

Séries chronologiques

Jean-Marc Azais d'après un poly de Jean-Claude Fort

22 janvier 2009

Module B0 année 2008/2009

1 Exemples.

Souvent les séries qui sont étudiées appartiennent au domaine économique et présentent donc deux caractéristiques universelles : la périodicité, en l'occurrence annuelle, et l'existence d'une tendance, à la hausse à la baisse.

Voici quelques exemples pêchés sur internet. Sur ces exemple on a d'abord fait apparaître la tendance par lissage sur une période ou une fenêtre arbitrairement choisie. On a alors désaisonnalisé empiriquement ce qui laisse un "bruit". Si ce bruit "semble" stationnaire on calcule sa fonction d'autocorrélation empirique et on constate généralement qu'il présente des dépendances traduites par des corrélations non nulles.

Le premier exemple est la production mensuelle d'énergie électrique australienne de janvier 1956 à août 1995.

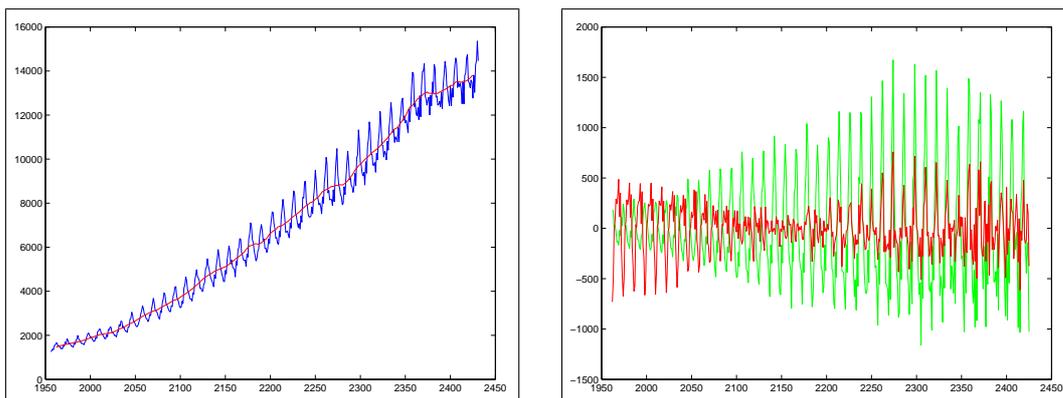


Figure 5. A gauche les données et la tendance, à droite le "bruit" avant (vert et croissant) et après désaisonnalisation (rouge) .

Le deuxième exemple, toujours en Australie, est la production de briques (de construction), les données sont trimestrielles.

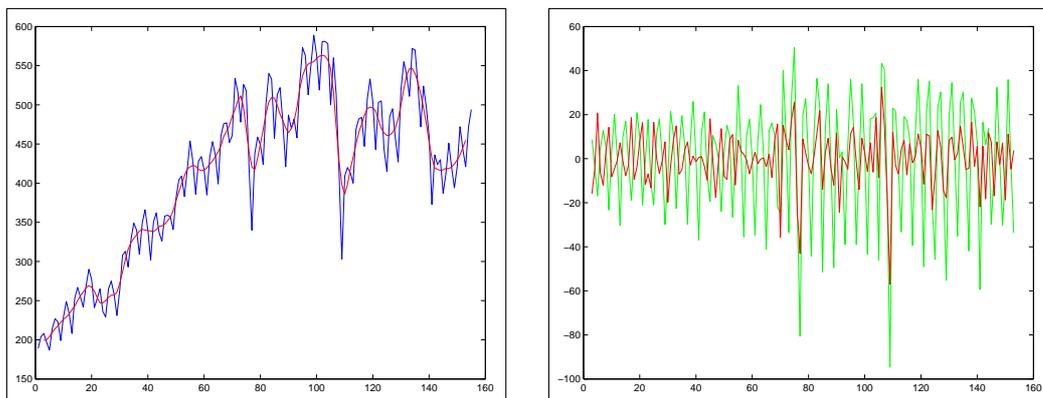


Figure 6 . A gauche les données et la tendance, à droite le "bruit" avant (vert) et après désaisonnalisation (rouge).

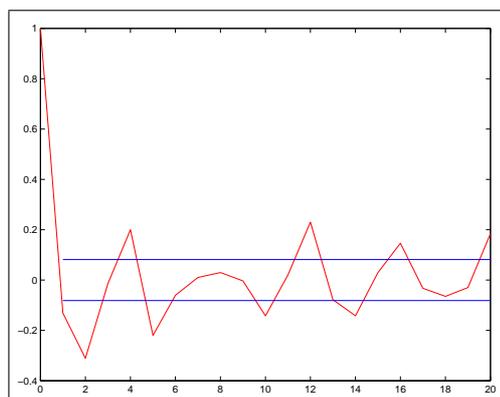


Figure 7. La fonction d'autocorrélation après suppression de la tendance et du bruit.

Le dernier exemple économique concerne les contrats de bon du Trésor des USA.

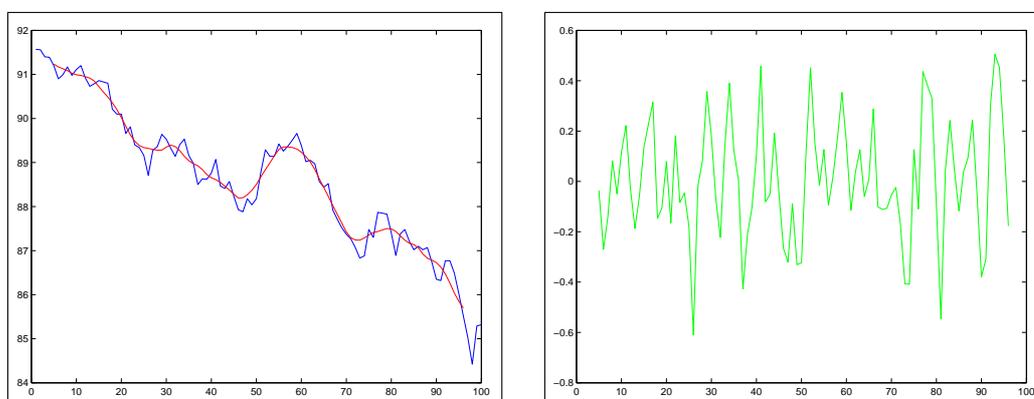


Figure 8. A gauche les données et la tendance, à droite le "bruit" .

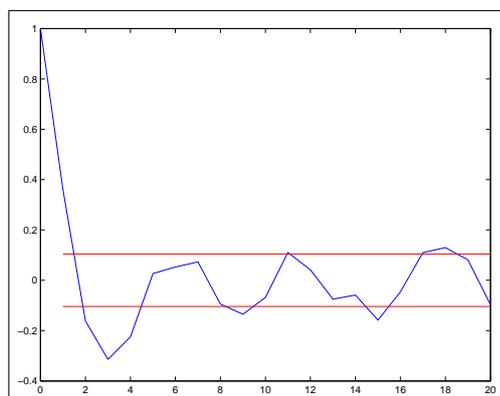


Figure 9. La fonction d'autocorrélation.

Donnons un exemple non économique : le nombre de lynx capturés par des trappeurs sur la rivière Mackenzie Nord-ouest du Canada pour la période 1821 - 1934, (Elton and Nicholson, 1942).

On donne la série brute et la série obtenue par passage au log et centrément.

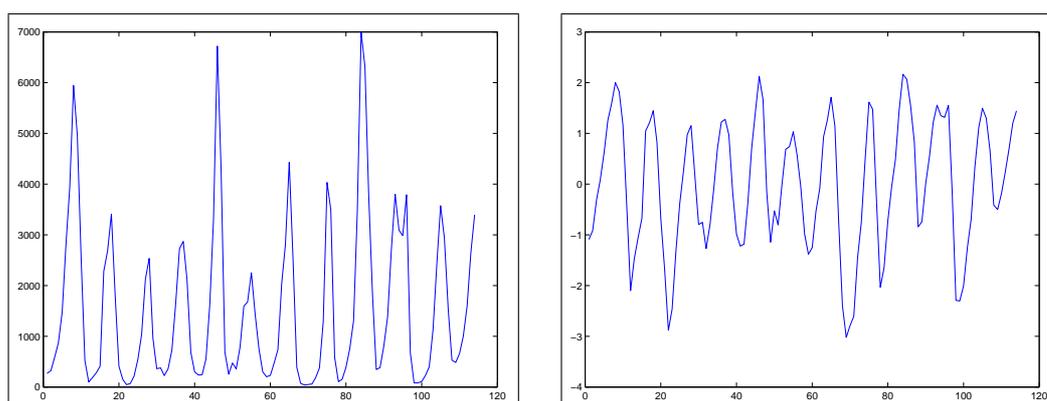


Figure 8. A gauche les données brutes, à droite après passage en log et centrément.

Le graphique montre une pseudo-périodicité pluriannuelle due à un modèle “proie-prédateur”.

L'objectif de ce qui suit est de modéliser ce “bruit” stationnaire à partir de ses caractéristiques à l'ordre 2, la modélisation conduisant à la réduction de ce “bruit” comme transformé (linéaire) d'un véritable “bruit blanc”, c'est à dire une suite stationnaire de variables décorréées.

2 Processus stationnaires au second ordre

Nous nous intéressons ici exclusivement aux processus indexés par \mathbb{Z} .

Définition 1 *Un processus stationnaire au second ordre est une suite de variables aléatoires réelles de carré intégrable $(X_i, i \in \mathbb{Z})$ dont l'espérance est constante et la covariance entre deux variables X_i et X_j ne dépend que de $|i - j|$. C'est à dire que la matrice de variance de X_{i_1}, \dots, X_{i_p} est la même que celle de $X_0, X_{i_2-i_1}, \dots, X_{i_p-i_1}$.*

Quelques exemples :

- une suite de v.a. i.i.d
- un processus gaussien de fonction de covariance $r(i, j) = \Gamma(|j - i|)$.
- un processus du type $A_1 \cos(\omega t) + A_2 \sin(\omega t)$, où A_1 et A_2 sont centrées, de même variance et non corrélées. (Dans le cas gaussien c'est un cas particulier du cas précédent).

- on se donne $(\varepsilon_i, i \in \mathbb{Z})$ une suite variables aléatoires non corrélées, alors $X_i = \varepsilon_i + b_1 \varepsilon_{i-1} + \dots + b_q \varepsilon_{i-q}$ est un processus stationnaire au second ordre.

- Dans les mêmes conditions si $(a_i, i \in \mathbb{Z})$ est une suite de carré sommable, et si $(\varepsilon_i, i \in \mathbb{Z})$ $X_i = \sum_{k=0}^{\infty} a_k \varepsilon_{i-k}$ est un processus stationnaire au second ordre.

Pour ces processus on définit la fonction d'auto-covariance, la fonction d'auto-corrélation et la fonction d'auto-corrélation partielle.

Définition 2 *On appelle fonction d'auto-covariance, $k \rightsquigarrow \Gamma(k) = \text{Cov}(X_i, X_{i+k})$. La fonction d'auto-corrélation est $k \rightsquigarrow \gamma(k) = \Gamma(k)/\Gamma(0)$. La fonction d'auto-corrélation partielle $\rho(k)$, entre X_0 et X_k est définie comme le coefficient de X_0 dans la régression de X_k sur X_0, \dots, X_{k-1} .*

Bien entendu ces quantités ne seront jamais connues mais seulement estimées. De plus calculer la corrélation partielle ne semble pas très simple. Il existe cependant un algorithme (de Durbin) qui permet de le faire récursivement à partir de la fonction d'auto-corrélation.

Commençons par une autre définition de la corrélation partielle.

Lemme 1 *Notons $P_{1,(k-1)}$ le projecteur orthogonal dans $L^2(\Omega)$ sur $\text{vect}(X_1, \dots, X_{k-1})$. L'auto-corrélation partielle $\rho(k)$ est également le coefficient de corrélation entre $X_0 - P_{1,(k-1)}X_0$ et $X_k - P_{1,(k-1)}X_k$.*

Preuve : Cela nous donne l'occasion de faire quelques calculs qui nous seront utiles par la suite. Par définition de la projection pour $\ell = 1, \dots, (k - 1)$

$$\langle X_\ell, P_{1,(k-1)}X_0 \rangle = \langle X_\ell, X_0 \rangle.$$

et donc

$$\langle X_k, P_{1,(k-1)}X_0 \rangle = \langle P_{1,(k-1)}X_k, P_{1,(k-1)}X_0 \rangle.$$

Donc, avec la nouvelle définition du lemme

$$\rho(k) = \frac{\text{Cov}(X_0 - P_{1,(k-1)}X_0, X_k)}{\sqrt{\text{Var}(X_0 - P_{1,(k-1)}X_0) \text{Var}(X_k - P_{1,(k-1)}X_k)}} = \frac{\text{Cov}(X_0 - P_{1,(k-1)}X_0, X_k)}{\text{Var}(X_0 - P_{1,(k-1)}X_0)}.$$

La dernière relation utilise un argument d'inversion du temps. Maintenant utilisons le changement de variable triangulaire suivant

$$\tilde{X}_0 := X_0 - P_{1,(k-1)}X_0 \quad (1)$$

$$\tilde{X}_1 := X_1 \quad (2)$$

$$\dots\dots \quad (3)$$

$$\tilde{X}_{k-1} := X_{k-1}. \quad (4)$$

Le fait que le changement soit triangulaire implique que les coefficients de X_0 ou \tilde{X}_0 dans la régression de X_k sont égaux. Mais comme \tilde{X}_0 est orthogonal aux autres régresseurs, son coefficient vaut

$$\frac{\langle X_k, \tilde{X}_0 \rangle}{\text{Var}(\tilde{X}_0)}$$

Ce qui donne le résultat. ■

Nous allons en déduire un algorithme numérique pour calculer les auto-corrélations partielles. Introduisons les notations suivantes :

$$- \Sigma_k = \begin{pmatrix} \Gamma(0) & \Gamma(1) & \dots & \Gamma(k-2) & \Gamma(k-1) \\ \Gamma(1) & \Gamma(0) & \dots & \Gamma(k-3) & \Gamma(k-2) \\ \Gamma(2) & \Gamma(1) & \dots & \Gamma(k-4) & \Gamma(k-3) \\ \dots & \dots & \dots & \dots & \dots \\ \Gamma(k-2) & \Gamma(k-3) & \dots & \Gamma(0) & \Gamma(1) \\ \Gamma(k-1) & \Gamma(k-2) & \dots & \Gamma(1) & \Gamma(0) \end{pmatrix} \text{ est la matrice de}$$

variance de $(X_0, \dots, X_k)^T$ (entre autres).

$$- \Gamma_k = (\Gamma(1), \dots, \Gamma(k))^T$$

$$- \overleftarrow{\Gamma}_k = (\Gamma(k), \dots, \Gamma(1))^T$$

- On pose $P_{0,(k-1)}(X_k) = \phi_{kk}X_0 + \dots + \phi_{k1}X_{k-1}$, ϕ_{kk} . ϕ_{kk} est la corrélation partielle recherchée.

$$- \Phi^k = \{\phi_{k,i}, i = 1, \dots, k\}$$

Les equations de régression donnent

$$\Sigma_k \Phi^k = \Gamma_k \quad (5)$$

On cherche un algorithme récursif pour trouver résoudre le système (5). On initialise l'algorithme par $\phi_{11} = \gamma(1)$ et on considère le passage de l'étape k à l'étape $k+1$. Nous avons à résoudre le système suivant que l'on écrit "par blocs" :

$$\begin{pmatrix} \Gamma_k \\ \Gamma(k+1) \end{pmatrix} = \begin{pmatrix} \Sigma_k & | \overleftarrow{\gamma}_k \\ (\overleftarrow{\gamma}_k)^T & | \Gamma(0) \end{pmatrix} \begin{pmatrix} Z \\ \phi_{k+1,k+1} \end{pmatrix},$$

où Z est le vecteur $\{\phi_{k+1,i}, 1 = 1, \dots, k\}$. Le système correspondant aux k premières lignes peut s'écrire

$$\Gamma_k = \Sigma_k Z + \phi_{k+1,k+1} \overleftarrow{\Gamma}_k$$

Compte tenu de la solution à l'ordre k qui s'écrit en reversant le temps

$$\overleftarrow{\Gamma}_k = \Sigma_k \overleftarrow{\Phi}_k$$

on obtient

$$Z = \Phi^k - \phi_{k+1,k+1} \overleftarrow{\Phi}_k. \quad (6)$$

On introduit ce résultat dans la dernière équation du système :

$$\Gamma(k+1) = (\overleftarrow{\Gamma}_k)^T Z + \Gamma(0) \phi_{k+1,k+1}$$

pour obtenir

$$\phi_{k+1,k+1} = \frac{(\overleftarrow{\Gamma}_k)^T \Phi^k}{(\overleftarrow{\Gamma}_k)^T \overleftarrow{\Phi}^k - \Gamma_0}$$

Les valeurs des ϕ_{k+1j} se déduisent immédiatement de l'équation (6).

Sur les deux avant-derniers exemples cela donne :

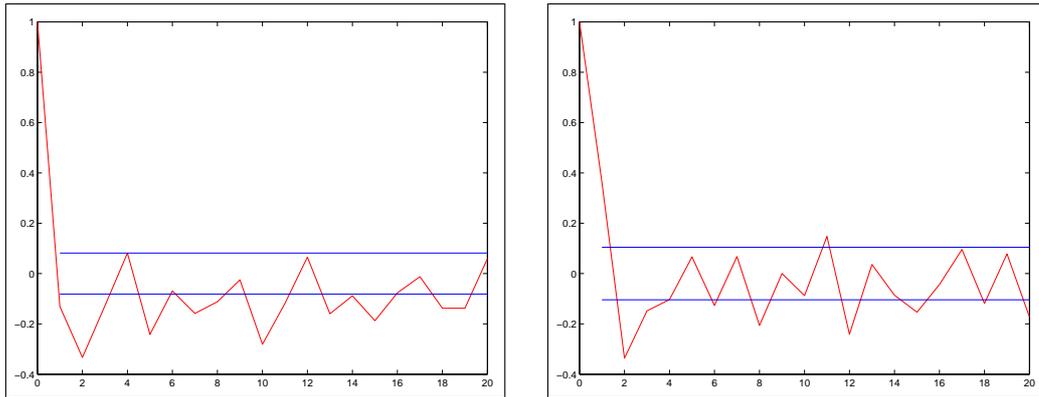


Figure 10. auto-corrélations partielles, à gauche les "briques", à droite le "trésorUS" .

Après ces quelques éléments empiriques descriptifs d'un processus stationnaire L^2 , nous allons en donner une représentation dite spectrale, car fondée sur une décomposition dans la base des fréquences pures (les $e^{it\nu}$).

3 Représentation spectrale.

Pour commencer, notons que la matrice d'auto-covariance Σ_k est semi-définie positive, ce qui se traduit par : pour tout $k+1$ -uplet non nul de nombres (prenons les complexes nous en aurons besoin) : $\sum_{j=0}^k \sum_{l=0}^k a_j \bar{a}_l \Gamma(j-l) \geq 0$.

On dit alors que la famille $\Gamma(j), j \in \mathbb{N}$ est semi-définie positive et on dispose d'un théorème de représentation de cette famille comme une série de Fourier.

Théorème 1 (Herglotz) Une famille (complexe) $\Gamma(i), i \in \mathbb{N}$ est semie définie positive si et seulement si il existe une mesure positive F sur $[-\pi, \pi]$ telle que :

$$\Gamma(h) = \int_{-\pi}^{\pi} e^{ih\nu} F(d\nu).$$

La série des $\Gamma(h)$ est donc la série de Fourier de la distribution $F(d\nu)$. La mesure $F(d\nu)$ est la mesure spectrale associée au processus X , et si elle possède une densité on parle de densité spectrale.

Remarque : On a $\text{Var}(X_t) = \int_{-\pi}^{\pi} F(d\nu)$. Ceci donne une décomposition de la variance du processus en fonction des fréquences. La mesure $F(d\nu)$ donne le poids de la fréquence $2\pi/\nu$ dans une possible décomposition (aléatoire) du processus sur les fréquences. C'est ce que nous allons voir par la suite.

Preuve : Il est clair que la formule proposée définit une famille positive (positif veut dire pour nous positif ou nul) car :

$$\sum_{l=0}^p \sum_{h=0}^p a_l \bar{a}_h \int_{-\pi}^{\pi} e^{i(l-h)\nu} F(d\nu) = \int_{-\pi}^{\pi} \left| \sum_{l=0}^p a_l e^{il\nu} \right|^2 F(d\nu) \geq 0.$$

Réciproquement posons $f_N(\nu) = \frac{1}{2\pi N} \sum_{r=0}^N \sum_{s=0}^N e^{-ir\nu} e^{is\nu} \Gamma(r-s)$. Cette fonction est positive lorsque la famille Γ est positive. Nous récrivons f_N en posant $r-s = m$:

$$f_N(\nu) = \frac{1}{2\pi N} \sum_{|m|=0}^{N-1} (N - |m|) e^{-im\nu} \Gamma(m).$$

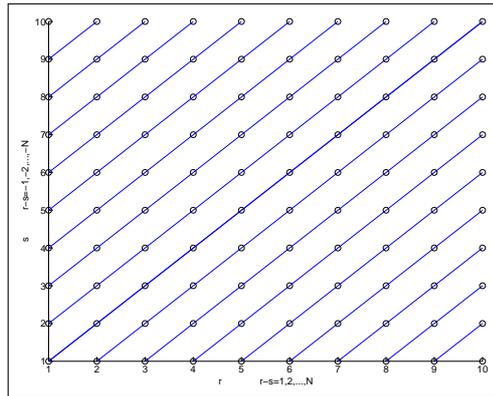


Figure 11. Le changement de variable $m=r-s$.

Posons $F_N(d\nu) = f_N(\nu)d\nu$, on a :

$$\int_{-\pi}^{\pi} e^{ih\nu} F_N(d\nu) = \frac{1}{2\pi} \sum_{|m|=0}^N (1 - \frac{|m|}{N}) \Gamma(m) \int_{-\pi}^{\pi} e^{-i(h-m)\nu} d\nu = (1 - \frac{|h|}{N}) \Gamma(h) \text{ si } |h| < N, (0 \text{ sinon.})$$

Donc en particulier

$$\int_{-\pi}^{\pi} F_N(d\nu) = \Gamma(0).$$

Ce qui montre que la masse est conservée et que la suite de mesures $F_N(d\nu)$ est donc tendue. On peut donc en extraire une sous-suite $F_{N_k}(d\nu)$ qui converge étroitement ("en loi") vers F , qui vérifie alors : $\int_{-\pi}^{\pi} e^{ih\nu} F(d\nu) = \Gamma(h)$. ■

Remarque : Une condition immédiate d'existence d'une densité spectrale est la convergence de la série des auto-covariances, car alors f_N converge uniformément (le vérifier) vers une fonction f continue, positive ou nulle qui est clairement la densité spectrale. Cela donne une formule d'inversion de Fourier :

$$f(\nu) = \frac{1}{2\pi} \sum_{m \in \mathbb{Z}} e^{im\nu} \Gamma(m) \quad (7)$$

Représentation spectrale

Considérons un processus stationnaire au second ordre (centré), sa mesure spectrale F , et construisons une isométrie T d'espaces de Hilbert complexes entre $\overline{\mathcal{H}} = \overline{\text{vect}_{\mathbb{C}}(X_t, t \in \mathbb{Z})}_{L^2}$ et $L^2(F)$ l'espace des fonctions sur $[-\pi, \pi]$ (à valeurs complexes) de carré intégrable pour $F(d\nu)$.

Pour cela on pose $T(X_t)(\nu) = e^{it\nu}$ et on prolonge par linéarité sur \mathcal{H} , puis par continuité (uniforme) sur \mathcal{H} puisque T est isométrique :

$$\begin{aligned} \|T(\sum_{j=1}^p a_j X_{t_j})\|^2 &= \int_{-\pi}^{\pi} |\sum_{j=1}^p a_j e^{it_j \nu}|^2 F(d\nu) = \int_{-\pi}^{\pi} \sum_{j=1}^p \sum_{k=1}^p a_j \bar{a}_k e^{it_j \nu} e^{-it_k \nu} F(d\nu) \\ &= \sum_{j=1}^p \sum_{k=1}^p a_j \bar{a}_k \Gamma(t_j - t_k) = \mathbb{E} \|\sum_{k=1}^p a_k X_{t_k}\|^2. \end{aligned}$$

L'application T est bien sûr injective car isométrique, mais aussi surjective car la famille des $\{\nu \rightarrow e^{it\nu}\}$ est une famille totale de $L^2(F(d\nu))$. On a donc bien l'isomorphisme annoncé.

On pose $\mathcal{I} = T^{-1}$. On définit le processus Z aléatoire sur $[-\pi, \pi]$ par : $Z(\lambda) = \mathcal{I}(\mathbb{I}_{(-\pi, \lambda]})$.

Théorème 2 *Le processus Z est un processus de carré intégrable centré et à accroissements orthogonaux. On a la représentation de X : $X_t = \int_{-\pi}^{\pi} e^{it\nu} Z(d\nu)$*

Preuve : Pour chaque λ , $Z(\lambda)$ est bien défini et se trouve dans L^2 . De plus, λ étant fixé, il existe une suite $(Y_n, n \in \mathbb{N})$ dans \mathcal{H} qui converge (L^2) vers $Z(\lambda)$. On en déduit que $\mathbb{E}(Z(\lambda)) = 0$ et pour $-\pi \leq \lambda_1 \leq \mu_1 < \lambda_2 \leq \mu_2 \leq \pi$:

$$\mathbb{E}[(Z(\mu_1) - Z(\lambda_1))\overline{(Z(\mu_2) - Z(\lambda_2))}] = \int_{-\pi}^{\pi} \mathbb{I}_{(\lambda_1, \mu_1]}(\nu) \overline{\mathbb{I}_{(\lambda_2, \mu_2]}(\nu)} F(d\nu) = 0,$$

$$\mathbb{E}|Z(\mu_1) - Z(\lambda_1)|^2 = F(\mu_1) - F(\mu_2).$$

Il est alors licite de poser $\mathcal{I}(\mathbb{I}_{(\lambda, \mu]}) = Z(\mu) - Z(\lambda) = \int_{\lambda}^{\mu} Z(d\nu)$, et pour toute fonction étagée $g = \sum_{i=1}^p a_j \mathbb{I}_{(\lambda_j, \lambda_{j+1}]}$: $\mathcal{I}(g) = \sum_{i=1}^p a_j (Z(\lambda_{j+1}) - Z(\lambda_j)) = \int_{-\pi}^{\pi} g(\nu) Z(d\nu)$. On prolonge alors par densité cette écriture pour toute fonction de $L^2(F(d\nu))$. L'application \mathcal{I} est donc une intégrale stochastique, par rapport à la mesure aléatoire $Z(d\nu)$.

Alors en se souvenant que $X_t = \mathcal{I}(e^{it \cdot})$ on obtient la représentation de X en intégrale stochastique. ■

Remarque : Pour tout couple de fonction (g, h) étagées, puis de $L^2(F(d\nu))$ par densité, on a :

$$\mathbb{E}\left(\int_{-\pi}^{\pi} g(\nu) Z(d\nu) \overline{\int_{-\pi}^{\pi} h(\nu) Z(d\nu)}\right) = \int_{-\pi}^{\pi} g(\nu) \overline{h(\nu)} F(d\nu) = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} g(\nu) \overline{h(\lambda)} \mathbb{E}[Z(d\nu) \overline{Z(d\lambda)}],$$

ce qui justifie l'écriture : $\mathbb{E}(Z(d\nu) \overline{Z(d\lambda)}) = \delta_{\lambda\nu} F(d\nu)$.

Quelques exemples :

- une suite stationnaire de *v.a.* non corrélées (bruit blanc L^2) a une densité spectrale constante.

- $X_i = U(i)$, où U est un processus d'Ornstein-Uhlenbeck de covariance $\Gamma(t) = e^{-t}$ X admet une densité spectrale : $f(\nu) = \frac{1}{2\pi} (1 - 2 \sum_{k=1}^{\infty} e^{-k} \sin k\nu)$

- Soit $X_t = A_1 \cos \nu_0 t + A_2 \sin \nu_0 t$, où A_1 et A_2 sont centrées, de même variance et non corrélées, correspond à $F(d\nu) = \frac{\sigma^2}{2} (\delta_{-\nu_0} + \delta_{\nu_0})$.

On peut lire la présence de composantes périodiques sur la représentation spectrale, et par voie de conséquence sur la mesure spectrale.

Proposition 1 Si $F(d\nu)$ admet un atome $\lambda_0 \in (-\pi, \pi)$ alors le processus X s'écrit :

$$X_t = \int_{[-\pi, \pi] \setminus \{\lambda_0\}} e^{it\nu} Z(d\nu) + (Z(\lambda_0) - Z(\lambda_0^-)) e^{it\lambda_0}.$$

On a de plus : $\text{Var}(Z(\lambda_0) - Z(\lambda_0^-)) = F(\lambda_0) - F(\lambda_0^-)$.

4 Filtrage. Processus ARMA

Un filtre invariant est simplement la donnée d'une famille h_k qui sont les poids du filtre, l'opération de filtrage faisant passer de X_t processus au second ordre à

$$Y_t = \sum_{k=-\infty}^{\infty} h_{t-k} X_k = \sum_{k=-\infty}^{\infty} h_k X_{t-k}.$$

On sait décrire simplement le passage de X à Y en terme de représentation spectrale :

Théorème 3 (Filtrage d'un processus au second ordre) Soit X_t un processus stationnaire au second ordre de mesure spectrale F_X et de représentation spectrale

$$X_t = \int_{[-\pi, \pi]} e^{it\nu} Z_X(d\nu).$$

Soit h un filtre linéaire dans $l^2(\mathbb{Z})$ (i.e. $\sum_{k=-\infty}^{\infty} |h_k|^2 < \infty$) alors le processus Y_t est un processus stationnaire au second ordre de représentation spectrale donnée par $Z_Y(d\nu) = h(e^{-i\nu})Z_X(d\nu)$, où $h(e^{-i\nu}) = \sum_{k=-\infty}^{\infty} h_k e^{-ik\nu} (\in L^2(F_X(d\nu)))$, et de mesure spectrale $F_Y(d\nu) = |h(e^{-i\nu})|^2 F_X(d\nu)$.

Si $h(e^{-i\nu})$ est p.s. non nulle on a la formule inverse :

$$X_t = \int_{[-\pi, \pi]} e^{it\nu} g(e^{-i\nu}) Z_Y(d\nu) = \int_{[-\pi, \pi]} \frac{e^{it\nu}}{h(e^{-i\nu})} Z_Y(d\nu).$$

Preuve : On a :

$$Y_t^N = \sum_{k=-N}^N h_k X_{t-k} = \int_{[-\pi, \pi]} e^{it\nu} \sum_{k=-N}^N h_k e^{-ik\nu} Z_X(d\nu).$$

La convergence $L^2(F_X(d\nu))$ permet de passer à la limite en N (dans L^2) et nous donne la représentation de Y . La définition de l'isomorphisme T conduit à la formule de la mesure spectrale. Bien sûr $Y \in \tilde{\mathcal{H}}_X$. La fonction $g(e^{-i\nu})$ est bien p.s. définie et appartient à $L^2(F_Y(d\nu))$ puisque : $\int_{[-\pi, \pi]} |g(e^{-i\nu})|^2 |h(e^{-i\nu})|^2 F_X(d\nu) = 1$. On écrit alors :

$$\|X_t - \int_{[-\pi, \pi]} g(e^{-i\nu}) e^{it\nu} Z_Y(d\nu)\| = \left\| \int_{[-\pi, \pi]} e^{it\nu} (1 - g(e^{-i\nu})h(e^{-i\nu})) Z_X(d\nu) \right\| \quad (8)$$

$$= \int_{[-\pi, \pi]} |1 - g(e^{-i\nu})h(e^{-i\nu})|^2 F_X(d\nu) = 0. \quad (9)$$

■

L'effet du filtre h est de transformer le mélange des fréquences donnant X en remplaçant la fréquence pure $e^{it\nu}$ par $h(e^{-i\nu})e^{it\nu}$, ce qui revient à multiplier l'amplitude par $|h(e^{-i\nu})|$ (amplification atténuation) et à déphaser cette fréquence pure de $\text{Arg}(h(e^{-i\nu}))$.

Il nous est maintenant commode de définir les processus ARMA.

Définition 3 i) Un filtre MA_q (moyenne mobile finie) est un filtre $h = (1, b_1, b_2, \dots, b_q)$. Un processus MA_q est le filtré d'un bruit blanc L^2 par un filtre MA_q .

ii) Un filtre AR_p est l'inverse d'un filtre MA_p . Un processus AR_p est le filtré d'un bruit blanc L^2 par un filtre AR_p .

iii) Un filtre ARMA(p, q) est le composé (commutatif) d'un filtre AR_p et d'un filtre MA_q. Un processus ARMA(p, q) est le filtré d'un bruit blanc L² par un filtre ARMA(p, q). Les formules respectives définissant ces processus sont :

$$X_t = \varepsilon_t + \sum_{k=1}^q b_k \varepsilon_{t-k}, X_t + \sum_{k=1}^p a_k X_{t-k} = \varepsilon_t, X_t + \sum_{k=1}^p a_k X_{t-k} = \varepsilon_t + \sum_{k=1}^q b_k \varepsilon_{t-k}.$$

Remarque : Un filtre MA_q est inversible si et seulement si le polynôme 1 + b₁x + ... + b_qx^q n'admet pas de racines de module 1 :

on introduit θ l'opérateur retard, défini sur l'espace des suites complexes (u_k, k ∈ ℤ) par : θ(u) = v, v_k = u_{k-1}. Un filtre est une série (convergente comme opérateur

sur l[∞](ℤ)) en θ : h(θ) = ∑_{-∞}[∞] h_kθ^k. L'ensemble de ces séries constitue une algèbre

commutative. Les filtres MA_q sont les polynômes avec coefficient constant égal à 1.

On a P(θ) = ∏_{j=1}^q (I - α_jθ), où les α_j sont les zéros (dans ℂ) de P. Un tel filtre est inversible si et seulement si tous les filtres I - α_jθ le sont, c'est-à-dire si tous les |α_j| sont différents de 1. Les inverses sont "causaux" ou "réalisables" si les |α_j| sont plus grands que 1.

Proposition 2 i) La densité spectrale d'un ARMA(p, q) défini par X_t + ∑_{k=1}^p a_kX_{t-k} =

ε_t + ∑_{k=1}^q b_kε_{t-k} où la variance de ε est σ_ε² est :

$$f_x(\nu) = \frac{\sigma_\varepsilon^2 |1 + \sum_{k=1}^q b_k e^{-ik\nu}|^2}{2\pi |1 + \sum_{k=1}^p a_k e^{-ik\nu}|^2}.$$

ii) Si f continue est la densité spectrale d'un processus réel (symétrie de f) stationnaire L², alors pour tout ε > 0 il existe un processus MA (filtre fini et inversible), X_ε, et un processus AR (fini) causal, Y_ε tels que :

$$\sup_{[-\pi, \pi]} |f(\nu) - f_{X_\varepsilon}(\nu)| \leq \varepsilon \text{ et } \sup_{[-\pi, \pi]} |f(\nu) - f_{Y_\varepsilon}(\nu)| \leq \varepsilon.$$

Preuve : Le calcul de la densité spectrale d'un ARMA est immédiat par le théorème de filtrage. Le reste est une conséquence directe du résultat suivant :

Théorème 4 Si f continue est la densité spectrale d'un processus réel (symétrie de f) stationnaire L² alors pour tout ε > 0 il existe un entier q et un polynôme réel de

degré q, b(z) = 1 + b₁z + ... + b_qz = ∏_{j=1}^q (1 - $\frac{z}{\beta_j}$) tel que :

$$\left| \frac{(\int_{-\pi}^{\pi} f(u) du) |b(e^{-i\nu})|^2}{2\pi(1 + b_1^2 + \dots + b_q^2)} - f(\nu) \right| \leq \varepsilon.$$

Ce théorème est fondé sur la convergence Césaro uniforme de la série de Fourier d'une fonction continue et périodique sur $[-\pi, \pi]$. Nous l'admettons faute de temps.

■

L'intérêt du résultat que nous venons *de facto* d'admettre est la densité des processus ARMA parmi les processus réels à densité spectrale continue. D'où le fait de se contenter le plus souvent pour modéliser un processus stationnaire qui semble "régulier" d'utiliser des ARMA. En effet nous avons vu qu'une singularité dans la répartition spectrale correspond à la présence d'une partie fréquence pure, qui hormis son amplitude, n'a rien d'aléatoire. Pour confirmation nous avons le théorème de Wold. Posons $\mathcal{H}_t = \text{vect}(X_u, u \leq t)$, $\mathcal{H}_{-\infty} = \bigcap_{t=-\infty}^{\infty} \mathcal{H}_t$, et P_U la projection orthogonale sur le sous-espace vectoriel U .

Théorème 5 (Wold) *Soit X un processus stationnaire L^2 alors il admet la décomposition unique suivante :*

$$X_t = U_t + V_t, \quad U_t = \sum_{j=0}^{\infty} h_j Z_{t-j},$$

avec

i) $h_0 = 1$ et h est une suite de l^2 ,

ii) Z est un bruit blanc L^2 de variance $\sigma^2 = \text{var}(X_t - P_{\mathcal{H}_{t-1}} X_t)$ et tel que $Z_t \in \mathcal{H}_t$.

iii) $V \in \mathcal{H}_{-\infty}$, V est déterministe ($V_t = P_{\mathcal{H}_{t-1}} V_t$ p.s.), et Z et V sont orthogonaux.

Preuve : elle repose en réalité sur le théorème de Pythagore, ce qui nous ramène quelques années en arrière. On prend très naturellement $Z_t = X_t - P_{\mathcal{H}_{t-1}} X_t$, $h_j = \frac{\langle X_t, Z_{t-j} \rangle}{\sigma^2}$ et $V_t = X_t - \sum_{j=0}^{\infty} h_j Z_{t-j}$.

Pour cela il faut tout de même noter que σ^2 ne dépend pas de t . On le voit en écrivant que $X_j, j \leq t$ est une famille totale de \mathcal{H}_t et que par conséquent il existe une suite dans l^2 telle que $P_{\mathcal{H}_{t-1}} X_t = \sum_{j=1}^{\infty} a_j X_{t-j}$. Les a_j dépendent a priori de t mais sont entièrement déterminés par les équations de Yule-Walker :

$$\Gamma(l) - \sum_{j=1}^{\infty} a_j \Gamma(l-j) = 0,$$

équations qui ne dépendent pas de t .

Ensuite par définition il est clair que les Z_t forment un bruit blanc, donc une base orthogonale de l'espace qu'ils engendrent. D'autre part les Z_t sont tous orthogonaux à $\mathcal{H}_{-\infty}$. Donc $\mathcal{H}_{-\infty} \subset \overline{\text{vect}(Z_t, t \in \mathbb{Z})}^{\perp}$. Réciproquement si $v \in \overline{\text{vect}(Z_t, t \in \mathbb{Z})}^{\perp}$, on a $v \in \mathcal{H}_{\infty} \cap \overline{\text{vect}(Z_t, t \in \mathbb{Z})}^{\perp}$, donc v appartient à tous les \mathcal{H}_t (car $\mathcal{H}_t = \text{vect}(Z_t, Z_{t-1}, \dots, Z_{t-k}) \oplus^{\perp} \mathcal{H}_{t-k-1}$) et donc se trouve dans $\mathcal{H}_{-\infty}$. Par conséquent $\mathcal{H}_{\infty} = \overline{\text{vect}(Z_t, t \in \mathbb{Z})} \oplus^{\perp} \mathcal{H}_{-\infty}$.

Il est clair que les $h_j = \frac{\langle X_t, Z_{t-j} \rangle}{\sigma^2}$ sont bien indépendants de t par stationnarité (même méthode que pour les a_j). On a bien $h_0 = \frac{\langle X_t, Z_t \rangle}{\sigma^2} = \frac{\|X_t - P_{\mathcal{H}_{t-1}} X_t\|^2}{\sigma^2} = 1$.

Comme V_t est orthogonal aux $Z_s, s \leq t$ et évidemment aux $Z_s, s > t$, il vient $\overline{\text{vect}(V_t, t \in \mathbb{Z})} \subset \mathcal{H}_{-\infty}$. Réciproquement, on voit que $\mathcal{H}_t = \overline{\text{vect}(Z_j, j \leq t)} \oplus^\perp \overline{\text{vect}(V_j, j \leq t)}$, par conséquent si $Y \in \mathcal{H}_{-\infty}$, comme $\langle Y, Z_t \rangle = 0$ pour tout t , on a $Y \in \overline{\text{vect}(V_j, j \leq t)}$ ce qui permet d'affirmer que $\mathcal{H}_{-\infty} \subset \overline{\text{vect}(V_t, t \in \mathbb{Z})}$. Ceci montre que V est déterministe puisqu'il est équivalent de dire qu'il est dans $\mathcal{H}_{-\infty}$.

Reste l'unicité. Pour cela on observe que tout processus Z vérifiant les propriétés indiquées vérifie Z_t orthogonal à \mathcal{H}_{t-1} . En utilisant alors l'expression de X_t et en projetant sur \mathcal{H}_{t-1} on obtient $Z_t = X_t - P_{\mathcal{H}_{t-1}} X_t$, par produit scalaire on obtient l'expression de h_j , puis $V_t = X_t - \sum_{j=1}^{\infty} h_j Z_{t-j}$. Ce qui montre bien l'unicité presque sûre. ■

Remarque : on dit que Z est l'innovation du processus X .

On définit le processus $U_t = \sum_{j=1}^{\infty} h_j Z_{t-j}$. On dit que U est la partie régulière de X et V la partie singulière. Si $V = 0$, X est dit régulier. On a le théorème de décomposition suivant conséquence du précédent :

Théorème 6 *La mesure spectrale de X , F_X , admet la décomposition : $F_X = F_U + F_V$, où F_U est la mesure spectrale de U et est absolument continue par rapport à la mesure de Lebesgue et F_V est la mesure spectrale de V et est étrangère à la mesure de Lebesgue. On obtient donc la décomposition de Lebesgue de la mesure spectrale.*

Preuve très partielle : par orthogonalité $F_X = F_U + F_V$. La densité spectrale de U est $\frac{\sigma^2 |h(e^{-i\lambda})|^2}{2\pi}$. En ce qui concerne F_V , il faut démontrer qu'elle n'admet pas de partie absolument continue. Nous l'admettons. ■

5 Estimation.

5.1 Estimation des paramètres au second ordre .

Nous commençons par un résultat qui apparaît comme élémentaire et qui pourtant nécessite déjà un outillage non négligeable.

Dans toute la suite nous supposons que le processus stationnaire au second ordre qui nous intéresse s'écrit (il est régulier) :

$$X_t = \mu + \sum_{-\infty}^{\infty} \varphi_j \varepsilon_{t-j},$$

où ε est un bruit blanc *i.i.d.* centré de variance σ_ε^2 , φ un filtre linéaire de l^1 (sommable) (il est donc dans l^2) et de somme non nulle.

Le premier résultat (illustratif) est le suivant :

Théorème 7 (*Normalité asymptotique de la moyenne empirique !*)

Soit $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$ alors :

$$\sqrt{n}(\bar{X}_n - \mu) \implies \mathcal{N}(0, v), \text{ où } v = \left(\sum_{-\infty}^{\infty} \varphi_j \right)^2 = \sum_{-\infty}^{\infty} \Gamma(j).$$

La preuve (comme toute celles qui suivront) consiste à démontrer le résultat pour un filtre à portée finie puis à l'étendre au cas d'un filtre quelconque. Cela repose sur le lemme général suivant :

Lemme 2 soit $(U_n, n \in \mathbb{N})$ et $(V_{nk}, n \in \mathbb{N}, k \in \mathbb{N})$, des suites de vecteurs aléatoires vérifiant :

- a) pour tout k $V_{nk} \implies_{n \rightarrow \infty} V_k$,
- b) $V_k \implies_{k \rightarrow \infty} V$,
- c) $\forall \eta > 0, \lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} P(\|U_n - V_{nk}\| > \eta) = 0$.

Alors U_n converge en loi vers V quand n tend vers l'infini.

Preuve : On note ϕ_Z la fonction caractéristique d'une v.a. Z . Alors :

$$|\phi_{U_n} - \phi_V| \leq |\phi_{U_n} - \phi_{V_{nk}}| + |\phi_{V_{nk}} - \phi_{V_k}| + |\phi_{V_k} - \phi_V|.$$

Quand n et puis k tendent vers l'infini dans cet ordre, les deux derniers termes du second membre tendent vers 0 par a) et b). La condition c) permet de montrer que le premier terme tend aussi vers 0. ■

Preuve du théorème :

Supposons pour l'instant le filtre φ de portée m finie, i.e. $\varphi_j = 0$ pour $|j| > m$. Sans perte de généralité nous supposons aussi par commodité $\mu = 0$.

Soit $X_1 + \dots + X_\ell$, un calcul très rapide montre que sa variance vaut

$$R_\ell := \ell\Gamma(0) + 2(\ell - 1)\Gamma(1) + \dots + 2\Gamma(\ell)$$

Soit maintenant $k > 2m$ et $r = \lfloor \frac{n}{k} \rfloor$, posons :

$$V_{nk} = \frac{(X_1 + \dots + X_{k-2m}) + (X_{k+1} + \dots + X_{2k-2m}) + \dots + (X_{(r-1)k+1} + \dots + X_{rk-2m})}{\sqrt{n}}.$$

Les r termes de cette somme étant indépendants, par le théorème de la limite centrale V_{nk} converge en loi, quand $n \rightarrow \infty$ vers une $\mathcal{N}(0, \frac{R_{k-2m}}{k})$. Comme

$$\frac{R_{k-2m}}{k} \rightarrow \sum_{|j| \leq m} \Gamma(j) \text{ quand } k \rightarrow \infty,$$

il reste à vérifier ici l'hypothèse c) du lemme. Mais :

$$\sqrt{n}\bar{X}_n - V_{nk} = \frac{\sum_{j=1}^{r-1} (X_{jk-2m+1} + \dots + X_{jk})}{\sqrt{n}} + \frac{X_{kr-2m+1} + \dots + X_n}{\sqrt{n}}.$$

C'est une somme de r termes indépendants, de variance

$$\frac{(r-1)R_{2m} + R_{n-k\lfloor \frac{n}{k} \rfloor + 2m}}{n}.$$

On a donc $\limsup_{n \rightarrow \infty} \text{Var}(\sqrt{n}\bar{X}_n - V_{nk}) = \frac{R_{2m}}{k}$. L'inégalité de Bienaimé-Tchébitcheff nous donne c) et nous permet d'appliquer le lemme.

Pour finir, une seconde application du lemme nous fait passer du cas à portée finie au cas général : on pose

$$V_{nm} = \frac{\sum_{t=1}^n \sum_{-m}^m \varphi_j \varepsilon_{t-j}}{\sqrt{n}}.$$

Nous venons de voir que V_{nm} converge en loi vers $V_m \sim \mathcal{N}(0, \sigma_\varepsilon^2 (\sum_{-m}^m \varphi_j)^2)$ quand $n \rightarrow \infty$. De plus, quand $m \rightarrow \infty$ V_m converge en loi vers une variable $\mathcal{N}(0, \sigma_\varepsilon^2 (\sum_{-\infty}^{\infty} \varphi_j)^2)$.

Enfin :

$$\begin{aligned} \text{Var}(\sqrt{n}\bar{X}_n - V_{nm}) &= n \times \text{Var}\left(\frac{\sum_{t=1}^n \sum_{|j|>m} \varphi_j \varepsilon_{t-j}}{n}\right) \\ &= \frac{1}{n} \sum_{0 \leq |k| \leq n-1} (n - |k|) \sum_{|j|>m, |j'|>m} \varphi_j \varphi_{j'} \mathbb{E} \varepsilon_j \varepsilon_{j'+k} \rightarrow \left(\sum_{|j|>m} \varphi_j\right)^2 \sigma_\varepsilon^2. \end{aligned} \quad (10)$$

On a donc

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \text{Var}(\sqrt{n}\bar{X}_n - V_{nm}) = 0,$$

ce qui conclut via l'inégalité de Bienaimé-Tchébitcheff. ■

Nous passons maintenant aux résultats plus sérieux concernant l'estimation des auto-covariances et auto-corrélations. Nous avons le théorème :

Théorème 8 (*Normalité asymptotique des auto-covariances et auto-corrélations empiriques*) Soit

$$\hat{\Gamma}(h) = \frac{\sum_{i=1}^{n-h} (X_i - \bar{X}_n)(X_{i+h} - \bar{X}_n)}{n}$$

et

$$\hat{\gamma}(h) = \frac{\hat{\Gamma}(h)}{\hat{\Gamma}(0)}.$$

Posons $\hat{\Gamma}_h = (\hat{\Gamma}(0), \dots, \hat{\Gamma}(h))^T$ et $\hat{\gamma}_h = (\hat{\gamma}(1), \dots, \hat{\gamma}(h))^T$.

Alors si ε admet un moment d'ordre 4 : $\mathbb{E} \varepsilon_t^4 = \kappa \sigma_\varepsilon^4$,

i) $\sqrt{n}(\hat{\Gamma}_h - \Gamma_h) \implies \mathcal{N}(0, V_h)$, où l'élément p, q de V_h vaut

$$V_h^{pq} = (\kappa - 3)\Gamma(p)\Gamma(q) + \sum_{k=-\infty}^{\infty} \Gamma(k)\Gamma(k+p-q) + \Gamma(k+q)\Gamma(k-p).$$

ii) $\sqrt{n}(\hat{\gamma}_h - \gamma_h) \implies \mathcal{N}(0, W_h)$, où l'élément p, q de W_h vaut

$$W_h^{pq} = \sum_{k=-\infty}^{\infty} \gamma(k+p)\gamma(k+q) + \gamma(k-p)\gamma(k+q) \\ + 2(\gamma(p)\gamma(q)\gamma(k)^2 - \gamma(p)\gamma(k)\gamma(k+q) - \gamma(q)\gamma(k)\gamma(k+p)).$$

Preuve : On commence par le passage de i) à ii). Pour cela il suffit d'appliquer la δ -méthode en écrivant que $\hat{\gamma}_h = g(\hat{\Gamma}_h)$, où $g(x_0, x_1, \dots, x_h) = (\frac{x_1}{x_0}, \dots, \frac{x_h}{x_0})$. Par suite il est connu que $\hat{\gamma}_h$ est asymptotiquement normal centré sur γ_h et de matrice de covariance $DV_h D^T$, où D est la matrice du gradient de g ,

$$D = \frac{1}{\Gamma(0)} \begin{pmatrix} -\gamma(1) & 1 & 0 & \cdot & \cdot & 0 \\ -\gamma(2) & 0 & 1 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -\gamma(h) & 0 & \cdot & \cdot & 0 & 1 \end{pmatrix}.$$

Le résultat annoncé en découle immédiatement.

Reste le cas des auto-covariances. Le processus est long en voici les étapes :

- En supposant le processus centré, considérer $\Gamma^*(h) = \frac{\sum_{t=1}^n X_t X_{t+h}}{n}$ (non calculable).
- Montrer le résultat pour Γ_h^* dans le cas d'un filtre à portée finie m , auquel cas le processus vectoriel $Y_t := (X_t^2, X_t X_{t+1}, \dots, X_t X_{t+h})^T$ est $2m + h$ dépendant. On applique alors le lemme à toute combinaison linéaire des éléments de Y_t .
- Passer du cas fini au cas général. C'est encore une application du lemme.
- Passer de Γ_h^* à $\hat{\Gamma}_h$. On remarque que :

$$\sqrt{n}(\Gamma_h^* - \hat{\Gamma}_h) = \sqrt{n}\bar{X}_n \left(\frac{\sum_{t=1}^{n-h} X_{t-h}}{n} + \frac{\sum_{t=1}^{n-h} X_t}{n} + \frac{(n-h)\bar{X}_n}{n} \right) + \frac{\sum_{t=n-h+1}^n X_t X_{t-h}}{\sqrt{n}}$$

Le second terme tend presque sûrement vers 0. Le premier terme s'écrit $\sqrt{n}\bar{X}_n \times P_n$, et on voit sans difficulté (toujours dans le cas $\mu = 0$) que P_n converge *p.s.* vers 0. Par conséquent l'ensemble converge en probabilité vers 0.

- il reste à passer de $\mu = 0$ au cas d'une espérance quelconque, ce qui est très facile.

Les étapes 2 et 3 nécessitent quelques calculs (en particulier celui de $\text{Cov}(\Gamma^*(p), \gamma^*(q))$), mais reposent sur les mêmes techniques que la normalité asymptotique de la moyenne empirique. ■

5.2 Spectrogramme

X_1, \dots, X_n représente les observations, le plus souvent réelles, mais que l'on considérera comme appartenant à \mathbb{C}^n . On rappelle que si $(a_j)_j$ est une suite de n

complexes, sa transformée de Fourier discrete DFT est la suite $(b_k)_k$ appartenant à \mathbb{C}^n définie par

$$b_k = \sum_{j=1}^n a_j \exp\left(\frac{i2\pi jk}{n}\right), \quad k = 1, \dots, n.$$

Il est bien connu qu'il existe un algorithme rapide, l'algorithme FFT pour effectuer ce calcul qui est une approximation de la transformée de Fourier continue.

Soit donc $(b_k)_k$ la DFT de X_1, \dots, X_n , alors, dans un premier temps le spectrogramme est défini aux **fréquences de Fourier** $\omega_k := 2\pi k/n$ par

$$I_n(\omega_k) := \frac{1}{n} |b_k|^2.$$

En fait on n'a besoin des valeurs de $I_n(\omega_k)$ que pour k tel que $\omega_k \in]-\pi, \pi[$ mais en fait la formule se prolonge à tout k par périodicité. De cette relation on déduit immédiatement que

$$\|X\|^2 = \sum_{k=1}^n I_n(\omega_k)$$

ce qui donne une sorte "d'analyse de la variance" de la série en fonction des fréquences.

Dans un second temps le spectrogramme est prolongé en une fonction périodique de période 2π "au plus proche voisin" c'est à dire que $I_n(\omega) = I_n(\omega_j)$ où ω_j est la fréquence de Fourier la plus proche.

Convergence du spectrogramme

On suppose dans cette partie que la série X_1, \dots, X_n, \dots est centrée et que la suite des covariances est absolument sommable. Nous avons déjà vu que cela implique que X admet une densité spectrale f continue qui est donné par

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

On a

$$I_n(0) = n \overline{X}^2,$$

et plus généralement pour $\omega_k \neq 0$

$$I_n(\omega_k) = \frac{1}{n} \left[\sum_{|j| \leq n} \widehat{\Gamma}(j) e^{ij\omega_k} \right].$$

Un estimateur naturel de la densité spectrale est donc $\frac{1}{2\pi} I_n(\omega)$

Proposition 3 *Sous les hypothèses ci-dessus, $\mathbb{E}(I_n(\omega)) \rightarrow 2\pi f(\omega)$ uniformément*

Preuve : On prend l'espérance dans la formule ci dessus, pour ω_k fréquence de Fourier

$$\mathbb{E}(I_n(\omega_k)) = \sum_{|j| \leq n} (1 - |j|/n) \Gamma(j) e^{ij\omega_k}.$$

Si maintenant ω est quelconque, n tend vers l'infini et $\omega_{k,n}$ est la fréquence de Fourier la plus proche (qui dépend de n)

$$\mathbb{E}(I_n(\omega_k)) - 2\pi f(\omega) \leq \left| \sum_{|j| \leq n} (|j|/n) \Gamma(j) e^{ij\omega_{k,n}} \right| + 2\pi |f(\omega) - f(\omega_{k,n})|.$$

La convergence dominée montre que le premier terme tend vers zero, La continuité de f permet de régler le cas du second. ■