \sum_{k=1}^{\infty} \bar{\beta}^{-k} X_{n+1-k}$$

Maintenant l'erreur de prédiction vaut $\sigma = |\beta|$. □

Remarque 5.27. Dans le cas $s > 1$ on effectue plusieurs projections successives pour trouver $\hat{X}_{n+s,n}$. Remarquons néanmoins que, exactement comme dans la remarque 5.11, le calcul de la prédiction est très facile si on se contente d'exprimer $\hat{X}_{n+s,n}$ en fonction du bruit blanc $(U_n)_n$. En effet, utilisant le filtre constructeur a ,

$$X_{n+s} = \sum_{k=0}^{\infty} a_k U_{n+s-k} = \underbrace{\sum_{k=0}^{s-1} a_k U_{n+s-k}}_{\text{orthogonal à } H_n(X)} + \underbrace{\sum_{k=s}^{\infty} a_k U_{n+s-k}}_{\text{appartenant à } H_n(X)}$$

Donc $\hat{X}_{n+s,n} = \sum_{k=s}^{\infty} a_k U_{n+s-k}$. Ceci permet de faire tout de suite le calcul de l'erreur car

$$\|X_{n+s} - \hat{X}_{n+s,n}\|_2^2 = \left\| \sum_{k=0}^{s-1} a_k U_{n+s-k} \right\|_2^2 = \sigma^2 \sum_{k=0}^{s-1} |a_k|^2$$

On peut aussi partir d'une mesure spectrale donnée pour construire un processus ARMA, lorsque celle-ci possède une densité de forme particulière:

Théorème 5.28. Soit μ une mesure finie sur \mathbb{T} de densité $|Q(\gamma_{-1})|^2/|P(\gamma_{-1})|^2$ où P et Q sont des polynômes avec $P(0) \neq 0$, $Q(0) \neq 0$, Q n'ayant pas de racines de module 1. Soit $(\sigma, \tilde{P}, \tilde{Q})$ le triplet canonique associé à (P, Q) et a et b les filtres constructeur et blanchisseur respectivement. Soit V un bruit blanc (quelconque) de variance σ^2 . On a alors:

1. Le processus X défini par $X = a * V$ vérifie $V = b * X$ et donc $H_n(X) = H_n(V)$, $\forall n$.
2. Le processus X vérifie $\tilde{\mathbf{p}} * X = \tilde{\mathbf{q}} * V$ (c'est donc un processus ARMA)
3. la mesure spectrale de X est égale à μ .

Ce théorème c'est un premier pas dans la direction de la *simulation* d'un processus ARMA. En général il est facile de simuler un bruit blanc et il est affirmé que un modèle ARMA est toujours un filtré d'un bruit blanc.

5.3.4 La fonction de covariance des processus ARMA

Dans cette section on étudie la fonction de covariance c_X d'un processus ARMA stationnaire.

Proposition 5.29. Soit $(X_n)_n$ un processus ARMA(p, q) stationnaire de triplet canonique $(\sigma, \tilde{P}, \tilde{Q})$ et soit a le filtre constructeur Alors il existe $0 \leq \rho < 1$ tel que

$$|a_j| \leq K \rho^j, \quad j \geq 0 \quad (5.2)$$

Et donc $|c_X(k)| \leq K' \rho^{|k|}$ pour tout k .

Démonstration. La fonction rationnelle \tilde{Q}/\tilde{P} est holomorphe dans le disque $|z| < R$, où $R > 1$ est le plus petit module des zéros de \tilde{P} . Donc sa série de Taylor est absolument convergente pour $|z| \leq R - \varepsilon$, pour tout $\varepsilon > 0$. Comme le filtre constructeur est donné par les coefficients de cette série, la série à termes positifs $\sum_j |a_j|(R - \varepsilon)^j$ est convergente et il existe donc $K < \infty$ tel que $|a_j| \leq K(R - \varepsilon)^{-j}$. Si on choisit ε assez petit, de façon à avoir $R - \varepsilon > 1$, on obtient (5.2). L'inégalité $|c_X(k)| \leq K'(R - \varepsilon)^{|k|}$ suit maintenant de la remarque 3.22 où il est indiqué que pour $k \geq 0$ on a $c_X(k) = \sigma^2 \sum_{\ell=0}^{\infty} a_{j+k} \bar{a}_j$ et on utilise le fait que pour $k < 0$ on a $c_X(k) = \overline{c_X(-k)}$. \square

Remarque 5.30. La décroissance exponentielle des covariances d'un processus ARMA est une propriété importante. Elle doit être prise en compte quand on essaie de modéliser des données. Les modèles ARMA ne sont pas adaptés à la modélisation dans des situations de "mémoire longue" où la corrélation reste importante à grande distance, ce qui indique la présence probable d'une partie ponctuelle dans la mesure spectrale.

Remarque 5.31. On a vu dans le théorème 5.19 que dans un modèle MA(q) la fonction de covariance non seulement décroît à l'infini exponentiellement vite, mais $c_X(k)$ est nulle pour $|k| > q$. On voit même sans peine que les processus MA sont les seuls parmi les ARMA avec cette propriété. En effet un processus ARMA a toujours une densité spectrale qui ne s'annule pas et on applique à nouveau le théorème 5.19.

5.4 Matrices de Toeplitz

On ne considère maintenant, que des processus réels à indices dans \mathbb{Z} .

Définition 5.32. On appelle *matrice de Toeplitz* une matrice R_n carrée d'ordre $n \geq 1$ à coefficients réels telle que:

- $R_n(i, j)$ ne dépend que de $|i - j|$.
- R_n est de type positif.

Un exemple de matrice de Toeplitz est fourni par la matrice $R_n(i, j) = c(i - j)$ où $(c(0), c(1), \dots, c(n - 1))$ sont donnés par la fonction de covariance c d'un p.s.c. réel. Dans la suite on désigne par $R_n(X)$ la matrice de Toeplitz associée à un p.s.c. X . En fait comme le montre le théorème ci dessous toute matrice de Toeplitz est de ce type et ceci pour une infinité de p.s.c. Le problème abordé dans ce chapitre est de construire certains d'entre eux.

On rappelle que, étant donnée une matrice R , $n \times n$ de type positif, il existe toujours un vecteur aléatoire (X_1, \dots, X_n) ayant R comme matrice de covariance. Il suffit en effet de le choisir gaussien.

Théorème 5.33. Soit R_n une matrice de Toeplitz. Alors il existe au moins un prolongement de R_n en une matrice de Toeplitz d'ordre $n + 1$ (en fait il y a une infinité de prolongements, sauf si R_n est singulière auquel cas ce prolongement est unique). On en conclut qu'il existe au moins un p.s.c. réel dont la matrice de covariance d'ordre n coïncide avec R_n et donc qu'il existe une mesure positive, finie μ sur \mathbb{T} telle que:

$$R_n(i, j) = \int \gamma_{i-j}(t) d\mu(t), \text{ pour } 1 \leq i, j \leq n$$

Démonstration. Soit (X_1, \dots, X_n) un vecteur aléatoire réel centré admettant R_n comme matrice de covariance c'est à dire $R_n(i, j) = c_X(i - j) = \langle X_i, X_j \rangle$ pour $1 \leq i, j \leq n$. On pose $\text{Proj}(X_n | (X_1, \dots, X_{n-1})) = \sum_{k=1}^{n-1} a_{n-k}^{(n-1)} X_k$. On a alors pour $r = 2, \dots, n$ la relation:

$$c_X(n + 1 - r) = \langle X_n, X_{r-1} \rangle = \sum_{k=1}^{n-1} a_{n-k}^{(n-1)} \langle X_k, X_{r-1} \rangle \quad (5.3)$$

On va chercher à ajouter un vecteur Y au système (X_1, \dots, X_n) de telle façon que la matrice de covariance de (X_1, \dots, X_n, Y) prolonge R_n . Pour ceci on va chercher Y sous la forme $Y = Z + \sum_{k=1}^{n-1} a_{n-k}^{(n-1)} X_{k+1}$ et l'on a bien un prolongement si:

1. $\langle Y, X_r \rangle = c_X(n + 1 - r)$ pour $r = 2, \dots, n$
2. $\langle Y, Y \rangle = c_X(0)$

Or $\langle Y, X_r \rangle = \langle Z, X_r \rangle + \sum_{k=1}^{n-1} a_{n-k}^{(n-1)} \langle X_{k+1}, X_r \rangle = \langle Z, X_r \rangle + c_X(n + 1 - r)$ Donc les conditions 1. ci-dessus sont satisfaites si Z est orthogonal à (X_2, \dots, X_n) . D'autre part

$$\begin{aligned} \langle Y, Y \rangle &= \|Z\|_2^2 + \left\| \sum_{k=1}^{n-1} a_{n-k}^{(n-1)} X_{k+1} \right\|_2^2 = \|Z\|_2^2 + \left\| \sum_{k=1}^{n-1} a_{n-k}^{(n-1)} X_k \right\|_2^2 = \\ &= \|Z\|_2^2 + \|\text{Proj}(X_n | (X_1, \dots, X_{n-1}))\|_2^2 = \\ &= \|Z\|_2^2 + \|X_n\|_2^2 - \|X_n - \text{Proj}(X_n | (X_1, \dots, X_{n-1}))\|_2^2 \end{aligned}$$

Or, puisque $\|X_n\|_2^2 = c_X(0)$, pour satisfaire aux conditions (1) et (2) il suffit de choisir Z orthogonal à (X_2, \dots, X_n) et tel que

$$\|Z\|_2^2 = \|X_n - \text{Proj}(X_n | (X_1, \dots, X_{n-1}))\|_2^2$$

ceci est toujours possible d'une infinité de façons sauf si X_n est dans l'espace engendré par (X_1, \dots, X_{n-1}) auquel cas on doit choisir $Z = 0$. \square

Le problème du prolongement nous amène donc au calcul des coefficients de régression $a_1^{(n-1)}, \dots, a_{n-1}^{(n-1)}$, tels que

$$\widehat{X}_n = \text{Proj}(X_n | (X_1, \dots, X_{n-1})) = \sum_{k=1}^{n-1} a_{n-k}^{(n-1)} X_k,$$

et de l'erreur de prédiction $\sigma_n = \|\widehat{X}_n - X_n\|$. Puisque les vecteurs \widehat{X}_n et $X_n - \widehat{X}_n$ sont orthogonaux, ce dernier vaut

$$\sigma_n^2 = \langle X_n - \widehat{X}_n, X_n - \widehat{X}_n \rangle = \left\langle X_n - \sum_{k=1}^{n-1} a_{n-k}^{(n-1)} X_k, X_n \right\rangle = c_X(0) - \sum_{k=1}^{n-1} a_k^{(n-1)} c_X(k)$$

Donc les coefficients de régression sont solutions du système (5.3), qui peut s'écrire

$$\boxed{\sum_{k=1}^{n-1} a_k^{(n-1)} c_X(r-k) = c_X(r), r = 1 \dots (n-1)}$$

et on a

$$\boxed{\sigma_n^2 = c_X(0) - \sum_{k=1}^{n-1} a_k^{(n-1)} c_X(k)}$$

Ces relations sont appelées *équations de Yule-Walker* (ou de *Wiener-Hopf*) et permettent de calculer les coefficients de régression ainsi que l'erreur de prédiction. Elles sont très utiles, puisque dans la pratique on ne dispose pas de l'observation de X_k pour tout $k \leq n$ (ce serait un nombre infini d'observations). On peut d'ailleurs, en posant $\mathbf{c} = (c_X(1), c_X(2), \dots, c_X(n-1))'$ et $\mathbf{a} = (a_1^{(n-1)}, \dots, a_{n-1}^{(n-1)})'$ leur donner la forme matricielle:

$$\boxed{R_{n-1} \mathbf{a} = \mathbf{c} \quad , \quad \sigma_n^2 = c_X(0) - \mathbf{c}' \mathbf{a}}$$

Le résultat suivant indique que dans beaucoup de cas d'intérêt la matrice R_n est inversible.

Théorème 5.34. *Soit X un p.s.c. réel et notons R_n sa matrice de Toeplitz à l'ordre n . Alors si $\det R_N = 0$, la mesure spectrale est ponctuelle et charge au plus N points.*

Démonstration. $\det R_N = 0$ entraîne l'existence de $\lambda \in \mathbb{R}^N$, non nul et tel que $\langle R_N \lambda, \lambda \rangle = 0$. Mais on a, pour tout $n \in \mathbb{Z}$,

$$\text{Var} \left(\sum_{k=0}^{N-1} \lambda_k X_{n-k} \right) = \sum_{k,\ell=0}^{N-1} \lambda_k \lambda_\ell \mathbb{E}[X_{n-k} X_{n-\ell}] = \sum_{k,\ell=0}^{N-1} \lambda_k \lambda_\ell c_X(k-\ell) = \langle R_N \lambda, \lambda \rangle = 0$$

donc le processus filtré $Y_n = \sum_{k=0}^{N-1} \lambda_k X_{n-k}$, qui est centré, est identiquement nul. Donc $\mu_Y = 0$. Mais $d\mu_Y(t) = |P(\gamma_{-k}(t))|^2 d\mu_X(t)$, où $P(z) = \lambda_0 + \lambda_1 z + \dots + \lambda_{N-1} z^{N-1}$. Pour

que μ_Y soit nulle il faut donc que μ_X soit portée par l'ensemble des zéros de P , qui est composé de N points au plus. \square

En particulier les équations de Yule-Walker peuvent être résolues en calculant l'inverse de la matrice de Toeplitz R_{n-1} , qui, d'après le théorème 5.42 est très souvent inversible. En particulier si, par exemple, le p.s.c. $(X_n)_n$ a une densité spectrale, comme c'est le cas des processus *ARMA*. Dans la pratique, pourtant, il est beaucoup plus intéressant d'utiliser un procédé de calcul récursif des coefficients $a_k^{(n)}$ appelé algorithme de Levinson, plutôt que d'inverser la matrice R_{n-1} . On le verra au prochain paragraphe.

Remarque 5.35. On remarque que si on a

$$\text{Proj}(X_{n+1} | (X_1, X_2, \dots, X_n)) = \sum_{k=1}^n a_k^{(n)} X_{n+1-k},$$

alors, utilisant l'opérateur unitaire \mathcal{T} , on a aussi, pour tout $n \in \mathbb{Z}$ et $s > 0$,

$$\text{Proj}(X_n | (X_{n-s}, \dots, X_{n-1})) = \sum_{k=1}^s a_k^{(s)} X_{n-k},$$

donc le coefficient $a_s^{(s)}$ est la corrélation partielle (définition 5.13).

Remarque 5.36. Soit $(X_n)_n$ un processus *AR*(p), de couple canonique (σ, \tilde{P}) , $\tilde{P}(z) = 1 - \sum_{k=1}^p \alpha_k z^k$. Grâce à la remarque 5.10, si $n \geq p$, la projection $\hat{X}_{n,n-1}$ de X_n sur $H_{n-1}(X)$ coïncide avec $\hat{X}_n = \text{Proj}(X_n | (X_1, \dots, X_n))$. Celle ci par ailleurs (proposition 5.9) vaut

$$\hat{X}_{n,n-1} = \sum_{k=1}^p \alpha_k X_{n-k}$$

Dans ce cas donc, si $n \geq p$, la solution de l'équation de Yule-Walker est

$$a_k^{(n-1)} = \begin{cases} \alpha_k & \text{si } k \leq p \\ 0 & \text{sinon} \end{cases}$$

Donc les coefficients $\alpha_1, \dots, \alpha_n$ apparaissent comme solution de l'équation de Yule-Walker. Cette remarque est à la base d'une méthode d'estimation des paramètres des modèles *AR*, qu'on verra au prochain chapitre.

5.4.1 L'Algorithme de Levinson

On remarque que, si on pose

$$\text{Proj}(X_{n+1} | (X_1, X_2, \dots, X_n)) = \sum_{k=1}^n a_{n+1-k}^{(n)} X_k = \sum_{k=1}^n a_k^{(n)} X_{n+1-k},$$

alors on a

$$\text{Proj}(X_0 | (X_{-n}, \dots, X_{-1})) = \sum_{k=1}^n a_{n+1-k}^{(n)} X_{k-n+1} \quad (5.4)$$

$$\text{Proj}(X_0 | (X_1, \dots, X_n)) = \sum_{k=1}^n a_k^{(n)} X_k \quad (5.5)$$

En effet pour la première relation il suffit d'utiliser l'opérateur unitaire de translation \mathcal{T} . Pour la seconde on constate que, dans le cas réel, l'opérateur de retournement du temps $\mathcal{S}X_n = X_{-n}$ est aussi un opérateur unitaire sur $H(X)$. \square

Définition 5.37. Soit X un p.s.c. réel et $n \geq 1$. On pose alors:

- $\widehat{X}_{n+1} = \text{Proj}(X_{n+1} | (X_1, X_2, \dots, X_n)) = \sum_{k=1}^n a_{n+1-k}^{(n)} X_k = \sum_{k=1}^n a_k^{(n)} X_{n+1-k}$
- $p_n = \|X_{n+1} - \widehat{X}_{n+1}\|^2$

Théorème 5.38. (Algorithme de Levinson). On a le schéma récursif suivant:

1. $a_k^{(n+1)} = a_k^{(n)} - k_{n+1} a_{n+1-k}^{(n)}$ pour $1 \leq k \leq n$
2. $a_{n+1}^{(n+1)} = k_{n+1}$
3. $p_{n+1} = p_n(1 - k_{n+1}^2)$

le coefficient k_{n+1} étant déterminé par $k_{n+1} = 0$ si $p_n = 0$ et sinon:

$$k_{n+1} = \frac{1}{p_n} \mathbb{E}(X_0(X_{n+1} - \widehat{X}_{n+1})) = \frac{1}{p_n} \left(c_X(n+1) - \sum_{k=1}^n a_{n+1-k}^{(n)} c_X(k) \right)$$

On obtient donc le système $(p_N, a_1^{(N)}, \dots, a_N^{(N)})$ à partir de $R_{N+1}(X)$.

Démonstration. On suppose la relation satisfaite au rang n . On a évidemment

$$\text{Proj}(X_0 | ((X_{n+1} - \widehat{X}_{n+1}))) =$$

en utilisant (5.5)

$$\begin{aligned} \text{Proj}(X_0 | (X_1, \dots, X_{n+1})) &= \text{Proj}(X_0 | (X_1, \dots, X_n)) + \text{Proj}(X_0 | (X_{n+1} - \widehat{X}_{n+1})) = \\ &= \sum_{k=1}^n a_k^{(n)} X_k + k_{n+1}(X_{n+1} - \widehat{X}_{n+1}) = \sum_{k=1}^n (a_k^{(n)} - k_{n+1} a_{n+1-k}^{(n)}) X_k + k_{n+1} X_{n+1} \end{aligned}$$

On obtient donc les relations (1) et (2) en utilisant de nouveau (5.5).

(3): En utilisant un décalage de $n+1$ et un retournement du temps on peut écrire:

$$\begin{aligned} p_n &= \|X_{n+1} - \widehat{X}_{n+1}\|^2 = \|X_0 - \text{Proj}(X_0 | (X_1, \dots, X_n))\|^2 \\ p_{n+1} &= \|X_0 - \text{Proj}(X_0 | (X_1, \dots, X_n)) - \text{Proj}(X_0 | (X_{n+1} - \widehat{X}_{n+1}))\|^2 \end{aligned}$$

On remarque que $\text{Proj}(X_0 | (X_{n+1} - \widehat{X}_{n+1})) = k_{n+1}(X_{n+1} - \widehat{X}_{n+1})$ et donc le carré de la norme de cette projection est égal à $p_n k_{n+1}^2$. La variable aléatoire

$$W = X_0 - \text{Proj}(X_0 | (X_1, \dots, X_n)) - \text{Proj}(X_0 | (X_{n+1} - \widehat{X}_{n+1}))$$

est orthogonale à (X_1, \dots, X_n) et par conséquent:

$$\begin{aligned} p_n &= \|W + \text{Proj}(X_0 | (X_{n+1} - \widehat{X}_{n+1}))\|^2 = \\ &= \|W\|^2 + \|\text{Proj}(X_0 | (X_{n+1} - \widehat{X}_{n+1}))\|^2 = p_{n+1} + p_n k_{n+1}^2 \end{aligned}$$

d'où la relation demandée. \square

Remarque 5.39. En fait l'algorithme est constitué par la boucle suivante:

Initialisation: $p_0 = c_X(0)$

Ensuite on trouve

$$k_1 = \frac{c_X(1)}{c_X(0)}, \quad a_1^{(1)} = k_1$$

Pour $m = 2 \dots n$ il faut faire ensuite, dans l'ordre

- le calcul de k_m puis de p_m
- le calcul de $a_k^{(m)}$ pour $k = 1, \dots, m$

Ce calcul demande $O(n^2)$ opérations.

5.4.2 Prolongement AR d'un segment de covariance

Proposition 5.40. (*Inversion de l'Algorithme de Levinson*).

Si l'on dispose du système $(p_N, a_1^{(N)}, \dots, a_N^{(N)})$ (obtenu par l'algorithme de Levinson ou par tout autre moyen) avec $p_N \neq 0$ alors on peut reconstruire $R_{N+1}(X)$. Autrement dit il n'existe qu'une seule matrice de covariance de dimension $N + 1$ possédant les spécifications ci-dessus.

Démonstration. Supposons connu le système $(p_{n+1}, a_1^{(n+1)}, \dots, a_{n+1}^{(n+1)})$. Le calcul des coefficients à l'ordre n est un simple problème linéaire associé à la matrice $M_n = (I - k_{n+1}D)$ d'ordre n où les termes de la matrice D sont nuls sauf ceux de la seconde diagonale qui sont égaux à 1. On vérifie facilement que M_n est inversible si $k_{n+1}^2 \neq 1$ ce qui est impliqué par $p_{n+1} \neq 0$. On a aussi bien sûr $p_n \neq 0$. On peut donc calculer tous les systèmes de coefficients ainsi que les p_k jusqu'à l'ordre 1 et il est alors facile de constater par récurrence que cela permet de reconstruire la matrice R_{N+1} . En effet à l'ordre 1 on a $a_1^{(1)} = \frac{c(1)}{c(0)}$ et $p_1 = c(0)(1 - |a_1^{(1)}|^2)$. De plus si l'on connaît $c(k)$ pour $k = 1..n - 1$ alors

$$c(n) = \langle X_{n+1}, X_1 \rangle = \langle \widehat{X_{n+1}}, X_1 \rangle = \sum_{k=1}^n a_{n+1-k}^{(n)} \langle X_k, X_1 \rangle = \sum_{k=1}^n a_{n+1-k}^{(n)} c(k-1)$$

□

Le théorème suivant fournit un prolongement AR d'une matrice de Toeplitz. Ce sont ces coefficients de régression a_k ainsi que la variance p_N qui sont transmis par exemple dans les communications GSM. L'hypothèse de base est alors que sur un intervalle de temps très court la parole peut s'interpréter comme un bruit blanc modulé par le filtre des cordes vocales assimilé à un polynôme. L'appareil émetteur échantillonne donc le signal sur cet intervalle de temps, effectue un calcul de type algorithme de Levinson et transmet les a_k et p_N (En fait les coefficients de réflexion k_n et $c(0)$). L'appareil récepteur génère alors un bruit blanc qu'il module avec ces coefficients. Du fait de l'incertitude reposant sur la génération des bruits blancs les processus source et destination peuvent être fort différents du point de vue des trajectoires mais fort heureusement l'oreille n'est sensible qu'aux propriétés statistiques qui elles sont quasi identiques. . . . Comme on peut le supposer, ce procédé de base est amélioré pour restituer les différentes qualités de voix.

Théorème 5.41. Soit R_{N+1} une matrice de Toeplitz inversible et $(p_N, a_1^{(N)}, \dots, a_N^{(N)})$, $p_N \neq 0$, calculés à partir de R_{N+1} par l'algorithme de Levinson ou par tout autre moyen. On a alors:

1. Le polynôme $Q(z) = 1 - \sum_{k=1}^N a_k^{(N)} z^k$ a toutes ses racines de module > 1 .
2. Le processus AR solution de $q * Y = B$ où B est un bruit blanc de variance p_N a une fonction de covariance qui prolonge R_{N+1} .

Démonstration. (1)

On le prouve par récurrence. Si $N = 1$ alors $Q(z) = 1 - \frac{c(1)}{c(0)}z$ et par conséquent si $Q(z) = 0$ alors $|z| \geq 1$ d'après l'inégalité de Schwarz. Si $|z| = 1$ cela signifie que R_2 est singulière ce qui est impossible puisque $p_1 \neq 0$.

Ensuite on vérifie que pour $n \leq N - 1$:

$$Q_{n+1}(z) = Q_n(z) - k_{n+1}z^{n+1}Q_n(1/z)$$

et on pose $H(z) = \frac{z^{n+1}Q_n(1/z)}{Q_n(z)}$. D'après l'hypothèse de récurrence $H(z)$ est holomorphe au voisinage du disque unité et donc atteint son maximum en module dans ce disque en un point du cercle unité. Or $|H(z)| = 1$ sur le cercle unité. Si on suppose $Q_{n+1}(u) = 0$ en un point u du disque unité ceci implique que $|H(u)| = \frac{1}{|k_{n+1}|}$ or $|k_{n+1}| < 1$ et donc Q_{n+1} a toutes ses racines de module > 1 .

(2)

Soit b le filtre causal inverse de q . On a donc $Y = b * B$ et d'après le calcul de projection 5.9 fait pour les processus AR on a:

$$\text{Proj}(Y_{N+1} | H_N(Y)) = \sum_{k=1}^N a_k^{(N)} Y_{N-k+1}$$

Cette dernière quantité ne dépend que des variables aléatoires (Y_1, \dots, Y_N) elle est donc égale à $\text{Proj}(Y_{N+1} | (Y_1, \dots, Y_N))$. De plus $\|Y_{N+1} - \widetilde{Y}_{N+1}\|^2 = \|B_N\|^2 = p_N$. Par conséquent la matrice de covariance de Y et R_{N+1} possèdent les mêmes spécifications $(p_N, a_1^{(N)}, \dots, a_N^{(N)})$ avec $p_N \neq 0$. \square

5.4.3 Prolongement de Pisarenko d'un segment de covariance

Théorème 5.42. Soit X un p.s.c. réel. Les assertions suivantes sont équivalentes:

1. $H(X)$ est de dimension finie N .
2. La mesure spectrale μ_X est ponctuelle et charge exactement N points.
3. $\det(R_1(X)) > 0, \dots, \det(R_N(X)) > 0, \det(R_{N+p}(X)) = 0$ pour $p \geq 1$.

Démonstration. (3) \Rightarrow (1)

$\det(R_s(X)) = 0$ est équivalent à l'existence de $\lambda \in \mathbb{R}^s$ non nul tel que :

$$\sum_{i=0}^{s-1} \sum_{j=0}^{s-1} \lambda_i \lambda_j c_X(i-j) = 0$$

$$\text{Var}\left(\sum_{i=0}^{s-1} \lambda_i X_{n-i}\right) = 0 \quad \forall n \in \mathbb{Z}$$

$$\sum_{i=0}^{s-1} \lambda_i X_{n-i} = 0 \quad \forall n \in \mathbb{Z}$$

Si $\det(R_{N+1})(X) = 0$ et $\det(R_N(X)) \neq 0$ alors $\sum_{i=0}^N \lambda_i X_{n-i} = 0$ avec $\lambda_0 \neq 0$ et X_n s'exprime en fonction de $(X_{n-1}, \dots, X_{n-N})$ et donc $H(X)$ est engendré par exemple par (X_1, \dots, X_N) . La dimension de $H(X)$ ne peut être inférieure à N car alors on aurait $\det(R_N(X)) = 0$.

(1) \Rightarrow (2)

Il existe $\lambda \in \mathbb{R}^{N+1}$ non nul tel que $\sum_{k=0}^N \lambda_k X_{-k} = 0$. En utilisant l'isométrie \mathcal{T} on obtient que $\sum_{k=0}^N \lambda_k X_{n-k} = 0$ pour tout n et il existe donc un polynôme non nul P de degré $\leq N$ tel que $p * X = 0$. Cette relation implique $|\hat{p}(t)|^2 \mu_X = 0$ et donc que μ_X est portée par les racines de module 1 de P qui sont en nombre $\leq N$. Supposons que μ_X soit portée par r points. Alors il existe un polynôme q de degré r tel que $|\hat{q}(t)|^2 \mu_X = 0$ et donc $q * X = 0$. Cette dernière relation implique immédiatement $\dim(H(X)) \leq r$ et on a donc $r = N$.

(2) \Rightarrow (3)

En reprenant la preuve ci dessus on obtient un polynôme q de degré N tel que $q * X = 0$ ceci implique $\det(R_{N+1}(X)) = 0$. Si on suppose que $\det(R_N(X)) = 0$ alors $H(X)$ serait de dimension $\leq N - 1$ et donc μ serait portée par au plus $N - 1$ points. \square

Le prolongement AR d'une matrice de covariance a une mesure spectrale absolument continue. Le théorème suivant fournit un prolongement (prolongement de Pisarenko) possédant une mesure spectrale qui est la somme d'une mesure ponctuelle et d'un multiple de la mesure de Lebesgue.

Théorème 5.43. *Soit R_N une matrice de Toeplitz, σ^2 sa plus petite valeur propre et M le rang de la matrice $(R_N - \sigma^2 I)$. Il existe alors une mesure positive μ sur $[0, 1[$ qui est la somme d'exactly M masses de Dirac et de σ^2 fois la mesure de Lebesgue et dont la fonction de covariance prolonge R_N .*

Démonstration. La matrice $G = (R_N - \sigma^2 I)$ est positive et symétrique de rang M . On sait donc qu'il existe un p.s.c. Y dont la fonction de covariance prolonge G et dont la mesure spectrale μ_Y est une mesure ponctuelle chargeant exactement M points. D'autre part si l'on considère un bruit blanc B de variance σ^2 sa matrice de covariance est $\sigma^2 I$ et sa mesure spectrale μ_B est σ^2 fois la mesure de Lebesgue. En choisissant ce bruit blanc dans $H(Y)^\perp$ alors la matrice de covariance de $X = B + Y$ prolonge R_N et $\mu_X = \mu_Y + \mu_B$. Pour déterminer $\mu_Y = \sum_{k=1}^M \lambda_k \delta_{t_k}$ on peut poser $S(l, r) = \gamma_l(t_r)$ pour $l = 1 \dots N$, $r = 1 \dots M$ et on note Λ la matrice diagonale d'ordre M fabriquée avec les poids λ_k . Alors un calcul simple montre que $S \Lambda \overline{S'} = G$. \square

Chapitre 6

Statistique des processus du second ordre

Dans la pratique on a souvent affaire avec des observations d'une même quantité au fil du temps. Les modéliser avec un p.s.c. permettrait de mettre en évidence certaines particularités de son comportement et fournirait des outils pour prédire son évolution future. Les modèles *ARMA*, qui ne dépendent que d'un nombre fini de paramètres peuvent constituer un cadre convenable (il faut néanmoins se rappeler de la remarque 5.30).

Exemple 6.1. Le graphique ci dessous montre les valeurs du nombre de taches solaires observées entre 1770 et 1869. On peut tenter de modéliser ces données en disant que la

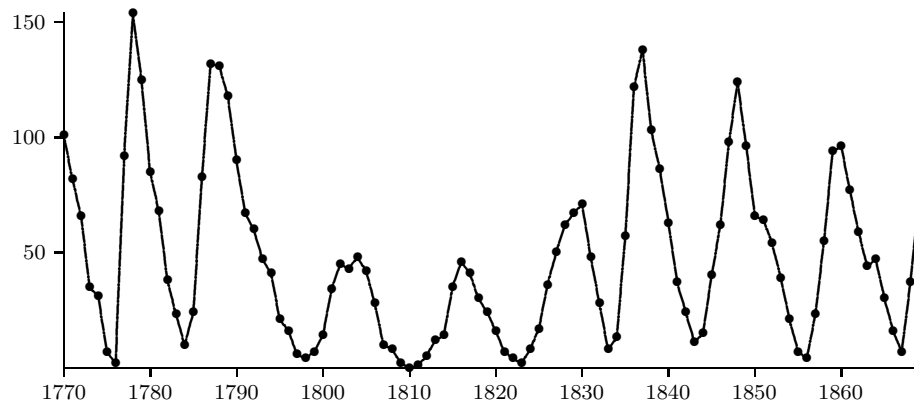


Figure 6.1 La série du nombre de taches solaires de 1770 à 1869 de Wölfer.

valeur Y_n du nombre observé dans l'année n -ième soit de la forme

$$Y_n = m + X_n$$

où m est une valeur moyenne, alors que le processus $(Y_n)_n$ est un processus *ARMA*. Dans ce but on est donc amenés de façon naturelle à la question d'estimer la valeur moyenne m , ainsi que les valeurs de p et q et, finalement des polynômes \tilde{P} et \tilde{Q} , qui caractérisent un processus *ARMA*(p, q). Une fois cette étape franchie, on pourra se servir du modèle ainsi construit pour faire de la prévision ou bien pour déduire des propriétés du phénomène. Par exemple, y a-t-il une périodicité? Quelle est cette période?

Dans ce chapitre on va aborder ces questions. Il s'agit-là d'un sujet très vaste et dont on ne verra que les premiers éléments. Néanmoins on parviendra à maîtriser un certain nombre de questions intéressantes.

6.1 Théorèmes limites

Dans ce paragraphe on se pose le problème de l'estimation des propriétés statistiques plus simples d'un processus du second ordre: de sa moyenne et de sa covariance. Comme on ne dispose le plus souvent que d'une seule réalisation et de plus sur un intervalle de temps fini, c'est pourquoi l'étude de théorèmes ergodiques fournissant la convergence presque sûre d'estimateurs de la covariance ou de la mesure spectrale est tout à fait primordiale.

On pose $M_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$.

Proposition 6.2. *Soit $(X_n)_n$ un p.s.c. La suite M_n converge dans $L^2(\Omega, \mathbb{P})$ vers $\mathcal{W}_X(\chi)$ où χ est la fonction indicatrice du point $0 \in \mathbb{T}$. De plus*

$$\|\mathcal{W}_X(\chi)\|_2^2 = \mu_X(\{0\}) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n c_X(k).$$

Si la suite $c_X(n)$ tend vers 0 en moyenne, $\mu_X(\{0\}) = 0$ et la limite de M_n est nulle.

Démonstration. On a $M_n = \mathcal{W}_X(\varphi_n)$ avec $\varphi_n = \frac{1}{n}(\gamma_1 + \gamma_2 + \dots + \gamma_n)$. Or il est clair que $\varphi_n(0) = 1$, alors que si $t \neq 0$

$$\gamma_1 + \gamma_2 + \dots + \gamma_n = \gamma_1 + \gamma_1^2 + \dots + \gamma_1^n = \gamma_1 \frac{1 - \gamma_1^n}{1 - \gamma_1}$$

Cette quantité étant bornée en n , $\varphi_n(t) \rightarrow_{n \rightarrow \infty} 0$ si $t \neq 0$. Donc φ_n converge vers χ ponctuellement. Puisque $|\varphi_n| \leq 1$, d'après le théorème de Lebesgue φ_n converge vers χ dans $L^2(\mathbb{T}, \mu_X)$ et donc $M_n \rightarrow \mathcal{W}_X(\chi)$ dans $L^2(\mathbb{T}, \mu_X)$. Puisque la convergence dans L^2 entraîne celle dans L^1 , on a, par l'égalité de Parseval,

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^n c_X(k) &= \mathbb{E}[X_0 M_n] = \langle 1, \varphi_n \rangle_{L^2(\mu_X)} = \\ &= \int_{\mathbb{T}} \varphi_n(t) d\mu_X(t) \xrightarrow{n \rightarrow \infty} \int_{\mathbb{T}} \chi(t) d\mu_X(t) = \mu_X(\{0\}) \end{aligned}$$

□

Pour $\lambda \in \mathbb{T}$ le processus $X_n^{(\lambda)} = \gamma_n(\lambda)X_n$ est un p.s.c. de covariance $c_X^{(\lambda)}(n) = \gamma_n(\lambda)c_X(n)$ et a donc pour mesure spectrale la mesure μ_X translatée par λ . On en déduit donc que si la suite $(c_X(n))_n$ tend vers 0 en moyenne, alors $\mu_X(\lambda) = 0$ pour tout $\lambda \in \mathbb{T}$ et donc la mesure spectrale μ_X est diffuse. On cherche maintenant à établir des résultats de convergence p.s. de la suite $(M_n)_n$. Dans ce but il est naturel de considérer le comportement de sa variance. Or un calcul simple donne

$$\mathbb{E}(|M_n|^2) = \frac{1}{n} \sum_{k=-n+1}^{n-1} \left(1 - \frac{|k|}{n}\right) c_X(k) \leq \frac{1}{n} \sum_{k=-n+1}^{n-1} |c_X(k)| \quad (6.1)$$

et donc, si $c_X \in \ell^1(\mathbb{Z})$,

$$\text{Var}(M_n) \leq \frac{1}{n} \|c_X\|_1. \quad (6.2)$$

Théorème 6.3. *Soit $(X_n)_n$ un p.s.c. S'il existe $K > 0$ et $\alpha > 0$ avec $\mathbb{E}(|M_n|^2) \leq K/n^\alpha$ alors la suite $(M_n)_n$ converge \mathbb{P} presque sûrement vers 0. Ceci a lieu en particulier si la suite $c_X(n)$ est sommable (on prend $\alpha = 1$).*

Démonstration. La preuve de ce théorème est tout à fait analogue à celle de la loi forte des grands nombres 2.12. Soit $\beta > 0$ tel que $\alpha\beta > 1$, on définit la sous suite n_r par $n_r = \inf\{n; n \geq r^\beta\}$. On a alors $r^\beta \leq n_r \leq r^\beta + 1$. L'inégalité sur $\mathbb{E}(|M_n|^2)$ jointe au lemme de Borel-Cantelli montre que la sous suite M_{n_r} converge presque sûrement vers 0. Il reste à contrôler les termes M_k pour $n_r \leq k \leq n_{r+1}$.

$$\begin{aligned} M_k - M_{n_r} &= \frac{X_1 + X_2 + \dots + X_k}{k} - \frac{X_1 + X_2 + \dots + X_{n_r}}{n_r} = \\ &= \left(\frac{n_r}{k} - 1\right)M_{n_r} + \frac{X_{n_r+1} + \dots + X_k}{k} \end{aligned}$$

et donc

$$|M_k - M_{n_r}| \leq |M_{n_r}| + Z_r,$$

avec

$$Z_r = \sup_{n_r \leq k \leq n_{r+1}} \left| \frac{X_{n_r+1} + \dots + X_k}{k} \right| \leq \frac{|X_{n_r+1}| + \dots + |X_{n_{r+1}}|}{n_r}$$

et donc en utilisant $\mathbb{E}(|X_i||X_j|) \leq c_X(0)$ on obtient:

$$\mathbb{E}(Z_r^2) \leq \frac{1}{n_r^2} (n_{r+1} - n_r)^2 c_X(0) \leq \left(\frac{(r+1)^\beta + 1}{r^\beta} - 1 \right)^2 c_X(0) = f(r)^2 c_X(0)$$

En utilisant le fait que $f(r) \sim \beta/r$, une application de l'inégalité de Bienaymé-Chebyshev à Z_r montre que Z_r converge presque sûrement vers 0. Il ne reste plus qu'à écrire $|M_k| \leq |M_k - M_{n_r}| + |M_{n_r}| \leq 2|M_{n_r}| + Z_r$ pour conclure. \square

On peut tenter de préciser la vitesse de convergence dans cette loi forte des grands nombres par un théorème de type théorème central limite. Celui-ci peut-être obtenu sans peine dans certains cas.

Lemme 6.4. *Soit a un filtre causal tel que la série $\sum_{n \geq 1} n|a_n|$ soit convergente. Alors la série $A(z) = \sum_{n \geq 0} a_n z^n$ (qui converge pour $|z| \leq 1$) peut s'écrire:*

$$A(z) = A(1) + (1-z)B(z)$$

où $B(z) = \sum_{n \geq 0} b_n z^n$ avec $b \in \ell^1(\mathbb{Z})$.

Démonstration. On a

$$\begin{aligned} A(z) &= \sum_{n \geq 0} a_n z^n = A(1) + \sum_{n \geq 1} a_n (z^n - 1) = A(1) - (1-z) \sum_{n \geq 1} a_n (z^{n-1} + \dots + 1) = \\ &= A(1) - (1-z) \sum_{n \geq 0} z^n \sum_{k \geq n+1} a_k = A(1) + (1-z)B(z) \end{aligned}$$

où on a posé $b_n = -\sum_{k \geq n+1} a_k$. Il reste à vérifier que la série définissant b_n est convergente et que le filtre $(b_n)_n$ est sommable, ce qui justifiera les échanges dans les sommes qu'on vient d'effectuer. Ces deux découlent, grâce au théorème de Fubini de la relation

$$\sum_{n \geq 0} \sum_{k \geq n+1} |a_k| = \sum_{k \geq 1} \sum_{n \leq k} |a_k| = \sum_{k \geq 1} k|a_k| < +\infty$$

Définition 6.5. On dit qu'un bruit blanc B_n est un bruit blanc fort, si la suite B_n est une suite de variables aléatoires indépendantes et de même loi.

Théorème 6.6. Soit $(X_n)_n$ un p.s.c. de type $X = a * Z$, où Z est un bruit blanc fort de variance σ^2 dans le cas réel, de matrice de covariance Σ dans le cas complexe (\mathbb{C} étant identifié à \mathbb{R}^2) et a est un filtre causal avec $\sum_{n \geq 0} n|a_n| < \infty$. On pose $\alpha = \sum_{n \geq 0} a_n$, alors la suite $n^{-1/2} \sum_{k=1}^n X_k$ converge en loi vers une loi gaussienne centrée, de variance $\alpha^2 \sigma^2$ dans le cas réel et de matrice de covariance $L\Sigma L'$ dans le cas complexe où l'on a posé $\alpha = u + iv$ et $L = \begin{pmatrix} u & -v \\ v & u \end{pmatrix}$.

Démonstration. D'après le lemme 6.4, si l'on pose $Y = b * Z$, alors Y est un p.s.c. et

$$X_n = (a * B)_n = A(1)Z_n + Y_n - Y_{n-1}$$

Puisque $A(1) = \alpha$, on conclut que

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n X_k = \frac{\alpha}{\sqrt{n}} \sum_{k=1}^n Z_k + \frac{1}{\sqrt{n}} (Y_n - Y_0)$$

Le théorème central limite, théorème 2.30, assure que le premier terme du membre de droite converge en loi vers la loi indiquée. Le second terme converge en probabilité vers 0 puisque, par l'inégalité de Bienaymé-Tchebishev, pour tout $\varepsilon > 0$

$$\mathbb{P}(|Y_n - Y_0| > \varepsilon \sqrt{n}) \leq \frac{\mathbb{E}[|Y_n - Y_0|^2]}{n\varepsilon^2} \leq \frac{\mathbb{E}[2|Y_n|^2 + 2|Y_0|^2]}{n\varepsilon^2} = \frac{4\mathbb{E}[|Y_0|^2]}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0$$

Il suffit alors d'appliquer la proposition 2.5. □

Remarque 6.7. • D'après 3.3 l'hypothèse faite implique que ce processus est nécessairement purement aléatoire. Dans la plupart des cas rencontrés dans la pratique, la condition de décroissance du filtre a est satisfaite, mais l'hypothèse de bruit blanc fort souvent difficile à vérifier...

• L'hypothèse de sommabilité du filtre a dans le théorème 6.6 est certainement satisfaite pour le filtre constructeur d'un modèle *ARMA*, puisque celui-ci est à décroissance exponentielle (proposition 5.29). Grâce au théorème 5.23, ce résultat s'applique à tout processus *ARMA* associé à un bruit blanc fort.

• Dans le cas réel, la quantité $\sigma^2 \alpha^2$ qui apparaît dans l'expression de la variance asymptotique peut s'exprimer en fonction des covariances. On a en effet, d'après la remarque 3.22

$$\sum_{k \in \mathbb{Z}} c_X(k) = \sigma^2 \sum_{k \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} a(n)a(n-k) = \sigma^2 \sum_{k \leq n} a(n-k) \sum_{n \geq 0} a(n)a(n-k) = \sigma^2 \alpha^2$$

6.2 Étude statistique d'un processus stationnaire

Soit X un p.s.c. On suppose que l'on n'observe pas X directement, mais un processus Y décentré de la forme

$$Y_n = m + X_n \quad (6.3)$$

où $m \in \mathbb{C}$ est inconnu. On se pose donc la question de l'estimation de la moyenne m et de la covariance c_X

6.2.1 Estimation de la moyenne

Dans le but d'estimer m , on peut considérer l'estimateur

$$\mathcal{Y}_n = \frac{1}{n} (Y_1 + \cdots + Y_n) = m + M_n \quad (6.4)$$

Le Théorème 6.3 affirme donc que, si la suite des covariances est sommable, alors $\mathcal{Y}_n \rightarrow m$ p.s. et \mathcal{Y}_n est donc un estimateur consistant de m . Grâce à la proposition 5.29 la condition de sommabilité des covariances est certainement satisfaite si $(X_n)_n$ est un processus *ARMA* associé à un bruit blanc fort. Si de plus on suppose que $(X_n)_n$ est de la forme $X = a * B$, où B_n est un bruit blanc fort et a un filtre causal avec $\sum_{n \geq 0} n|a_n| < \infty$, le Théorème 6.6 garantit que

$$\sqrt{n}(\mathcal{Y}_n - m) = \frac{1}{\sqrt{n}} M_n$$

converge en loi vers une v.a. gaussienne ce qui permet de construire des intervalles de confiance pour m (dans le cas $\alpha \neq 0$). On rappelle que, grâce au théorème 5.23 et à la proposition 5.29, les conditions du Théorème 6.6 sont satisfaites si $(X_n)_n$ est un processus *ARMA* associé à un bruit blanc fort.

6.2.2 Estimation de la covariance

Soit $(X_n)_n$ un p.s.c. Pour tout $p \in \mathbb{Z}$ on définit le processus $Z_n^{(p)} = X_{n+p} \overline{X_n}$. Si l'on dispose de n observations (X_1, X_2, \dots, X_n) on définit pour $|p| \leq n$

$$C_X^{(n)}(p) = \frac{1}{n} \sum_{j=1}^{n-p} Z_j^{(p)} \text{ si } p \geq 0, \quad C_X^{(n)}(p) = \frac{1}{n} \sum_{j=1-p}^n Z_j^{(p)} \text{ si } p \leq 0$$

On a $\mathbb{E}(Z_j^{(p)}) = \mathbb{E}(X_{j+p} \overline{X_j}) = \text{Cov}(X_{j+p}, X_j) = c_X(p)$ et donc

$$\mathbb{E}(C_X^{(n)}(p)) = \frac{n - |p|}{n} c_X(p)$$

Donc $\mathbb{E}(C_X^{(n)}(p)) \rightarrow_{n \rightarrow \infty} c_X(p)$. L'estimateur $C_X(p)$ est biaisé, mais il est asymptotiquement sans biais. Le théorème ergodique ci-dessus permet d'affirmer que $C_X^{(n)}(p)$ est un estimateur consistant de $c_X(p)$ si on sait que le processus $Z_n^{(p)}$ est stationnaire et vérifie les hypothèses de moments nécessaires. Nous allons indiquer une situation simple dans laquelle cette hypothèse est satisfaite.

Corollaire 6.8. *Soit $(X_n)_n$ un p.s.c. et p un entier tels que $(Z_n^{(p)})_n$ soit stationnaire de carré intégrable et que $\text{Var}(C_X^{(n)}(p)) \leq K/n^\alpha$ alors la suite $(C_X^{(n)}(p))_n$ converge \mathbb{P} presque sûrement vers $c_X(p)$.*

Ces conditions sont difficiles à vérifier sur un processus quelconque, mais dans le cas gaussien réel, on peut préciser les propriétés du processus $(Z_n^{(p)})_n$.

Lemme 6.9. Soit (X_1, X_2, X_3, X_4) un vecteur gaussien centré réel. Alors on a :

$$\mathbb{E}(X_1 X_2 X_3 X_4) = \mathbb{E}(X_1 X_2) \mathbb{E}(X_3 X_4) + \mathbb{E}(X_2 X_3) \mathbb{E}(X_1 X_4) + \mathbb{E}(X_1 X_3) \mathbb{E}(X_2 X_4)$$

Démonstration. Il suffit de dériver quatre fois la fonction caractéristique du vecteur (X_1, X_2, X_3, X_4) . \square

Proposition 6.10. Si $(X_n)_n$ un p.s.c. gaussien réel, alors $(Z_n^{(p)})_n$ est un processus stationnaire de covariance $c_Z(n) = c_X(n+p)c_X(n-p) + c_X^2(n)$. Par conséquent :

$$\text{Var}(C_X^{(n)}(p)) = \frac{1}{n} \sum_{k=-n+1}^{n-1} \left(1 - \frac{|k|}{n}\right) (c_X(k+p)c_X(k-p) + c_X^2(k))$$

En particulier si $c_X \in \ell^2(\mathbb{Z})$ alors $C_X^{(n)}(p) \rightarrow_{n \rightarrow \infty} c_X(p)$ en moyenne quadratique et p.s.

Démonstration. Grâce au lemme 6.9

$$\begin{aligned} \mathbb{E}[Z_n^{(p)} Z_m^{(p)}] &= \mathbb{E}[X_{n+p} X_n X_{m+p} X_m] = \\ &= \mathbb{E}[X_{n+p} X_n] \mathbb{E}[X_{m+p} X_m] + \mathbb{E}[X_n X_{m+p}] \mathbb{E}[X_{n+p} X_m] + \mathbb{E}[X_n X_m] \mathbb{E}[X_{n+p} X_{m+p}] = \\ &= c_X(p)^2 + c_X(n-m-p)c_X(n-m+p) + c_X(n-m)^2 \end{aligned}$$

Donc

$$\text{Cov}(Z_n^{(p)} Z_m^{(p)}) = \mathbb{E}[Z_n^{(p)} Z_m^{(p)}] - \mathbb{E}[Z_n^{(p)}] \mathbb{E}[Z_m^{(p)}] = c_X(n-m)^2 + c_X(n-m-p)c_X(n-m+p)$$

Cette quantité ne dépendant que de $n-m$, $(Z_n^{(p)})_n$ est un processus stationnaire et il suffit d'appliquer les résultats du paragraphe précédent, en tenant compte du fait que c_Z étant sommable, μ_Z possède une densité et donc $\mu_Z(\{0\}) = 0$. \square

Si l'on se trouve dans un cas décentré du type décrit dans (6.3), l'observation étant donnée par Y_1, \dots, Y_n on a un problème, car on n'a pas accès aux valeurs de X_1, \dots, X_n , la moyenne m étant inconnue. Il est assez naturel alors d'utiliser l'estimateur

$$C_Y^{(n)}(p) = \frac{1}{n} \sum_{j=1}^{n-p} (Y_{j+p} - \mathcal{Y}_n) (\overline{Y_j - \mathcal{Y}_n})$$

où $\mathcal{Y}_n = \frac{1}{n} \sum_{j=1}^n Y_j$. Il est facile de montrer que, du moins dans les hypothèses de la Proposition 6.10, ceci donne un estimateur consistant de $c_X(p)$. En effet on peut écrire

$$\begin{aligned} C_Y^{(n)}(p) &= \frac{1}{n} \sum_{j=1}^{n-p} (Y_{j+p} - \mathcal{Y}_n) (\overline{Y_j - \mathcal{Y}_n}) = \\ &= \frac{1}{n} \sum_{j=1}^{n-p} ((Y_{j+p} - m) + (m - \mathcal{Y}_n)) ((\overline{Y_j - m}) + (\overline{m - \mathcal{Y}_n})) = \\ &= \frac{1}{n} \sum_{j=1}^{n-p} (Y_{j+p} - m) (\overline{Y_j - m}) + (\overline{m - \mathcal{Y}_n}) \frac{1}{n} \sum_{j=1}^{n-p} (Y_{j+p} - m) + \\ &\quad + (m - \mathcal{Y}_n) \frac{1}{n} \sum_{j=1}^{n-p} (\overline{Y_j - m}) + |m - \mathcal{Y}_n|^2 = \\ &= I_1 + I_2 + I_3 + I_4 \end{aligned}$$

On a immédiatement

$$I_1 = \frac{1}{n} \sum_{j=1}^{n-p} (Y_{j+p} - m)(\overline{Y_j - m}) = \frac{1}{n} \sum_{j=1}^{n-p} X_{j+p} \bar{X}_j = \underline{C}_X^{(n)}(p)$$

Donc ce premier terme converge vers $c_X(p)$ p.s. et dans L^2 pour $n \rightarrow \infty$. Pour les autres termes, on voit facilement que

$$I_3 = (m - \mathcal{Y}_n) \frac{n-p}{n} (\overline{\mathcal{Y}_{n-p} - m})$$

Donc I_3 converge vers 0 p.s., puisque $\mathcal{Y}_{n-p} \rightarrow m$ p.s. pour $n \rightarrow \infty$. Le même argument s'applique à I_2 , alors que la convergence vers 0 de I_4 est évidente.

6.3 Estimation des paramètres d'un processus ARMA réel

6.3.1 Estimation dans les modèles AR: les équations de Yule-Walker

On suppose que $(X_n)_n$ est un processus autorégressif $AR(p)$. Il est donc déterminé par la variance σ^2 du bruit blanc ainsi que de son polynôme canonique

$$\tilde{P}(z) = 1 - \sum_{k=1}^p \alpha_k z^k,$$

On veut trouver un estimateur convenable de la variance σ^2 du bruit blanc ainsi que des coefficients $\alpha_1, \dots, \alpha_p$. Dans la remarque 5.36 on a vu que le vecteur $\alpha = (\alpha_1, \dots, \alpha_p)'$ est solution de l'équation de Yule-Walker

$$R_p \alpha = c \tag{6.5}$$

où $c = (c_X(1), \dots, c_X(n))'$ et R_p est la matrice de Toeplitz des covariances à l'ordre p

$$R_p = \begin{pmatrix} c_X(0) & c_X(1) & c_X(2) & \dots & c_X(p-1) \\ c_X(1) & c_X(0) & c_X(1) & \dots & c_X(p-2) \\ & & \dots & & \\ c_X(p-1) & c_X(p-2) & & \dots & c_X(0) \end{pmatrix}$$

Pour estimer les paramètres $\alpha_1, \dots, \alpha_p$ on peut penser de remplacer dans les équations de Yule-Walker (6.5) les covariances $c_X(k)$, $k = 0, \dots, n$ avec leurs estimateurs empiriques

$$C_X^{(n)}(k) = \frac{1}{n} \sum_{j=1}^{n-k} X_{j+k} X_j$$

Le vecteur des estimateurs, $\alpha^{(n)}$, est donc obtenu en résolvant l'équation de Yule-Walker "empirique"

$$R_p^{(n)} \alpha^{(n)} = C_{X,p}^{(n)} \tag{6.6}$$

où $R_p^{(n)}$ est du même type que R_p avec les $c_X(k)$ remplacés par les covariances empiriques $C_X^{(n)}(k)$ et $C_{X,p}^{(n)}$ désigne le vecteur colonne de composantes $C_X^{(n)}(k)$, $k = 1, \dots, p$. On peut montrer que cette équation a toujours solution. De plus la matrice R_p est inversible pour tout p , sous des hypothèses très faibles.

Proposition 6.11. *On suppose $n > p$. Alors la matrice $R_p^{(n)}$ est semi-définie positive. Elle est définie positive et donc inversible dès que $C_X^{(n)}(0) > 0$.*

Démonstration. On note T la matrice $p \times (n + p)$

$$T = \begin{pmatrix} 0 & \dots & \dots & 0 & X_1 & \dots & \dots & \dots & \dots & X_n \\ 0 & \dots & 0 & X_1 & \dots & \dots & \dots & \dots & X_n & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ X_1 & \dots & \dots & \dots & \dots & X_n & 0 & \dots & \dots & 0 \end{pmatrix}$$

Il est facile de vérifier que

$$R_p^{(n)} = \frac{1}{n} TT'$$

Donc $R_p^{(n)} \geq 0$. Finalement il est facile de voir que T , et donc aussi $R_p^{(n)}$, a rang p si et (seulement si) un au moins des X_j , $j = 1, \dots, n$ est $\neq 0$. \square

Les estimateurs de Yule-Walker sont évidemment consistants dès que les $C_X^{(n)}(k)$ sont des estimateurs consistants des covariances $c_X(k)$, car alors $R_p^{(n)} \xrightarrow{n \rightarrow \infty} R_p$ et $C_X^{(n)} \rightarrow c_X$ p.s. Le résultat suivant donne plus de précision sur le comportement asymptotique de ces estimateurs. Il est utile même dans le cas où on aurait résolu le système (6.6) avec une valeur de p qui ne serait pas la bonne.

Proposition 6.12. *Soit $(X_n)_n$ un processus $AR(p)$ stationnaire, soit $m \geq p$ et notons $\alpha \in \mathbb{R}^m$ le vecteur $\alpha = (\alpha_1, \dots, \alpha_p, 0, \dots, 0)$. Notons $\hat{\alpha}^{(n)}$ la solution de l'équation de Yule-Walker empirique dans \mathbb{R}^m soit $R_m^{(n)} \hat{\alpha}^{(n)} = C_{X,m}^{(n)}$. Alors*

$$\sqrt{n}(\hat{\alpha}^{(n)} - \alpha) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, \sigma^2 R_m^{-1}) \quad (6.7)$$

Remarque 6.13. La matrice de covariance R_m dépend de σ^2 de façon linéaire (est égale à σ^2 multiplié par la matrice de covariance pour un bruit de variance 1). Donc la matrice de covariance asymptotique $\sigma^2 R_m^{-1}$ ne dépend pas de σ^2 .

La Proposition 6.12 précise la variance asymptotique des estimateurs de Yule-Walker. Nous verrons plus tard que ces estimateurs ont la même variance asymptotique que ceux du maximum de vraisemblance. Mais, évidemment, ils ne sont utiles que lorsque on sait déjà que l'on a affaire avec un processus AR (et donc que la partie MA est nulle).

Dans le cadre d'un processus AR , la Proposition 6.12 fournit une approche à l'étude de la meilleure valeur de p . En effet, si $m > p$ et $\hat{\alpha}^{(n)} = (\hat{\alpha}_1^{(n)}, \dots, \hat{\alpha}_m^{(n)})$, (6.7) affirme que $\sqrt{n} \hat{\alpha}_m^{(n)} \rightarrow N(0, \lambda)$ pour $n \rightarrow \infty$, où λ est égal à σ^2 multiplié par l'élément diagonal d'indice m , m de la matrice R_m^{-1} . Or il est remarquable que cette quantité vaut toujours 1. En effet

Corollaire 6.14. *Dans les hypothèses de la Proposition 6.12, si $m > p$, alors*

$$\sqrt{n} \hat{\alpha}_m^{(n)} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, 1)$$

Démonstration. On a déjà remarqué que $\hat{\alpha}_m^{(n)}$ ne dépend pas de σ^2 . Soit $(X_n)_n$ un processus $AR(p)$ gaussien associé au même polynôme canonique \tilde{P} et de variance $\sigma^2 = 1$. Pour un tel processus la loi de (X_1, \dots, X_m) est gaussienne $\mathcal{N}(0, R_m^{-1})$ et, donc a densité égale à

$$f_X(x) = \frac{1}{(2\pi)^{m/2}(\det R_m)^{1/2}} \exp\left(-\frac{1}{2}x'R_m^{-1}x\right) \quad (6.8)$$

La quantité qu'on cherche à évaluer est donc le coefficient de x_m^2 dans la forme quadratique qui apparaît dans l'exposant de (6.8), multiplié par 2 et changé de signe. On remarque que $X_m = B_m + \alpha_1 X_{m-1} + \dots + \alpha_p X_{m-p}$ et que les v.a. B_m et X_{m-1}, \dots, X_1 sont indépendantes. Donc, si $p \leq m+1$, on peut appliquer le lemme 6.16 suivant et on obtient que, si on note $\tilde{x} = (x_1, \dots, x_{m-p})$, la densité f_X est aussi égale à

$$\begin{aligned} & \frac{1}{(2\pi)^{(m-1)/2}(\det R_{m-1})^{1/2}} \exp\left(-\frac{1}{2}\tilde{x}'R_{m-1}^{-1}\tilde{x}\right) \times f_{B_m}(x_m + \alpha_1 x_{m-1} + \dots + \alpha_p x_{m-p}) = \\ & = \frac{1}{(2\pi)^{m/2}(\det R_{m-1})^{1/2}} \exp\left(-\frac{1}{2}\left\{\tilde{x}'R_{m-1}^{-1}\tilde{x} + (x_m + \alpha_1 x_{m-1} + \dots + \alpha_p x_{m-p})^2\right\}\right) \end{aligned}$$

Donc, si on compare avec (6.8), on voit que le coefficient de x_m^2 dans la forme quadratique vaut $-\frac{1}{2}$, ce qui termine la preuve. \square

Remarque 6.15. On remarque que dans la preuve précédente on a aussi montré que si $\sigma^2 = 1$, on a $\det R_m = \det R_{m-1}$ pour tout $m > p$. On en déduit, pour σ^2 quelconque, que pour un modèle $AR(p)$, on a $\det R_m = \sigma^2 \det R_{m-1}$ pour tout $m > p$.

Lemme 6.16. Soit Y et Z des v.a. à valeurs \mathbb{R} et \mathbb{R}^{d-1} respectivement. On suppose $Y = W + \alpha_1 Z_{d-1} + \dots + \alpha_{d-1} Z_1$, où W est une v.a. indépendante de Z . On suppose aussi que Z et W ont des densités f_Z et f_W . Alors le couple (Z, Y) a une densité donnée par

$$f_{(Z,Y)}(z, y) = f_Z(z) f_W(y + \alpha_1 z_{d-1} + \dots + \alpha_{d-1} z_1)$$

Le corollaire 6.14 fournit un outil pour chercher la bonne valeur de p quand on veut modéliser une série avec un processus AR . On peut en effet calculer les unes après les autres les corrélations partielles empiriques $\hat{\alpha}_{m,m}$. Si la série est assez longue et $p < m$, la loi de $\hat{\alpha}_{m,m}$ est donnée, approximativement, par le corollaire 6.14. Donc $\{|\hat{\alpha}_{m,m}| \geq z_{.975}/\sqrt{n}\}$ est la région de rejet d'un test de niveau α de l'hypothèse $p < m$.

Exemple 6.17. On considère une série $(X_n)_{n \leq 1024}$ et on se demande si on pourrait la modéliser avec un processus $AR(p)$. Et aussi avoir une idée de la valeur de p . Un calcul numérique donne pour les corrélations partielles

$$\begin{aligned} \hat{\alpha}_{11} &= 0.9637 \\ \hat{\alpha}_{22} &= 0.3922 \\ \hat{\alpha}_{33} &= 0.5995 \\ \hat{\alpha}_{44} &= -0.0478 \\ \hat{\alpha}_{55} &= -0.0352 \\ \hat{\alpha}_{66} &= -0.0289 \\ \hat{\alpha}_{77} &= 0.0090 \\ \hat{\alpha}_{88} &= 0.0136 \\ \hat{\alpha}_{99} &= -0.0548 \\ \hat{\alpha}_{10,10} &= 0.0216 \\ \hat{\alpha}_{11,11} &= 0.0414 \\ \hat{\alpha}_{12,12} &= 0.0404 \end{aligned}$$

Si on compare avec le quantile à l'ordre 0.975 de la gaussienne, divisé par $\sqrt{1024}$, soit $\bar{z} = 0.0612$, on voit que, à partir de $m = 4$ la valeur de $\hat{\alpha}_m^{(n)}$ reste plus petite en module. On ne peut donc pas repousser l'hypothèse que $p \leq 3$. La figure ci-dessous montre le comportement des corrélations partielles.

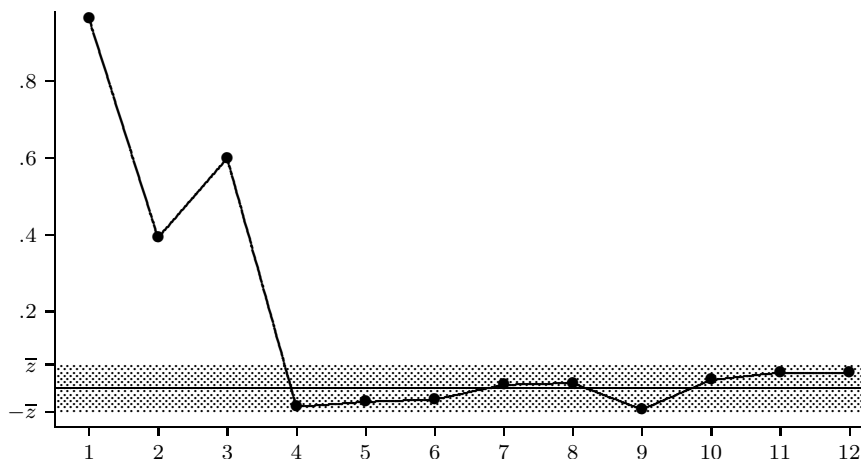


Figure 6.2 Le graphique des corrélations partielles. Les valeurs de α_{mm} qui restent dans la région ombragée ne permettent pas de repousser l'hypothèse $p \geq m$. Ce graphique suggère une modélisation avec un processus $AR(3)$.

Les données suggèrent donc l'emploi d'un modèle $AR(3)$. On peut alors résoudre les équations de Yule-Walker, ce qui donne, ici, $\bar{\sigma}^2 = 0.358$ et

$$\bar{\alpha}_1 = 0.351, \quad \bar{\alpha}_2 = 0.041, \quad \bar{\alpha}_3 = 0.6$$

La série avait été simulée avec le paramètres $\sigma^2 = 0.36$ et

$$\alpha_1 = 0.3, \quad \alpha_2 = 0, \quad \alpha_3 = 0.7$$

6.3.2 Les estimateurs du maximum de vraisemblance

En général, pour estimer les paramètres d'un modèle $ARMA(p, q)$, il est naturel de chercher à calculer l'estimateur du maximum de vraisemblance. Malheureusement la vraisemblance elle-même est ici difficile à calculer. Si on suppose que le bruit blanc est gaussien, il s'agit en effet de calculer

$$L_{\xi, \sigma^2} = \frac{1}{(2\pi)^{n/2} (\det R_n)^{1/2}} \exp\left(-\frac{1}{2} x' R_n^{-1} x\right)$$

où $\xi = (\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)$ est le vecteur des paramètres, R_n est la matrice de Toeplitz à l'ordre n et $x = (x_1, \dots, x_n)$ sont les valeurs observées de la série; n est donc le nombre d'observation, qui peut être très grand. Il s'agit donc de calculer les valeurs de ξ, σ^2 qui rendent la quantité L_{ξ, σ^2} maximum. Cette opération est ardue à cause du calcul de l'inverse de la "grosse" matrice R_n . Plusieurs méthodes numériques, notamment de type récursif, ont été développées pour mener à bien le calcul de l'estimateur et se trouvent dans tous les textes sur les séries chronologiques. Nous allons nous passer de leur description pour deux raisons. D'abord parce que tout logiciel statistique possède, déjà

programmé, ce genre de calcul. Deuxièmement parce qu'on verra plus loin un autre estimateur, plus simple à calculer, qui se trouve avoir le même bon comportement asymptotique. De toute façon on peut montrer que, pour un processus $ARMA(p, q)$, par rapport à un bruit blanc fort, l'estimateur du maximum de vraisemblance

$$(\hat{\xi}_n, \hat{\sigma}_n^2) = \operatorname{argmin}_{\xi, \sigma^2} L_{\xi, \sigma^2}$$

est consistant et qu'il existe une matrice Γ_ξ , symétrique et définie positive, telle que

$$\sqrt{n}(\hat{\xi}_n - \xi) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2 \Gamma_\xi)$$

La matrice Γ_ξ peut être précisée comme suit. Si $\xi = (\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)$, on peut considérer les deux processus *autorégressifs* $(Y_n)_n$ et $(Z_n)_n$, associés aux polynômes $\tilde{P}(z) = 1 - \alpha_1 z - \dots - \alpha_p z^p$ et $\tilde{Q}(z) = 1 + \beta_1 z + \dots + \beta_q z^q$ respectivement, par rapport au même bruit blanc $(U_n)_n$ de variance σ^2 . C'est-à-dire

$$\begin{aligned} Y_n - \alpha_1 Y_{n-1} - \dots - \alpha_p Y_{n-p} &= U_n \\ Z_n + \beta_1 Z_{n-1} + \dots + \beta_q Z_{n-q} &= U_n \end{aligned}$$

On considère les vecteurs

$$\begin{aligned} Y &= (Y_1, \dots, Y_p) \\ Z &= (Z_1, \dots, Z_q) \end{aligned}$$

alors

$$\Gamma_\xi = \begin{pmatrix} \Sigma_Y & \Sigma_{YZ} \\ \Sigma_{YZ} & \Sigma_Z \end{pmatrix}^{-1} \quad (6.9)$$

où Σ_Y et Σ_Z sont les matrices de covariance de Y et Z respectivement et Σ_{YZ} est la matrice des covariances croisées. On remarque que si les observations proviennent d'un modèle autorégressif ($q = 0$), alors

$$\Gamma_\xi = R_p^{-1}$$

R_p étant comme d'habitude la matrice de Toeplitz à l'ordre p . Une comparaison avec le théorème 6.12 montre que, dans le cas d'un processus AR , l'estimateur de Yule-Walker a une variance asymptotique identique à celle de l'estimateur du maximum de vraisemblance.

Exemple 6.18. On suppose que les observations proviennent d'un modèle $ARMA(1, 1)$ canonique. Alors $\xi = (\alpha, \beta)$ avec α, β de module < 1 . Un calcul simple montre que

$$Y_n = \sum_{k=0}^{\infty} \alpha^k U_{n-k}, \quad Z_n = \sum_{k=0}^{\infty} (-\beta)^k U_{n-k}$$

d'où, grâce à l'orthogonalité des v.a. U_n , $\mathbb{E}(Y_n^2) = \sigma^2(1 - \beta^2)^{-1}$, $\mathbb{E}(Z_n^2) = \sigma^2(1 - \alpha^2)^{-1}$ et

$$\mathbb{E}(Y_n Z_n) = \mathbb{E} \left[\sum_{k=0}^{\infty} \alpha^k U_{n-k} \sum_{j=0}^{\infty} (-\beta)^j U_{n-j} \right] = \mathbb{E} \left[\sum_{j=0}^{\infty} (-\alpha\beta)^k U_{n-k}^2 \right] = \frac{\sigma^2}{1 + \alpha\beta}$$

Donc

$$\sigma^2 \Gamma_\xi = \begin{pmatrix} (1 - \alpha^2)^{-1} & (1 + \alpha\beta)^{-1} \\ (1 + \alpha\beta)^{-1} & (1 - \beta^2)^{-1} \end{pmatrix}^{-1}$$

6.4 Estimation du spectre d'un p.s.c.

6.4.1 Estimation de la densité spectrale

Une des questions primordiales dans la théorie des processus du second ordre est l'estimation de la mesure spectrale. On le fait classiquement en utilisant le *périodogramme*.

Définition 6.19. Soit X_1, \dots, X_n des observations d'un p.s.c. X . On appelle *périodogramme* la fonction $I_n : \mathbb{T} \rightarrow \mathbb{R}^+$ définie par

$$I_n(t) = \frac{1}{n} \left| \sum_{k=1}^n X_k \gamma_{-k}(t) \right|^2$$

Proposition 6.20. Soit $(X_n)_n$ un p.s.c. tel que pour tout $p \in \mathbb{Z}$ la suite $(C_X^{(n)}(p))_n$ converge p.s. vers $c_X(p)$. Alors la suite $(\mu_n)_n$ de mesures positives sur le tore de densités $I_n(t)$ converge p.s. étroitement vers μ_X . De plus $\mathbb{E}(\mu_n(\mathbb{T})) = c_X(0)$ pour tout n .

Démonstration. Supposons par exemple $0 \leq p \leq n$. On a alors

$$\widehat{I}_n(-p) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n X_i \overline{X}_j \int_{\mathbb{T}} \gamma_{p-i+j}(t) dt$$

Puisque

$$\int_{\mathbb{T}} \gamma_{p-i+j}(t) dt = \begin{cases} 1 & \text{si } i = p + j \\ 0 & \text{sinon} \end{cases}$$

on a, si $0 \leq p \leq n-1$,

$$\widehat{I}_n(-p) = \frac{1}{n} \sum_{j=1}^{n-p} X_{p+j} \overline{X}_j = C_X^{(n)}(p)$$

Donc, si $p \geq 0$

$$\widehat{I}_n(-p) \xrightarrow{n \rightarrow \infty} c_X(p) = \widehat{\mu}_X(-p)$$

Le même argument étant valable pour $p \leq 0$, le théorème de P. Lévy 3.11 permet de conclure. \square

Un regard plus attentif à la preuve de la proposition 6.20 montre que

$$\widehat{I}_n(-p) = C_X^{(n)}(p), \quad \text{si } -(n-1) \leq p \leq n-1$$

et $\widehat{I}_n(-p) = 0$ sinon. On en déduit (théorème 1.31) l'expression alternative pour le périodogramme

$$I_n(t) = \sum_{j=-(n-1)}^{n-1} C_X^{(n)}(j) \gamma_{-j}(t)$$

Puisque la densité spectrale, si elle existe, vaut

$$g(t) = \sum_{k \in \mathbb{Z}} c_X(k) \gamma_{-k}(t)$$

on pourrait penser que le périodogramme puisse être utilisé comme estimateur de la densité spectrale. Pour étudier les qualités de l'estimateur ainsi construit on va supposer

que la fonction de covariance c_X est sommable et donc que la mesure spectrale μ_X possède une densité continue $g(x) = \hat{c}_X(x)$. On remarque que:

$$\mathbb{E}(I_n(t)) = \sum_{j=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) c_X(k) \gamma_{-k}(t) \xrightarrow{n \rightarrow \infty} \sum_{j \in \mathbb{Z}} c_X(k) \gamma_{-k}(t) = g(t)$$

Par conséquent $I_n(t)$ est un estimateur asymptotiquement sans biais de la densité spectrale $g(t)$. Ceci ne garantit pas que cet estimateur converge p.s. vers $g(t)$ ni qu'il soit convergent dans L^2 . Ceci n'est malheureusement pas le cas, comme le montre l'exemple des processus gaussiens.

Proposition 6.21. *Soit X_n un processus gaussien réel. On pose*

$$\Gamma_n(u, v) = \frac{1}{n} \sum_{k,p=1}^n c_X(p-k) \gamma_k(u) \gamma_p(v)$$

Alors

$$\text{Cov}(I_n(u), I_n(v)) = |\Gamma_n(u, -v)|^2 + |\Gamma_n(u, v)|^2$$

Démonstration.

$$\begin{aligned} \text{Cov}(I_n(u), I_n(v)) &= \frac{1}{n^2} \left(\sum_{k,k',p,p'} c(k-p') c(p-k') \gamma_{p-k}(u) \gamma_{p'-k'}(v) \right. \\ &\quad \left. + \sum_{k,k',p,p'} c(k-k') c(p-p') \gamma_{p-k}(u) \gamma_{p'-k'}(v) \right) \\ &= \frac{1}{n^2} \left(\left| \sum_{p,k'} c(p-k') \gamma_p(u) \gamma_{k'}(-v) \right|^2 + \left| \sum_{k,k'} c(k-k') \gamma_k(u) \gamma_{k'}(v) \right|^2 \right) \end{aligned}$$

□

Proposition 6.22. *Pour un processus gaussien réel on a $\text{Var}(I_n(t)) \geq (\mathbb{E}(I_n(t)))^2$*

Démonstration. On a $\Gamma_n(t, -t) = \mathbb{E}(I_n(t))$

□.

Cette proposition montre que $I_n(t)$ est un très mauvais estimateur de la densité spectrale puisque son écart type est de l'ordre de la quantité à estimer!

Si l'on suppose que le processus X est un bruit blanc gaussien de variance σ^2 , et donc de densité spectrale constante $g(t) = \sigma^2$ sur le tore $\mathbb{T} = [0, 1[$, on obtient:

$$\Gamma_n(t, -t) = \sigma^2, \quad \Gamma_n(t, t) = \frac{\sigma^2}{n} \sum_{k=1}^n \gamma_k(2t) = \begin{cases} \sigma^2 & \text{si } t = 0 \text{ ou } t = 1/2 \\ \frac{\sigma^2}{n} \gamma_n(2t) \frac{\gamma_n(2t) - 1}{\gamma_1(2t) - 1} & \text{sinon} \end{cases}$$

Par conséquent, $\mathbb{E}(I_n(t)) = \sigma^2 = g(t)$ pour tout $t \in \mathbb{T}$ et

$$\text{Var}(I_n(t)) = \begin{cases} 2\sigma^4 & \text{pour } t = 0 \text{ ou } t = 1/2 \\ \sigma^4 & \text{pour } t \text{ de la forme } \frac{k}{2n} \text{ avec } (k = 1, \dots, n-1, n+1, \dots, 2n-1) \end{cases}$$

En effet dans le premier cas on a $\gamma_1(2t) = 1$ et dans le second $\gamma_n(2t) = 1$ (et $\gamma_1(2t) \neq 1$). Pour ces dernières valeurs de t , on peut préciser la loi de $I_n(t)$. En notant

$$X' = (X_1, X_2, \dots, X_n), \quad \gamma'(t) = (\gamma_1(t), \gamma_2(t), \dots, \gamma_n(t))$$

on peut écrire $I_n(t) = \frac{1}{n} \|\gamma'(t)X\|^2$. En posant $\gamma(t) = u(t) + iv(t)$ avec u et v dans \mathbb{R}^n , la relation $\gamma'(t)\gamma(t) = 0$ est équivalente à $\|u(t)\|^2 = \|v(t)\|^2$ et $u(t)$ et $v(t)$ orthogonaux. Or $\gamma'(t)\gamma(t) = \sum_{k=1}^n \gamma_k(2t)$ et cette quantité est donc nulle pour les valeurs de t ci dessus. On en déduit que les variables aléatoires $U = u'(t)X$ et $V = v'(t)X$ sont des lois normales indépendantes de même variance égale à $n\sigma^2/2$. Il s'en suit que $I_n(t)$ suit une loi exponentielle de paramètre $1/\sigma^2$ et l'on retrouve bien les espérances et variances de $I_n(t)$ calculées plus haut. De plus, on peut remarquer que d'une part, $I_n(t)$ est très "instable" puisque $\text{Var}(I_n(0)) = 2\sigma^4$ alors que $\text{Var}(I_n(1/(2n))) = \sigma^4$ et que d'autre part il n'est pas du tout consistant puisque $I_n(0) = \sigma^2 U_n^2$ où U_n suit une loi gaussienne centrée réduite et il n'est bien sûr pas question que $I_n(0)$ puisse converger p.s. vers $g(0) = \sigma^2 \dots$

Pour construire, à partir du périodogramme, un estimateur consistant de la densité spectrale plusieurs méthodes ont été développées. Nous allons maintenant voir celle qu'on peut appeler du *lissage*. Elle consiste à remplacer I_n avec \tilde{I}_n , où $\tilde{I}_n(t)$ est obtenu en faisant une moyenne pondérée des valeurs de $I_n(s)$ pour des valeurs de s proches de t . On pose $F_n = \{t_k = k/(2n); k = 0, \dots, 2n-1\}$ et on définit t_r pour $r \in \mathbb{Z}$ quelconque par t_k où $k = r$ modulo $2n$.

Définition 6.23. On appelle *périodogramme régularisé* la fonction \tilde{I}_n définie (sur \mathbb{T}) par

$$\tilde{I}_n(t_j) = \sum w_n(k) I(t_{j-k})$$

et $\tilde{I}_n(t)$ interpolé linéairement si t est situé entre deux valeurs contigues de t_j . La suite de filtres $(w_n)_n$ satisfait aux propriétés suivantes

1. $w_n(k) = 0$ pour $|k| > m_n$, où $(m_n)_n$ est une suite telle que $m_n \rightarrow \infty$ et $m_n/n \rightarrow 0$ pour $n \rightarrow \infty$.
2. $\sum_{|k| \leq m_n} w_n(k) = 1$, $\sum_{|k| \leq m_n} w_n^2(k) \rightarrow 0$ pour $n \rightarrow \infty$.

Théorème 6.24. Soit $(X_n)_n$ un p.s.c. de la forme $X = a * B$, où B est un bruit blanc indépendant tel que $\mathbb{E}(B_j^4) < +\infty$ et le filtre a satisfait la condition

$$\sum_{j \in \mathbb{Z}} |a_j| |j|^{1/2} < +\infty \quad (6.10)$$

Alors, si \tilde{I}_n est un périodogramme régularisé, $\tilde{I}_n(t) \rightarrow_{n \rightarrow \infty} g(t)$ dans L^2 , pour tout $t \in \mathbb{T}$.

Dans la pratique on essaie de régulariser le périodogramme avec différents filtres régularisant I , jusque à ce que le graphique ne se stabilise. Les figures montrent le comportement du périodogramme et de son régularisé pour les données des taches solaires de Wölfer. On a utilisé le filtre $w_{-4} = \frac{1}{21}, w_{-3} = \frac{2}{21}, w_{-2} = w_{-1} = w_0 = w_1 = w_2 = \frac{3}{21}, w_3 = \frac{2}{21}, w_4 = \frac{1}{21}$.

6.4.2 Estimateurs des paramètres basés sur le spectre

Dans cette section on va introduire des estimateurs des paramètres d'un processus *ARMA* réel, qui sont plus simples à obtenir et qui ont les mêmes propriétés asymptotiques que les estimateurs du maximum de vraisemblance. On suppose que $(X_n)_n$ est un processus *ARMA*(p, q) réel associé au triplet canonique $(\tilde{P}, \tilde{Q}, \sigma^2)$. On pose donc

$$\begin{aligned} \tilde{P}(z) &= 1 - \alpha_1 z - \dots - \alpha_p z^p \\ \tilde{Q}(z) &= 1 + \beta_1 z + \dots + \beta_q z^q \end{aligned} \quad (6.11)$$

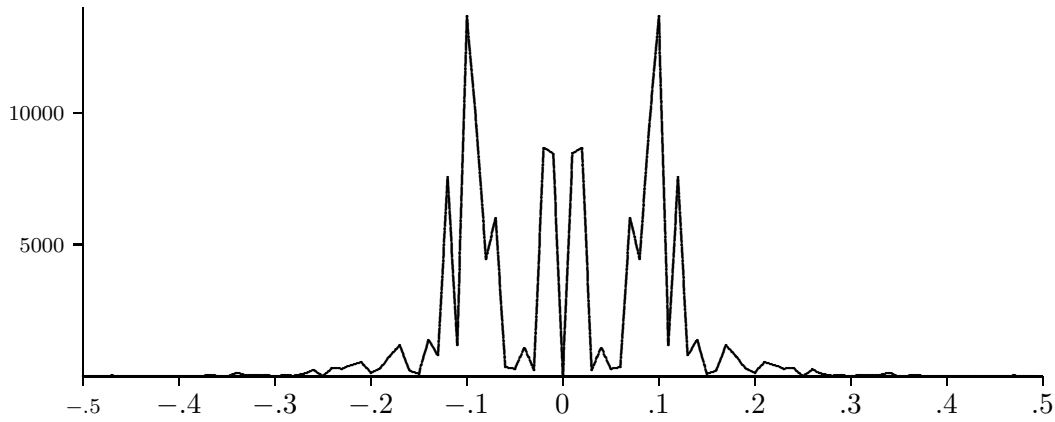


Figure 6.3 Le périodogramme pour la série des taches solaires de Wölfer.

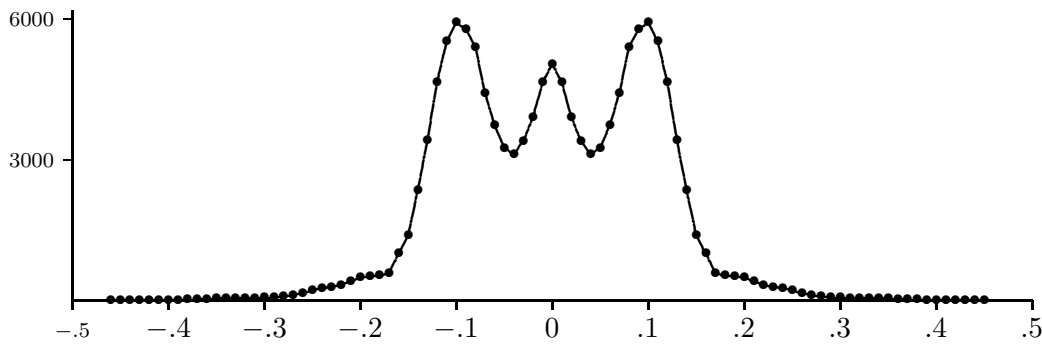


Figure 6.4 Le périodogramme régularisé pour la série des taches solaires de Wölfer. On remarque un maximum à la fréquence 0.1, ce qui suggère une périodicité de période $\frac{1}{0.1} = 10$ ans. Les petits boulets marquent les fréquences $\omega_j \in F_n$

avec les conditions

- a) $\alpha_p \neq 0, \beta_q \neq 0$.
- b) Toutes les racines de \tilde{P} et de \tilde{Q} sont de module > 1 .
- c) \tilde{P} et \tilde{Q} n'ont pas de racines communes.

On va noter $C \subset \mathbb{R}^{p+q}$ l'ensemble des vecteurs

$$\xi = (\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q) \quad (6.12)$$

satisfaisant aux propriétés a), b), c) ci-dessus. La donnée de $\xi \in C$ et $\sigma^2 > 0$ détermine donc de façon unique un processus $ARMA(p, q)$ canonique, associé au triplet $(\tilde{P}, \tilde{Q}, \sigma^2)$ déterminé grâce aux relations (6.11) et (6.12). La densité spectrale de $(X_n)_n$ peut s'écrire $\sigma^2 f(\xi, t)$, où

$$f(\xi, t) = \left| \frac{\tilde{Q}(e^{2\pi it})}{\tilde{P}(e^{2\pi it})} \right|^2$$

Proposition 6.25. Soit $\xi_0 \in C$ un vecteur fixé. Alors on a

$$\int_{\mathbb{T}} \frac{f(\xi_0, t)}{f(\xi, t)} dt > 1 \quad (6.13)$$

pour tout $\xi \in \bar{C}$, avec $\xi \neq \xi_0$.

Démonstration. Soit $(X_n)_n$ un processus $ARMA(p, q)$ avec coefficients ξ_0 et variance $\sigma^2 = 1$. Soit $\tilde{P}_{\xi_0}, \tilde{Q}_{\xi_0}$ les polynômes canoniques associés. On sait (proposition 5.24) que le prédicteur à un pas de X_{n+1} sachant $H_n(X)$ est donné par $-\sum_{j \geq 1} b_j^0 X_{n+1-j}$, où les coefficients b_j^0 sont déterminés par

$$1 + \sum_{j \geq 1} b_j^0 z^j = \frac{\tilde{P}_{\xi_0}(z)}{\tilde{Q}_{\xi_0}(z)}$$

De plus l'erreur de prédiction vaut 1. Si $\xi \in C$, avec $\xi \neq \xi_0$ et

$$1 + \sum_{j \geq 1} b_j z^j = \frac{\tilde{P}_{\xi}(z)}{\tilde{Q}_{\xi}(z)}$$

alors le prédicteur $-\sum_{j \geq 1} b_j X_{n+1-j}$ aurait une erreur de prédiction strictement plus grande que 1. C'est-à-dire

$$1 < \mathbb{E} \left[\left(X_{n+1} + \sum_{j \geq 1} b_j X_{n+1-j} \right)^2 \right] \quad (6.14)$$

On considère le processus

$$Y_n = X_n + \sum_{j \geq 1} b_j X_{n-j}$$

de façon que l'espérance mathématique qui apparaît dans (6.14) soit égale à $\mathbb{E}(Y_0^2)$. Le processus $(Y_n)_n$ est obtenu à partir de $(X_n)_n$ en appliquant le filtre causal ayant transformée en z donnée par $\tilde{P}_{\xi}/\tilde{Q}_{\xi}$. Sa densité spectrale vaut donc

$$\left| \frac{\tilde{Q}_{\xi_0}(e^{2\pi i t})}{\tilde{P}_{\xi_0}(e^{2\pi i t})} \right|^2 \left| \frac{\tilde{P}_{\xi}(e^{2\pi i t})}{\tilde{Q}_{\xi}(e^{2\pi i t})} \right|^2 = \frac{f(\xi_0, t)}{f(\xi, t)}$$

Il suffit maintenant de remarquer que la variance $\mathbb{E}(Y_0^2)$ est égale à l'intégrale sur \mathbb{T} de la densité spectrale. \square

On peut maintenant considérer, pour $\xi \in C$, la fonctionnelle

$$\Psi_n(\xi) = \frac{1}{n} \sum_{\omega_j \in F_n} \frac{I_n(t_j)}{f(\xi, t_j)} \quad (6.15)$$

où $t_j \in F_n$. Si on considère que $I_n(t_j)$ est un estimateur de $f(\xi_0, t_j)$, on peut voir Ψ_n comme une approximation de l'intégrale (6.13). On peut en effet montrer que $\Psi_n(\xi)$ converge p.s. vers l'intégrale (6.13) pour $n \rightarrow \infty$. Le vrai paramètre ξ_0 étant la valeur de ξ qui minimise (6.13), définissons $\tilde{\xi}_n$ comme la valeur de ξ qui minimise (6.15) dans \tilde{C} . La quantité dans (6.15) est parfois appelée la *vraisemblance de Whittle*. En effet l'estimateur $\tilde{\xi}_n$ jouit de très bonnes propriétés.

Théorème 6.26. *Si n est assez grand, il existe une valeur unique $\tilde{\xi}_n \in C$ telle que*

$$\tilde{\xi}_n = \operatorname{argmin}_{\xi \in C} \Psi_n(\xi)$$

De plus $\tilde{\xi}_n \rightarrow \xi_0$ p.s. et

$$\lim_{n \rightarrow \infty} \sqrt{n} (\tilde{\xi}_n - \xi_0) \stackrel{\mathcal{L}}{\underset{n \rightarrow \infty}{\rightarrow}} \mathcal{N}(0, \sigma^2 \Gamma_{\xi_0})$$

où Γ_{ξ_0} est la matrice $(p+q) \times (p+q)$, décrite dans (6.9). De plus $\lim_{n \rightarrow \infty} \Psi_n(\tilde{\xi}_n) = \sigma^2$ p.s.

En particulier l'estimateur de Whittle $\tilde{\xi}_n$ a même variance asymptotique que celui du maximum de vraisemblance.