

SIGMA205 : Prédiction d'une série temporelle localement stationnaire

Thomas EBOLI et Amaury DURAND

21 juin 2016

Table des matières

Chapitre 1

Rappels sur les processus AR

Equation AR(p) : On s'intéresse à l'équation AR(p) définie pour $p \geq 1$ et $(X_t)_{t \in \mathbb{Z}}$ un processus aléatoire à valeurs dans \mathbb{C} par

$$X_t = \sum_{i=1}^p \theta_i X_{t-i} + \epsilon_t \quad (\text{AR})$$

où $\theta_1, \dots, \theta_p \in \mathbb{C}$ et $\epsilon \sim BB(0, \sigma^2)$ (i.e ϵ est un processus stationnaire au second ordre, centré de fonction d'autocovariance γ vérifiant $\gamma(0) = \sigma^2 > 0$ et $\forall s \neq 0, \gamma(s) = 0$)

1.1 Construction d'une solution stationnaire au second ordre

Théorème 1.1.1. Soit $\Theta(z) = 1 - \sum_{k=1}^p \theta_k z^k$, supposons que $\forall |z| = 1, \Theta(z) \neq 0$
Alors il existe une solution stationnaire au second ordre de (AR)

Démonstration. Pour prouver cela on a besoin du lemme suivant :

Lemme 1.1.1. Pour $\alpha \in \ell^1$ et X un processus tel que $\sup_t \mathbb{E}[|X_t|] < +\infty$ on appelle filtrage de X le processus :

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

Alors si $\alpha, \beta \in \ell^1$ on a $F_\alpha(F_\beta(X)) = X$ si $\alpha \star \beta = \delta$
De plus $\alpha \star \beta = \delta \iff \forall |z| = 1 \left(\sum_{k \in \mathbb{Z}} \alpha_k z^k \right) \left(\sum_{k \in \mathbb{Z}} \beta_k z^k \right) = 1$
Enfin si X est stationnaire au second ordre $F(X)$ l'est aussi.

Remarque. $F_\alpha = \sum_{k \in \mathbb{Z}} \alpha_k B^k$ où B est l'opérateur de shift $B : (x_t)_{t \in \mathbb{Z}} \mapsto (x_{t-1})_{t \in \mathbb{Z}}$

Le polynôme Θ peut s'écrire : $\Theta(z) = \prod_{k=1}^p (1 - u_k z)$ où les u_k sont les inverses des racines de Θ . L'équation (AR) s'écrit $\Theta(B)(X) = \epsilon$. Or avec la factorisation obtenue on a :

$$\Theta(B) = (1 - u_1 B) \circ \dots \circ (1 - u_p B) = F_{\alpha^{(1)}} \circ \dots \circ F_{\alpha^{(p)}}$$

$$\text{où } \alpha_k^{(l)} = \begin{cases} 1 & k = 0 \\ -u_l & k = 1 \\ 0 & \text{sinon} \end{cases}$$

Le but désormais est de chercher pour tout $l \in \llbracket 1, p \rrbracket$ un $\beta^{(l)}$ tel que $\alpha^{(l)} \star \beta^{(l)} = \delta$ pour pouvoir inverser la relation. On sait d'après la remarque qu'il suffit de trouver $\beta^{(l)}$ tel que $\frac{1}{1 - u_l z} = \sum_{k \in \mathbb{Z}} \beta_k^{(l)} z^k$ pour tout $|z| = 1$

On sait de plus que $\forall l \in \llbracket 1, p \rrbracket |u_l| \neq 1$ par hypothèse sur les racines de Θ .

Prenons $l \in \llbracket 1, p \rrbracket$ alors deux cas sont possibles :

- si $|u_l| < 1$ on a $\forall |z| = 1 \frac{1}{1 - u_l z} = \sum_{k \geq 0} u_l^k z^k$ il suffit donc de prendre $\beta_k^{(l)} = \begin{cases} u_l^k & k \geq 0 \\ 0 & \text{sinon} \end{cases}$
- si $|u_l| > 1$ on a $\forall |z| = 1 \frac{1}{1 - u_l z} \frac{-u_l^{-1} z^{-1}}{1 - u_l^{-1} z^{-1}} = \sum_{k \leq -1} -u_l^k z^k$ il suffit donc de prendre $\beta_k^{(l)} = \begin{cases} -u_l^k & k \leq -1 \\ 0 & \text{sinon} \end{cases}$

Finalement prenons $F = F_{\beta(p)} \circ \dots \circ F_{\beta(1)}$, on prends alors

$$X = F(\epsilon)$$

□

1.2 Prédiction

Proposition 1.2.1. *On se donne X stationnaire au second ordre vérifiant (AR). On suppose de plus que X est causal i.e $\Theta(z) \neq 0$ pour tout $|z| \leq 1$ On note $\mathcal{H}_t^X = \text{Vect}\{X_s, s \leq t\}$. Alors*

$$\hat{X}_{t+1} = \text{proj}(X_{t+1} | \mathcal{H}_t^X) = \sum_{k=1}^p \theta_k X_{t+1-k}$$

Démonstration. $X_{t+1} = \sum_{k=1}^p \theta_k X_{t+1-k} + \epsilon_{t+1}$ de plus $\forall k \in \llbracket 1, p \rrbracket X_{t+1-k} \in \mathcal{H}_t^X$ donc

$$\hat{X}_{t+1} = \sum_{k=1}^p \theta_k X_{t+1-k} + \text{proj}(\epsilon_{t+1} | \mathcal{H}_t^X)$$

Or X est causal donc dans la construction de X (cf preuve d'avant) on a $\Theta(z) = \prod_{l=1}^p (1 - u_l z)$ où $\forall l |u_l| < 1$ ainsi les $\beta^{(l)}$ correspondants sont tous à support dans \mathbb{N} ce qui entraîne que $\psi = \beta^{(p)} \star \dots \star \beta^{(1)}$ est aussi à support dans \mathbb{N} et donc

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

On en déduit que $\mathcal{H}_t^X = \mathcal{H}_t^\epsilon$ et donc comme ϵ est un bruit blanc $\text{proj}(\epsilon_{t+1} | \mathcal{H}_t^X) = \text{proj}(\epsilon_{t+1} | \mathcal{H}_t^\epsilon) = 0$ d'où la solution □

1.3 Algorithme de Levinson-Durbin

On rappelle ici l'algorithme de Levinson Durbin et quelques conséquences. Le cadre d'étude est le suivant :

On se donne un processus stationnaire au second ordre, centré X de fonction d'autocovariance γ . On suppose que pour tout $p \geq 1$ la matrice de covariance de $\mathbf{X}_{t,p} = [X_t, \dots, X_{t-p+1}]^T$ $\Gamma_p = \mathbb{E}[\mathbf{X}_{t,p} \mathbf{X}_{t,p}^T]$ est **inversible**. On rappelle que cette matrice est la matrice de Toeplitz suivante :

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

Remarquons le lemme suivant :

Lemme 1.3.1. *Il existe $t \in \mathbb{Z}$ tel que $X_t \in \text{Vect}(X_{t-1}, \dots, X_{t-p+1})$ si et seulement si Γ_p est non inversible*

Démonstration. En effet Γ_p est la matrice de Gram de (X_t, \dots, X_{t-p+1}) associée au produit scalaire $\langle X, Y \rangle = \mathbb{E}[X \bar{Y}]$ qui est inversible si et seulement si (X_t, \dots, X_{t-p+1}) est libre. □

On note pour tout $p \geq 1$, pour tout $k \in \llbracket 1, p \rrbracket$, $\theta_{k,p}$ les coefficients de prédiction direct d'ordre p de X_t à partir de son passé et σ_p^2 l'erreur de prédiction i.e

$$\text{proj}(X_t | \mathcal{H}_{t-1,p}^X) = \sum_{k=1}^p \theta_{k,p} X_{t-k} \text{ et } \sigma_p^2 = \|X_t - \text{proj}(X_t | \mathcal{H}_{t-1,p}^X)\|_2^2$$

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

L'algorithme de Levinson permet de trouver tous les $(\theta_{k,p})_{p \in \llbracket 1, d \rrbracket, k \in \llbracket 1, p \rrbracket}$ où d est un ordre auquel on s'arrête, à partir des

$\gamma(1), \dots, \gamma(d)$. Il est basé sur un calcul de $(\theta_{1,p+1}, \dots, \theta_{p,p})$ connaissant $(\theta_{1,p-1}, \dots, \theta_{p-1,p-1})$

Algorithme 1.3.1 : Levinson-Durbin

Data : $\gamma(1), \dots, \gamma(d)$

Result : $(\theta_{i,p})_{p \in \llbracket 1, d \rrbracket, i \in \llbracket 1, p \rrbracket}$

Initialization :

└ $\kappa_1 = \frac{\gamma(1)}{\gamma(0)}, \theta_{1,1} = \kappa_1, \sigma_1^2 = \gamma(0)(1 - \kappa_1^2);$
for $p = 2, \dots, d$ **do**
└ $\kappa_p = \sigma_{p-1}^{-2} \left(\gamma(p) - \sum_{k=1}^{p-1} \theta_{k,p-1} \gamma(p-k) \right);$
└ $\theta_{p,p} = \kappa_p;$
└ $\sigma_p^2 = \sigma_{p-1}^2 (1 - \kappa_p^2);$
└ **for** $m = 1 \dots p-1$ **do**
└└ $\theta_{m,p} = \theta_{m,p-1} - \kappa_p \theta_{p-m,p-1}$

Proposition 1.3.1. Les coefficient κ_p s'appellent coefficients d'autocorrélation partielle et vérifient :

$$\forall p \geq 1, |\kappa_p| \leq 1$$

Proposition 1.3.2. Soit $p \geq 1$. Si pour tout $k \geq p, \kappa_k \in]-1, 1[$ alors $\forall k \leq p, \Gamma_p$ est inversible.

On peut alors en générant une famille $(\kappa_1, \dots, \kappa_d) \in]-1, 1[^d$ calculer des coefficients θ en inversant les équations de Levinson-Durbin grâce à l'algorithme suivant

Algorithme 1.3.2 : Construction des θ à partir des κ

Data : $(\kappa_1, \dots, \kappa_d) \in]-1, 1[^d$

Result : $(\theta_{i,p})_{p \in \llbracket 1, d \rrbracket, i \in \llbracket 1, p \rrbracket}$

Initialization :

└ **for** $p = 1, \dots, d$ **do**
└└ $\theta_{p,p} = \kappa_p;$
for $p = 2, \dots, d$ **do**
└ **for** $m = 1, \dots, p-1$ **do**
└└ $\theta_{m,p} = \theta_{m,p-1} - \kappa_p \theta_{p-m,p-1};$

Remarque. Dans cet algorithme, à chaque étape on calcule les $(\theta_{1,p}, \dots, \theta_{p-1,p})$ grâce aux $(\theta_{1,p-1}, \dots, \theta_{p-1,p-1})$ calculés à l'étape d'avant :

- $p = 2 : \theta_{1,2} = \theta_{1,1} - \kappa_2 \theta_{1,1}$
- $p = 3 : \theta_{1,3} = \theta_{1,2} - \kappa_3 \theta_{2,2}$ et $\theta_{2,3} = \theta_{2,2} - \kappa_3 \theta_{1,2}$
- ...

Théorème 1.3.1. Soit $d \geq 1$. Alors quels que soient $(\kappa_1, \dots, \kappa_d) \in]-1, 1[^d, \forall p \in \llbracket 1, d \rrbracket$, les coefficients $(\theta_{1,p}, \dots, \theta_{p,p})$ construits par l'algorithme ?? sont les coefficients d'un processus AR(p) causal i.e

$$1 - \sum_{k=1}^p \theta_{k,p} z^k \neq 0 \quad \forall |z| \leq 1$$

Démonstration.

Voyons tout d'abord le cas où $p = 1$:

On a $\theta_{1,1} = \kappa_1 = \frac{\gamma(1)}{\gamma(0)} \in]-1, 1[\setminus \{0\}$. Donc l'unique racine de $1 - \theta_{1,1} z$ est $\frac{1}{\theta_{1,1}}$ de module strictement supérieur à 1

Cas p quelconque : Soit maintenant $p \in \llbracket 1, d \rrbracket$ quelconque alors écrivons

$$1 - \sum_{k=1}^p \theta_{k,p} z^k = \prod_{k=1}^p (1 - \nu_{k,p} z)$$

où les $\nu_{k,p}$ sont les inverses des racines du polynôme. Le résultat cherché équivaut à $\forall k \in 1, \dots, p, |\nu_{k,p}| < 1$.

Construisons une bijection liant les $(\nu_{k,p})_{k \in \llbracket 1, p \rrbracket}$ aux $(\theta_{k,p})_{k \in \llbracket 1, p \rrbracket}$:

La première idée serait de considérer $\theta_p = [\theta_{1,p}, \dots, \theta_{p,p}]^T \mapsto [\nu_{1,p}, \dots, \nu_{p,p}]^T$ mais cela ne définit pas une application car une permutation des éléments dans le vecteur de droite donne le même polynôme et donc on aurait deux images pour le même θ_p .

La deuxième idée serait de considérer alors $\theta \mapsto \{\nu_{k,p}, k \in \llbracket 1, p \rrbracket\}$ mais le problème dans cette définition est que si il existe $i \neq j$ tels que $\nu_{i,p} = \nu_{j,p}$, cela a une importance dans le polynôme (racine multiple) mais cette redondance n'est pas visible dans $\{\nu_{k,p}, k \in \llbracket 1, p \rrbracket\}$.

Finalement une vision qui fonctionne est de considérer $\Phi : \theta_p = (\theta_{1,p}, \dots, \theta_{p,p}) \mapsto \bar{\nu}_p = \sum_{k=1}^p \delta_{\nu_{k,p}}$. Le fait de considérer la mesure $\sum_{k=1}^p \delta_{\nu_{k,p}}$ tient compte du fait qu'une interversion des $\nu_{k,p}$ ne change pas la valeur du polynôme et n'annule pas les redondances.

Soit X le processus stationnaire au second ordre associé aux $\kappa_1, \dots, \kappa_p$. Montrons que $\bar{\nu}_p \mapsto \mathbb{E} \left[\left| \left(\prod_{k=1}^p (1 - \nu_{k,p} B) \right) X \right|_t^2 \right]$

admet un unique minimiseur :

Il suffit de considérer $\theta_p = \Phi^{-1}(\bar{\nu}_p)$ et alors

$$\mathbb{E} \left[\left| \left(\prod_{k=1}^p (1 - \nu_{k,p} B) \right) X \right|_t^2 \right] = \mathbb{E} \left[\left| X_t - \sum_{k=1}^p \theta_{k,p} X_{t-k} \right|^2 \right]$$

Ce problème a bien une unique solution θ_p^* car Γ_p est inversible (puisque pour tout $k \in \llbracket 1, p \rrbracket, \kappa_k \in]-1, 1[$). Ainsi on a bien par bijection une unique solution

$$\bar{\nu}_p^* = \Phi(\theta_p^*) = \arg \min_{\bar{\nu}_p} \mathbb{E} \left[\left| \left(\prod_{k=1}^p (1 - \nu_{k,p} B) \right) X \right|_t^2 \right]$$

Ecrivons maintenant

$$\bar{\nu}_p^* = \sum_{k=2}^p \delta_{\nu_{k,p}^*} + \delta_{\nu_{1,p}^*} \quad (1.1)$$

Alors

$$\nu_{1,p}^* = \arg \min_{\nu \in \mathbb{C}} \mathbb{E} \left[\left| \left((1 - \nu B) \circ \prod_{k=2}^p (1 - \nu_{k,p}^* B) \right) X \right|_t^2 \right]$$

En effet si cela n'était pas le cas i.e si $\tilde{\nu} = \arg \min_{\nu \in \mathbb{C}} \mathbb{E} \left[\left| \left((1 - \nu B) \circ \prod_{k=2}^p (1 - \nu_{k,p}^* B) \right) X \right|_t^2 \right]$ était différent de $\nu_{1,p}^*$ on aurait

$$\mathbb{E} \left[\left| \left((1 - \tilde{\nu} B) \circ \prod_{k=2}^p (1 - \nu_{k,p}^* B) \right) X \right|_t^2 \right] < \mathbb{E} \left[\left| \left((1 - \nu_{1,p}^* B) \circ \prod_{k=2}^p (1 - \nu_{k,p}^* B) \right) X \right|_t^2 \right]$$

donc $\sum_{k=2}^p \delta_{\nu_{k,p}^*} + \delta_{\tilde{\nu}}$ fournirait une valeur plus petite que celle fournie par $\bar{\nu}_p^*$. Ceci est impossible par définition de ce dernier.

Considérons maintenant $Y = \prod_{k=2}^p (1 - \nu_{k,p}^* B) X$. Par filtrage d'un processus stationnaire au second ordre centré, Y est stationnaire au second ordre centré. De plus on a

$$\begin{aligned} \nu_{1,p}^* &= \arg \min_{\nu \in \mathbb{C}} \mathbb{E} \left[|(1 - \nu B) Y|_t^2 \right] \\ &= \arg \min_{\nu \in \mathbb{C}} \mathbb{E} \left[|Y_t - \nu Y_{t-1}|^2 \right] \end{aligned}$$

On est ramené à un cas d'ordre 1 dont la solution est le coefficient de prédiction d'ordre 1 associé à Y qui vaut (cf algorithme de Levinson ??) $\frac{\gamma_Y(1)}{\gamma_Y(0)}$. Donc

$$|\nu_{1,p}^*| = \left| \frac{\gamma_Y(1)}{\gamma_Y(0)} \right| \leq 1$$

Il reste à montrer que cette inégalité est stricte.

$|\nu_{1,p}^*| = 1 \iff \gamma_Y(1) = \gamma_Y(0) \iff \mathbb{E} [Y_1 \bar{Y}_0] = \mathbb{E} [Y_0^2] \iff Y_1 \in \text{Vect}(Y_0)$ (cas d'égalité dans Cauchy-Schwarz).

On a $Y_0 \in \text{Vect}(X_0, \dots, X_{1-p})$ donc

$$\text{Vect}(Y_0) \subset \text{Vect}(X_0, \dots, X_{1-p})$$

De plus en développant le produit dans l'expression de Y , on peut écrire $Y_1 = X_1 + Z$ où $Z \in \text{Vect}(X_0, \dots, X_{2-p}) \subset \text{Vect}(X_0, \dots, X_{1-p})$. Ainsi

$$Y_1 \in \text{Vect}(Y_0) \Rightarrow X_1 = Y_1 - Z \in \text{Vect}(X_0, \dots, X_{1-p}) \Rightarrow \Gamma_p \text{ non inversible}$$

Ce qui est contradictoire.

On a donc montré que $|\nu_{1,p}^*| < 1$ mais de plus dans l'équation (??), l'ordre des coefficients $\nu_{k,p}^*$ n'a pas d'importance, ainsi avoir montré l'inégalité pour $\nu_{1,p}^*$ la prouve aussi pour les autres (il suffit d'alterner $\nu_{1,p}^*$ avec un autre coefficient).

Finalement on a bien $\forall k \in \llbracket 1, p \rrbracket, |\nu_{k,p}^*| < 1$

□

1.4 Densité spectrale de puissance

On rappelle la formule de la DSP dans le cas d'un modèle AR(p) :

Définition 1.4.1 (DSP). La DSP est définie pour $\lambda \in [-\pi, \pi[$

$$S_x(\lambda) = \frac{\sigma^2}{2\pi |\Theta(e^{-i\lambda})|^2}$$

où $\Theta(z) = 1 - \sum_{k=1}^p \theta_k z^k$

Proposition 1.4.1. *Considérons z_1, \dots, z_p les inverses des racines de $\Theta(z)$ alors en notant $\forall n \in \llbracket 1, p \rrbracket, z_n = \rho_n e^{i\phi_n}$ avec $0 < \rho_n < 1$ et $\phi_n \in [-\pi, \pi[$ on a, si $\forall n \in \llbracket 1, p \rrbracket, \rho_n$ est assez proche de 1*

$$S_x(\lambda) \text{ admet un pic en } \lambda \iff \exists n \in \llbracket 1, p \rrbracket \lambda = \phi_n$$

Démonstration.

$$|\Theta(e^{-i\lambda})|^2 = \prod_{n=1}^p |1 - \rho_n e^{i(\phi_n - \lambda)}|^2$$

Or $\forall n \in \llbracket 1, p \rrbracket$

$$|1 - \rho_n e^{i(\phi_n - \lambda)}|^2 = 1 - 2\rho_n \cos(\phi_n - \lambda) + \rho_n^2$$

est minimal pour $\lambda = \phi_n \pmod{2\pi}$

Ceci ne veut pas dire que le produit est minimal ssi $\exists n \in \llbracket 1, p \rrbracket, \lambda = \phi_n(u) \pmod{2\pi}$. Mais si $\forall n \in \llbracket 1, p \rrbracket, \rho_n$ est assez proche de 1, pour chaque racine, les contributions des autres racines sont négligeables. □

Cas particulier où $p=1$: Dans le cas $p = 1$ quelle que soit la valeur de ρ , $|1 - \rho e^{i(\phi - \lambda)}|^2$ est minimal ssi $\lambda = \rho$

Cas particulier où $p=2$ à racines complexes conjuguées : On écrit les deux inverses des racines : $z_1 = \bar{z}_2 = \rho e^{i\phi}$, avec $0 < \rho < 1, \phi \in [-\pi, \pi[$ alors en développant le produit, $\forall \lambda \in [-\pi, \pi[$

$$f(\lambda) \stackrel{\text{def}}{=} |\Theta(e^{-i\lambda})|^2 = 1 + \rho^4 + 4\rho^2 \cos^2 \phi - 4\rho(1 + \rho^2) \cos \phi \cos \lambda + 2\rho^2 \cos 2\lambda$$

On dérive :

$$f'(\lambda) = 4\rho \sin(\lambda) [(1 + \rho^2) \cos \phi - 2\rho \cos \lambda]$$

ainsi

$$f'(\lambda) = 0 \iff \lambda = 0 \text{ ou } \cos \lambda = \frac{1 + \rho^2}{2\rho} \cos \phi$$

La seconde condition n'est possible que si $-1 \leq \frac{1 + \rho^2}{2\rho} \cos \phi \leq 1$ ce qui se traduit par

$$\begin{cases} \phi \in \left[\arccos\left(\frac{2\rho}{1+\rho^2}\right), \arccos\left(\frac{-2\rho}{1+\rho^2}\right) \right] & (\text{cas où } \phi \in [0, \pi]) \\ \text{ou} \\ \phi \in \left[-\arccos\left(\frac{-2\rho}{1+\rho^2}\right), -\arccos\left(\frac{2\rho}{1+\rho^2}\right) \right] & (\text{cas où } \phi \in [-\pi, 0]) \end{cases} \quad (1.2)$$

On a alors deux cas possibles :

1. Si ϕ vérifie (??) : alors

$$f'(\lambda) = 0 \iff \lambda = 0 \text{ ou } \lambda = \pm \arccos\left(\frac{1+\rho^2}{2\rho} \cos \phi\right)$$

et on a le tableau de variations suivant :

λ	$-\pi$	$-\arccos\left(\frac{1+\rho^2}{2\rho} \cos \phi\right)$	0	$\arccos\left(\frac{1+\rho^2}{2\rho} \cos \phi\right)$	π
$f'(\lambda)$	$-$	0	$+$	0	$+$
f		\searrow	\nearrow	\searrow	\nearrow

2. Si ϕ ne vérifie pas (??) alors là encore deux cas possibles

(a) Si $\frac{1+\rho^2}{2\rho} \cos \phi > 1$ i.e $\phi < \arccos\left(\frac{2\rho}{1+\rho^2}\right)$ ou $\phi > -\arccos\left(\frac{2\rho}{1+\rho^2}\right)$ alors $(1+\rho^2) \cos \phi - 2\rho \cos \lambda > 0$ et on a le tableau suivant :

λ	$-\pi$	0	π
$f'(\lambda)$	$-$	0	$+$
f		\searrow	\nearrow

(b) Si $\frac{1+\rho^2}{2\rho} \cos \phi < -1$ i.e $\phi < -\arccos\left(\frac{-2\rho}{1+\rho^2}\right)$ ou $\phi > \arccos\left(\frac{-2\rho}{1+\rho^2}\right)$ alors $(1+\rho^2) \cos \phi - 2\rho \cos \lambda < 0$ et on a le tableau suivant :

λ	$-\pi$	0	π
$f'(\lambda)$	$+$	0	$-$
f		\nearrow	\searrow

Voyons ce qu'il se passe quand $\rho \rightarrow 1$:

$\frac{1+\rho^2}{2\rho} \cos \phi \xrightarrow{\rho \rightarrow 1} \cos \phi$ donc on se retrouve dans le cas 1 et $\arccos\left(\frac{1+\rho^2}{2\rho} \cos \phi\right) \xrightarrow{\rho \rightarrow 1} \phi$. On aura donc bien nos pics à $\lambda = \phi$ et $\lambda = -\phi$

Chapitre 2

Processus TVAR

2.1 Construction d'une solution stable

Equation TVAR :

$$X_t = \sum_{i=1}^p \theta_i(t) X_{t-i} + \sigma(t) \epsilon_t \quad (\text{TVAR})$$

où $\epsilon \stackrel{\text{iid}}{\sim} BB(0, 1)$.

Condition de stabilité : Le critère qui nous intéresse est d'avoir une solution de (TVAR) vérifiant la condition de stabilité suivante :

$$\sup_{t \in \mathbb{Z}} \mathbb{E} [|X_t|^2] < +\infty \quad (\text{S})$$

2.1.1 Cas simple

On considère ici $\sigma(t) = 1 \forall t \in \mathbb{R}$ et $\theta_i(t) = \begin{cases} 0 & t < 0 \\ \theta_i & t \geq 0 \end{cases}$.

Proposition 2.1.1. Dans ce cas particulier, l'équation (TVAR) admet une unique solution

Démonstration. Tout d'abord pour $t < 0$ on a $X_t = \epsilon_t$

Pour $t \geq 0$, notons $\mathbf{X}_k = [X_k, \dots, X_{k-p+1}]^T$, $\mathbf{e}_1 = [1, 0, \dots, 0]^T$ et $A = \begin{pmatrix} \theta_1 & \theta_2 & \dots & \dots & \theta_p \\ 1 & 0 & \dots & \dots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}$. Alors l'équation

(TVAR) s'écrit :

$$\mathbf{X}_t = A\mathbf{X}_{t-1} + \epsilon_t \mathbf{e}_1$$

En itérant pour $t-1, t-2$ etc, on obtient

$$\forall k \geq 0 \mathbf{X}_t = A^{k+1} \mathbf{X}_{t-k-1} + \sum_{j=0}^k \epsilon_{t-j} A^j \mathbf{e}_1$$

Considérons $k \geq t$ alors $\mathbf{X}_{t-k-1} = [\epsilon_{t-k-1}, \dots, \epsilon_{t-k-p}]^T$. Ainsi

$$\forall k \geq t \mathbf{X}_t = A^{k+1} \begin{pmatrix} \epsilon_{t-k-1} \\ \vdots \\ \epsilon_{t-k-p} \end{pmatrix} + \sum_{j=0}^k \epsilon_{t-j} A^j \mathbf{e}_1 \quad (2.1)$$

ce qui fournit une définition $X_t = \mathbf{e}_1^T \mathbf{X}_t$ unique □

On cherche alors une condition sur les $(\theta_i)_{i=1}^p$ pour cette solution vérifie la condition de stabilité (S)

Cas où $p=1$

Proposition 2.1.2. Si $p = 1$ on note $\theta(t) = \begin{cases} 0 & t < 0 \\ \theta & t \geq 0 \end{cases}$ et (TVAR) devient $X_t = \theta(t)X_{t-1} + \sigma(t)\epsilon_t$. La solution de l'équation vérifie la condition de stabilité (S) si et seulement si $|\theta| < 1$

Démonstration. Si $t < 0$ $X_t = \epsilon_t$ donc $\sup_{t < 0} \mathbb{E}[|X_t|^2] = 1$. Si $t \geq 0$, la formule de X_t donnée par (1) se traduit pour $p = 1$ par $\forall k \geq t$ $X_t = \theta^{k+1}\epsilon_{t-k-1} + \sum_{j=0}^k \theta^j \epsilon_{t-j}$ i.e $\forall k \geq t$ $X_t = \sum_{j=0}^{k+1} \theta^j \epsilon_{t-j}$. Ainsi

$$\forall t \in \mathbb{N}, \forall k \geq t \mathbb{E}[|X_t|^2] = \sum_{i=0}^{k+1} \sum_{j=0}^{k+1} \theta^i \bar{\theta}^j \mathbb{E}[\epsilon_{t-i} \epsilon_{t-j}] = \sum_{i=0}^{k+1} |\theta|^{2i}$$

Ainsi en faisant tendre k vers $+\infty$ on obtient

$$\forall t \in \mathbb{N}, \mathbb{E}[|X_t|^2] = \sum_{i=0}^{+\infty} |\theta|^{2i} = \begin{cases} +\infty & |\theta| \geq 1 \\ \frac{1}{1-|\theta|^2} & |\theta| < 1 \end{cases}$$

Ce qui donne $\sup_{t \in \mathbb{N}} \mathbb{E}[|X_t|^2] < +\infty \iff |\theta| < 1$ et comme sur $\sup_{t < 0} \mathbb{E}[|X_t|^2] = 1$ la condition est valable pour le sup sur $t \in \mathbb{Z}$ \square

Cas p quelconque

Pour prouver ce cas, quelques lemmes d'algèbre linéaire sont nécessaires

Lemme 2.1.1. La matrice A définie dans la preuve de la propriété 3.1 a pour polynôme caractéristique :

$$\chi_A(X) = \det(A - XI_p) = (-1)^p \left(X^p - \sum_{i=1}^p \theta_i X^{p-i} \right)$$

Conséquence : Les valeurs propres de A sont les inverses des racines de $P = 1 - \sum_{i=1}^p \theta_i X^i$ car $\chi_A = (-1)^p X^p P(\frac{1}{X})$

Lemme 2.1.2. Soit $A \in \mathbb{C}^{p \times p}$ alors, on rappelle la définition de la norme subordonnée associée à une norme $\|\cdot\|$ sur \mathbb{C}^p :

$$\|A\| = \sup_{\|x\|=1} \|Ax\|$$

On rappelle aussi la définition du rayon spectral $\rho(A) = \max_{\lambda \in \text{spec}(A)} |\lambda|$. On a la propriété suivante :

Pour tout $A \in \mathbb{C}^{p \times p}$, pour tout $\epsilon > 0$ il existe une norme sur \mathbb{C}^p dépendant de ϵ et de A telle que la norme subordonnée correspondante $\|\cdot\|_{\epsilon, A}$ vérifie

$$\|A\|_{\epsilon, A} \leq \rho(A) + \epsilon$$

Proposition 2.1.3 (Condition suffisante). Dans ce cas particulier, en notant $P(z) = 1 - \sum_{i=1}^p \theta_i z^i$, si $\forall |z| \leq 1$ $P(z) \neq 0$ (i.e les racines de P sont hors du disque unité fermé) alors la solution de l'équation (TVAR) vérifie la condition de stabilité (S)

Démonstration. On part de la définition de \mathbf{X}_t pour $t \in \mathbb{N}$ donnée par (1)

De plus $X_t = \mathbf{e}_1^T \mathbf{X}_t$ donc $\mathbb{E}[|X_t|^2] = \mathbb{E}[X_t \bar{X}_t^T] = \mathbf{e}_1^T \mathbb{E}[\mathbf{X}_t \mathbf{X}_t^T] \mathbf{e}_1$ avec $\forall k \geq t$, comme ϵ est un bruit blanc

$$\begin{aligned} \mathbb{E}[\mathbf{X}_t \mathbf{X}_t^T] &= \sum_{j=0}^k \sum_{l=0}^k A^j \mathbf{e}_1 \mathbb{E}[\epsilon_{t-j} \epsilon_{t-l}] \mathbf{e}_1^T (A^l)^T + A^{k+1} \mathbb{E} \left[\begin{pmatrix} \epsilon_{t-k-1} \\ \vdots \\ \epsilon_{t-k-p} \end{pmatrix} [\epsilon_{t-k-1}, \dots, \epsilon_{t-k-p}] \right] (A^{k+1})^T \\ &= \sum_{j=0}^k A^j \mathbf{e}_1 \mathbf{e}_1^T (A^j)^T + A^{k+1} (A^{k+1})^T \end{aligned}$$

Or on a supposé que les racines de P sont de module strictement supérieur à 1 donc les valeurs propres de A sont de module strictement inférieur à 1. Ainsi $\rho(A) < 1$, il existe donc $\epsilon > 0$ tel que $\rho(A) + \epsilon < 1$. En appliquant le lemme 3 à ϵ et A on obtient (en notant juste $\|A\|$ au lieu de $\|A\|_{A, \epsilon}$ pour alléger les notations) $\|A\| < 1$.

Ceci implique dans un premier temps $\|A^k\| \leq \|A\|^k \xrightarrow[k \rightarrow +\infty]{} 0$ donc $A^k \xrightarrow[k \rightarrow +\infty]{} 0$ et ainsi $A^{k+1}(A^{k+1})^T \xrightarrow[k \rightarrow +\infty]{} 0$

De plus $\forall j \in \mathbb{N} \quad \left\| A^j \mathbf{e}_1 \mathbf{e}_1^T (A^j)^T \right\| \leq \|A\|^{2j} \|\mathbf{e}_1 \mathbf{e}_1^T\|$ qui est terme général d'une série convergente dans \mathbb{R} car $\|A\| < 1$ donc la série des $A^j \mathbf{e}_1 \mathbf{e}_1^T (A^j)^T$ est absolument convergente donc convergente dans $\mathbb{R}^{p \times p}$.

En faisant donc tendre k vers $+\infty$ dans l'expression de $\mathbb{E} [\mathbf{X}_t \mathbf{X}_t^T]$ on obtient :

$$\mathbb{E} [\mathbf{X}_t \mathbf{X}_t^T] = \sum_{j=0}^{+\infty} A^j \mathbf{e}_1 \mathbf{e}_1^T (A^j)^T \in \mathbb{R}^{p \times p}$$

Ainsi $\mathbb{E} [|X_t|^2]$ étant le premier coefficient de cette matrice, $\mathbb{E} [|X_t|^2] < +\infty$ □

2.1.2 Cas général avec $p=1$

L'équation (TVAR) devient $X_t = \theta(t)X_{t-1} + \sigma(t)\epsilon_t$

Proposition 2.1.4. *Si $\sup_t |\theta(t)| < 1$ et $\sup_t |\sigma(t)| < +\infty$ alors il existe un unique processus $(X_t)_{t \in \mathbb{Z}}$ vérifiant à la fois (TVAR) et la condition de stabilité (S)*

Démonstration. En itérant k fois l'équation on obtient

$$\forall k \geq 0 \quad X_t = \left(\prod_{j=0}^k \theta(t-j) \right) X_{t-k-1} + \sum_{j=0}^k \left(\prod_{i=0}^{j-1} \theta(t-i) \right) \sigma(t-j) \epsilon_{t-j} \quad (2.2)$$

- Supposons que X vérifie la condition de stabilité et appelons $M = \sup_t \mathbb{E} [|X_t|^2]$, $\theta_{\max} = \sup_t |\theta(t)|$ et $\sigma_{\max} = \sup_t |\sigma(t)|$ alors on a $\forall k \geq 0$:

$$\left\| \left(\prod_{j=0}^k \theta(t-j) \right) X_{t-k-1} \right\|_2^2 = \left(\prod_{j=0}^k |\theta(t-j)|^2 \right) \mathbb{E} [|X_{t-k-1}|^2] \leq \theta_{\max}^{2(k+1)} M$$

Ainsi $\lim_{k \rightarrow +\infty} \left\| \left(\prod_{j=0}^k \theta(t-j) \right) X_{t-k-1} \right\|_2^2 = 0$ car on a pris $\theta_{\max} < 1$.

De plus $\forall j \geq 0$

$$\left\| \left(\prod_{i=0}^{j-1} \theta(t-i) \right) \sigma(t-j) \epsilon_{t-j} \right\|_2^2 = \left(\prod_{i=0}^{j-1} |\theta(t-i)|^2 \right) |\sigma(t-j)|^2 \mathbb{E} [\epsilon_{t-j}^2] = \left(\prod_{i=0}^{j-1} |\theta(t-i)|^2 \right) |\sigma(t-j)|^2 \leq \theta_{\max}^{2j} \sigma_{\max}^2 \quad (2.3)$$

Ce qui donne

$$\left\| \left(\prod_{i=0}^{j-1} \theta(t-i) \right) \sigma(t-j) \epsilon_{t-j} \right\|_2 \leq \theta_{\max}^j \sigma_{\max}$$

Encore une fois, parce qu'on a pris $\theta_{\max} < 1$, on a à droite de l'inégalité le terme général d'une série convergente. Ainsi en notant $c_j = \left(\prod_{i=0}^{j-1} \theta(t-i) \right) \sigma(t-j)$ la série de terme général $c_j \epsilon_{t-j}$ est absolument convergente donc convergente.

On peut alors faire tendre k vers $+\infty$ dans (2), on obtient

$$X_t = \sum_{j=0}^{+\infty} c_j \epsilon_{t-j} \quad \text{avec} \quad c_j = \left(\prod_{i=0}^{j-1} \theta(t-i) \right) \sigma(t-j) \quad (2.4)$$

- Définissons maintenant X comme dans l'équation (4) et montrons qu'elle vérifie l'équation TVAR et la condition de stabilité.

Tout d'abord pour tout $t \in \mathbb{Z}$ on a

$$\begin{aligned}
\theta(t)X_{t-1} + \sigma(t)\epsilon_t &= \theta(t) \sum_{j=0}^{+\infty} \left(\prod_{i=0}^{j-1} \theta(t-1-i) \right) \sigma(t-1-j)\epsilon_{t-1-j} + \sigma(t)\epsilon_t \\
&= \sum_{j=-1}^{+\infty} \left(\prod_{i=-1}^{j-1} \theta(t-1-i) \right) \sigma(t-1-j)\epsilon_{t-1-j} \\
&\stackrel{(j \leftarrow j+1)}{=} \sum_{j=0}^{+\infty} \left(\prod_{i=-1}^{j-2} \theta(t-1-i) \right) \sigma(t-j)\epsilon_{t-j} \\
&\stackrel{(i \leftarrow i+1)}{=} \sum_{j=0}^{+\infty} \left(\prod_{i=0}^{j-1} \theta(t-i) \right) \sigma(t-j)\epsilon_{t-j} \\
&= X_t
\end{aligned}$$

- Remarquons tout d'abord que (3) prouve que $(c_j)_{j \in \mathbb{N}} \in \ell^2(\mathbb{N})$. On a alors $\forall t \in \mathbb{Z}$

$$\begin{aligned}
\mathbb{E}[|X_t|^2] &= \mathbb{E} \left[\left| \lim_{k \rightarrow +\infty} \sum_{j=0}^k c_j \epsilon_{t-j} \right|^2 \right] \\
&= \lim_{k \rightarrow +\infty} \mathbb{E} \left[\left| \sum_{j=0}^k c_j \epsilon_{t-j} \right|^2 \right] \quad (\text{par continuité de l'espérance}) \\
&= \lim_{k \rightarrow +\infty} \sum_{j=0}^k |c_j|^2 \mathbb{E}[|\epsilon_{t-j}|^2] \quad (\text{car } \epsilon \text{ est un bruit blanc}) \\
&= \lim_{k \rightarrow +\infty} \sum_{j=0}^k |c_j|^2 \\
&= \sum_{k=0}^{+\infty} |c_k|^2
\end{aligned}$$

Ce résultat étant indépendant de t on a bien

$$\sup_t \mathbb{E}[|X_t|^2] = \sum_{k=0}^{+\infty} |c_k|^2 < +\infty$$

□

2.1.3 Contre-exemple avec $p=2$

On va montrer que la condition obtenue pour le cas $p = 1$ n'est plus suffisante pour $p = 2$.

Pour ce contre-exemple, on va considérer le TVAR(2) défini comme ceci :

$$\begin{aligned}
X_{2t} &= aX_{2t-1} + \epsilon_{2t} \\
X_{2t+1} &= b_1X_{2t} + b_2X_{2t-1} + \epsilon_{2t+1}
\end{aligned}$$

pour $t > 0$ sinon $X_t = \epsilon_t$.

On a alors :

$$\forall t > 0, X_{2t+1} = (ab_1 + b_2)X_{2t-1} + b_1\epsilon_{2t} + \epsilon_{2t+1}$$

On a alors que le processus $Y_t = X_{2t+1}$ suit l'équation d'un AR(1) dont la condition de stabilité implique :

$$|ab_1 + b_2| < 1 \quad (*)$$

D'autre part, le polynôme caractéristique associé est $P(z) = 1 - b_1 z - b_2 z^2$. Posons $P(z) = (1 - bz)^2$. On a alors $b_1 = 2b$ et $b_2 = -b^2$. La condition (??) donne :

$$|2ba - b^2| < 1$$

Avec $a = -b$, celle-ci devient :

$$3b^2 < 1$$

Or ceci peut être faux. Il suffit de prendre $b = \frac{1}{\sqrt{2}}$ par exemple.

On a donc trouvé une sous-suite Y_t du processus X_t qui diverge. Donc la condition initiale, à savoir $\sup_t |\theta_i(t)| < 1$, n'est plus suffisante pour assurer la stabilité d'un TVAR(2).

2.1.4 Cas général

On introduit une nouvelle définition du modèle TVAR introduite dans [?] :

Définition 2.1.1. Soient $p \geq 1$, $\theta_1, \dots, \theta_p$ et σ des fonctions définies sur $] -\infty, 1]$ et $(\epsilon_t)_{t \in \mathbb{Z}}$ une suite i.i.d de variables aléatoires avec moyenne nulle et variance unitaire. Pour tout $T \geq 1$ on dit que $(X_{t,T})_{t \leq T}$ est un processus TVAR stable s'il vérifie les deux conditions suivantes

(i) $\forall -\infty < t \leq T$

$$X_{t,T} = \sum_{i=1}^p \theta_i \left(\frac{t}{T} \right) X_{t-i,T} + \sigma \left(\frac{t}{T} \right) \epsilon_t \quad (\text{TVAR'})$$

(ii)

$$\sup_{-\infty < t \leq T} \mathbb{E} \left[|X_{t,T}|^2 \right] < +\infty \quad (\text{S'})$$

Remarque. Revenons sur cette nouvelle définition du modèle TVAR. Supposons $u \in] -\infty, 1]$ fixé et $T \gg 1$. Si on veut étudier ce qu'il se passe pour t dans un intervalle du type $[(u-b)T, (u+b)T]$, alors on peut prendre b petit sans avoir un intervalle trop petit pour t . Alors $\frac{t}{T} \in [u-b, u+b]$ reste proche de u et si θ_i et σ sont suffisamment régulières, on peut approcher les approcher par des constantes. Ainsi l'équation s'approche par une équation AR.

Proposition 2.1.5. Supposons que les coefficients θ_i de l'équation (TVAR) sont uniformément continus sur $] -\infty, 1]$ et que σ est bornée sur $] -\infty, 1]$. Supposons de plus qu'il existe $\delta \in]0, 1[$ tel que $\Theta(z; u) \neq 0 \forall |z| < \delta^{-1}, u \in [0, 1]$ où

$$\Theta(z; u) = 1 - \sum_{i=1}^p \theta_i(u) z^i$$

Alors il existe $T_0 \geq 1$ tel que $\forall T \geq T_0$ il existe un unique processus $(X_{t,T})_{t \leq T}$ vérifiant (TVAR') et (S'). De plus cette solution s'écrit :

$$\forall -\infty < t \leq T \quad X_{t,T} = \sum_{j=0}^{+\infty} \phi_{t,T}(j) \sigma \left(\frac{t-j}{T} \right) \epsilon_{t-j} \quad (2.5)$$

avec $\forall \delta_1 \in]\delta, 1[$

$$\sup_{T \geq T_0} \sup_{-\infty < t \leq T} \sup_{j \geq 0} \delta_1^{-1} |\phi_{t,T}(j)| < +\infty$$

Démonstration. La preuve de cette proposition peut être trouvée dans [?] □

2.2 Prédiction sur un TVAR

Détail pratique pour l'implémentation : Dans cette partie, nous considérerons que les coefficients sont constantes pour $u \leq 0$: $\theta_i(u) = \theta_i(0)$ si $u \leq 0$. Dans la pratique nous prendrons 0. Nous allons donc nous restreindre à $[0, 1]$

2.2.1 Implémentation d'un TVAR

Coefficients AR abstraits associés Il va être intéressant pour la suite d'adopter la vision suivante :

On se donne p fonctions $\theta_1, \dots, \theta_p$ vérifiant les hypothèses de la proposition ?? . Alors à T donné et $-\infty < t \leq T$ fixé si on note $u = \frac{t}{T}$ les coefficients $(\theta_1(u), \dots, \theta_p(u))$ sont ceux d'un AR(p) causal car $\Theta(z; u) \neq 0 \forall |z| \leq 1$.

On peut donc à t fixé se ramener à ce que l'on connaît sur les processus AR en considérant un processus AR(p) abstrait associé à ces coefficients (notamment utiliser l'algorithme de Levinson-Durbin on considère la densité spectrale de puissance). Il y a un processus AR(p) abstrait local à chaque temps t et notamment une densité spectrale de puissance locale à chaque instant t .

Maintenant que nous considérons à chaque instant u un processus AR(p) local nous pouvons appliquer l'algorithme ?? et le théorème ?? pour chaque instant u à partir de p fonctions continues $\kappa_i : [0, 1] \rightarrow]-1, 1[$ pour $i = 1, \dots, p$

Théorème 2.2.1. *Quelles que soient les fonctions $\kappa_1, \dots, \kappa_p$ continues sur $[0, 1]$ et à valeurs dans $]-1, 1[$, les fonctions $\theta_{1,p}, \dots, \theta_{p,p}$ obtenues en appliquant l'algorithme ?? à chaque instant u sont les coefficients d'un processus TVAR(p) stable i.e elles sont uniformément continues sur $[0, 1]$ et il existe $\delta \in]0, 1[$ tel que $\Theta(z; u) \neq 0 \forall |z| < \delta^{-1}, u \in [0, 1]$ où*

$$\Theta(z; u) = 1 - \sum_{i=1}^p \theta_{i,p}(u) z^i$$

Démonstration. Tout d'abord les $(\theta_{i,p})_{i \in \llbracket 1, p \rrbracket}$ sont obtenus par sommes et produits des $(\kappa_i)_{i \in \llbracket 1, p \rrbracket}$ donc sont continus sur $[0, 1]$ donc uniformément continus sur ce segment.

Supposons maintenant $\forall \delta \in]0, 1[, \exists |z| < \delta^{-1}, u \in [0, 1], \Theta(z; u) = 0$. Considérons une suite $(\delta_n)_{n \in \mathbb{N}} \in]0, 1[^\mathbb{N}$ telle que $\delta_n \uparrow 1$. Alors il existe une suite $(z_n)_{n \in \mathbb{N}}$ et une suite $(u_n)_{n \in \mathbb{N}}$ telle que $\forall n \in \mathbb{N}$

$$|z_n| < \delta_n^{-1}, u_n \in [0, 1] \text{ et } \Theta(z_n; u_n) = 0$$

On peut alors prendre une sous suite $(z_{\alpha_n}, u_{\alpha_n})_{n \in \mathbb{N}}$ de $(z_n, u_n)_{n \in \mathbb{N}}$ qui converge vers un couple (z, u) tel que $|z| \leq 1$ et $u \in [0, 1]$. On a en effet $|z| \leq 1$ car $\forall n \in \mathbb{N}, |z_{\alpha_n}| < \delta_{\alpha_n}^{-1}$ et par passage à la limite $|z| \leq 1$

De plus par continuité de Θ , en passant à la limite dans $\forall n \in \mathbb{N}, \Theta(z_{\alpha_n}, u_{\alpha_n}) = 0$ on obtient finalement :

$$|z| \leq 1, u \in [0, 1] \text{ et } \Theta(z; u) = 0$$

Ce qui contredit le résultat du théorème ?? appliqué à $\kappa_1(u), \dots, \kappa_p(u)$ qui dit que $\Theta(z; u) \neq 0, \forall |z| \leq 1$ □

2.2.2 Définition du prédicteur TVAR

On considère un processus TVAR $X_{t,T}$ défini grâce à la proposition ?? . Notons que l'équation (TVAR') se réécrit de manière plus compacte par : $\forall t \in \llbracket 1, T \rrbracket$

$$X_{t,n} = \boldsymbol{\theta}_{t-1,T}^T \mathbf{X}_{t-1,T} + \sigma_{t,T} \epsilon_{t,T}$$

où

$$\begin{aligned} \mathbf{X}_{t,T} &= [X_{t,T}, X_{t-1,T}, \dots, X_{t-p+1,T}]^T \\ \boldsymbol{\theta}_{t,T} &= \boldsymbol{\theta} \left(\frac{t}{T} \right) = \left[\theta_1 \left(\frac{t}{T} \right), \theta_2 \left(\frac{t}{T} \right), \dots, \theta_p \left(\frac{t}{T} \right) \right]^T \\ \sigma_{t,T} &= \sigma \left(\frac{t}{T} \right) \end{aligned}$$

On cherche un estimateur de $\boldsymbol{\theta}_{t,T}$ défini par

$$\hat{\boldsymbol{\theta}}_{t,T} = \arg \min_{\boldsymbol{\theta} \in \mathbb{C}^p} \mathbb{E} \left[\left| X_{t+1,T} - \boldsymbol{\theta}^T \mathbf{X}_{t,T} \right|^2 \right]$$

On définit cet estimateur en ligne par l'algorithme des moindres carrés normalisé (NLMS) :

Définition 2.2.1 (Estimateur NLMS de $\boldsymbol{\theta}$). On se donne un pas μ , l'estimateur est le suivant :

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{0,T}(\mu) &= 0 \\ \forall t \in \llbracket 0, T-1 \rrbracket, \hat{\boldsymbol{\theta}}_{t+1,T}(\mu) &= \hat{\boldsymbol{\theta}}_{t,T}(\mu) + \mu (X_{t+1,T} - \hat{\boldsymbol{\theta}}_{t+1,T}(\mu)^T \mathbf{X}_{t,T}) \frac{\mathbf{X}_{t,T}}{1 + \mu \|\mathbf{X}_{t,T}\|^2} \end{aligned} \quad (\text{NLMS})$$

On définit ensuite l'estimateur de $\boldsymbol{\theta} : u \rightarrow \boldsymbol{\theta}(u) = [\theta_1(u), \theta_2(u), \dots, \theta_p(u)]^T$ pour $u \in [0, 1]$ en faisant une interpolation :

$$\hat{\boldsymbol{\theta}}_T(u; \mu) = \hat{\boldsymbol{\theta}}_{\lfloor uT \rfloor, T}(\mu)$$

A un instant u , l'estimateur est une fonction de $\mathbf{X}_{0,T}, X_{1,T}, \dots, X_{\lfloor uT \rfloor, T}$ et μ

Définition 2.2.2 (Prédicteur associé). On définit à partir de $(\boldsymbol{\theta}_{t,T}(\mu))_{1 \leq t \leq T}$ le prédicteur suivant : $\forall 1 \leq t \leq T$

$$\hat{X}_{t,T}(\mu) = \hat{\boldsymbol{\theta}}_{t-1,T}(\mu)^T \mathbf{X}_{t-1,T} \quad (2.6)$$

Définition 2.2.3 (β -Lipschitz semi-norme). Soit $\beta \in]0, 1]$ et $\mathbf{f} : [0, 1] \rightarrow \mathbb{R}^d$, on définit la semi-norme β -Lipschitz de f par :

$$|\mathbf{f}|_{\Lambda, \beta} = \sup_{t \neq s} \frac{\|\mathbf{f}(t) - \mathbf{f}(s)\|}{|t - s|^\beta}$$

On définit la boule β -Lipschitz de rayon $L > 0$:

$$\Lambda_d(\beta, L) = \left\{ \mathbf{f} : [0, 1] \rightarrow \mathbb{R}^d, |\mathbf{f}|_{\Lambda, \beta} \leq L, \sup_{t \in [0, 1]} \|\mathbf{f}(t)\| \leq L \right\}$$

Notations : On notera pour Z v.a à valeur dans un espace muni d'une norme $\|\cdot\|$: $\|Z\|_q = (\mathbb{E}[\|Z\|^q])^{1/q}$ et $\|Z\|_{q, \boldsymbol{\theta}, \sigma} = (\mathbb{E}_{\boldsymbol{\theta}, \sigma}[\|Z\|^q])^{1/q}$ dans le cas où l'on a une paramétrisation par $(\boldsymbol{\theta}, \sigma)$

On note de plus

- pour $\rho > 0$:

$$S(\rho) = \{ \boldsymbol{\theta} = [\theta_1, \dots, \theta_p]^T : [0, 1] \rightarrow \mathbb{R}^p, \Theta(z; u) \neq 0 \forall |z| < \delta^{-1}, u \in [0, 1] \}$$

- pour $\beta \in]0, 1], L > 0, \rho \in]0, 1[, 0 < \sigma_- \leq \sigma_+ < +\infty$:

$$C(\beta, L, \rho, \sigma_-, \sigma_+) = \{ (\boldsymbol{\theta}, \sigma), \boldsymbol{\theta} \in \Lambda_d(\beta, L) \cap S(\rho), \sigma : [0, 1] \rightarrow [\sigma_-, \sigma_+] \}$$

Théorème 2.2.2. On suppose que ϵ est un bruit blanc centré de variance unitaire indépendant de $\mathbf{X}_{0,T}$ et de plus que $\sup_{T \geq 1} \|\mathbf{X}_{0,T}\|_q < +\infty$ et $\sup_{1 \leq t \leq T} \|\epsilon_{t,T}\|_q < +\infty$ avec $q \geq 4$.

Soient $p \in [1, q/3], \beta \in]0, 1], L > 0, \rho \in]0, 1[$ et $0 < \sigma_- \leq \sigma_+$

Alors $\exists M, \delta > 0$ et $\mu_0 > 0$ tels que $\forall \mu \in]0, \mu_0], T \geq 1, u \in [0, 1]$ et $(\boldsymbol{\theta}, \sigma) \in C(\beta, L, \rho, \sigma_-, \sigma_+)$

$$\left\| \hat{\boldsymbol{\theta}}_T(u; \mu) - \boldsymbol{\theta}(u) \right\|_{p, \boldsymbol{\theta}, \sigma} \leq M (\|\boldsymbol{\theta}(0)\| (1 - \delta\mu)^{uT} + \sqrt{\mu} + (T\mu)^{-\beta})$$

Démonstration. Le preuve peut être trouvée dans [?] □

Remarque. Cette borne supérieure peut s'interpréter de la manière suivante :

- Le premier terme $\|\boldsymbol{\theta}(0)\| (1 - \delta\mu)^{uT}$ traduit l'oubli de l'erreur initiale de l'estimateur quand T augmente
- Le deuxième terme $\sqrt{\mu}$ est un terme de variance de l'estimateur qui serait là même si $u \mapsto \boldsymbol{\theta}(u)$ était constante
- Le dernier terme $(T\mu)^{-\beta}$ contrôle l'erreur due à l'évolution de $\boldsymbol{\theta}$ dans le temps. Elle décroît d'autant plus vite que β est grand, ce qui implique une bonne régularité de $\boldsymbol{\theta}$

2.2.3 Analyse en densité spectrale de puissance

Comme expliqué précédemment, on considère pour un processus TVAR donné, les processus AR abstraits à chaque instant $u \in [0, 1]$, on définit ainsi les DSP "locales" comme les DSP associées à chaque AR.

Définition 2.2.4 (DSP). La DSP locale théorique en $u \in [0, 1]$ est définie pour $\lambda \in [-\pi, \pi[$

$$S_x(\lambda; u) = \frac{\sigma^2(u)}{2\pi |\Theta(e^{-i\lambda}; u)|^2}$$

où $\Theta(z; u) = 1 - \sum_{k=1}^p \theta_k(u) z^k$

La DSP locale estimée est :

$$\hat{S}_x(\lambda; u) = \frac{\hat{\sigma}^2(u)}{2\pi \left| \hat{\Theta}(e^{-i\lambda}; u) \right|^2}$$

où $\hat{\Theta}(z; u) = 1 - \sum_{k=1}^p \hat{\theta}_k(u) z^k$

En appliquant la proposition ??, à chaque instant les modules des inverses racines sont assez proches de 1, les pics seront proches des phases des inverses des racines.

2.3 Agrégation des prédicteurs

2.3.1 Algorithme et stratégies

On applique dans cette section les méthodes d'agrégation en ligne fournies dans [?] au cadre des processus TVAR. On se donne des observations NLMS d'un processus sur une période de temps $\{1, \dots, T\}$ pour différentes valeurs de μ , le pas de descente de gradient : $(X_{t,T})_{1 \leq t \leq T}$ vérifiant (TVAR') et (S'). On considère N prédicteurs définis par (??) associés à N pas μ_1, \dots, μ_N : on notera $\forall i \in \llbracket 1, N \rrbracket, \forall 1 \leq t \leq T, \hat{X}_{t,T}^{(i)} = \hat{X}_{t,T}(\mu_i)$

Définition 2.3.1 (Agrégation). Agréger les prédicteurs $\left\{ \left(\hat{X}_{t,T}^{(i)} \right)_{1 \leq t \leq T}, i \in \llbracket 1, N \rrbracket \right\}$ signifie pour tout $1 \leq t \leq T$

- Choisir $\alpha_t = (\alpha_{1,t}, \dots, \alpha_{N,t}) \in \mathbb{R}_+^N$ tel que $\sum_{i=1}^N \alpha_{i,t} = 1$
- Calculer le nouvel estimateur $\hat{X}_{t,T} = \hat{X}_{t,T}^{[\alpha_t]} = \sum_{i=1}^N \alpha_{i,t} \hat{X}_{t,T}^{(i)}$

Comme expliqué dans [?], on peut approcher les poids d'agrégation optimaux α_t en calculant en ligne et récursivement des poids $\hat{\alpha}_t$ selon deux stratégies à partir d'un paramètre $\eta > 0$ nommé "pas d'apprentissage".

Stratégie 1 : à partir du gradient de l'erreur quadratique $\forall i \in \llbracket 1, N \rrbracket, \forall t \in \llbracket 1, T \rrbracket$

$$\hat{\alpha}_{i,t} = \frac{\exp \left(-2\eta \sum_{s=1}^{t-1} \left(\sum_{j=1}^N \hat{\alpha}_{j,s} \hat{X}_{s,T}^{(j)} - X_{s,T} \right) \hat{X}_{s,T}^{(i)} \right)}{\sum_{k=1}^N \exp \left(-2\eta \sum_{s=1}^{t-1} \left(\sum_{j=1}^N \hat{\alpha}_{j,s} \hat{X}_{s,T}^{(j)} - X_{s,T} \right) \hat{X}_{s,T}^{(k)} \right)}$$

Stratégie 2 : à partir de l'erreur quadratique $\forall i \in \llbracket 1, N \rrbracket, \forall t \in \llbracket 1, T \rrbracket$

$$\hat{\alpha}_{i,t} = \frac{\exp \left(-\eta \sum_{s=1}^{t-1} \left(\hat{X}_{s,T}^{(i)} - X_{s,T} \right)^2 \right)}{\sum_{k=1}^N \exp \left(-\eta \sum_{s=1}^{t-1} \left(\hat{X}_{s,T}^{(k)} - X_{s,T} \right)^2 \right)}$$

Dans les deux cas on prend la convention qu'une somme sur aucun élément est nulle donc $\forall i \in \llbracket 1, N \rrbracket, \hat{\alpha}_{i,1} = 1/N$.

Remarque. Les deux stratégies sont bien en ligne et récursives ce qui signifie

- Pour calculer α_t on n'a besoin que des valeurs des $\left\{ \hat{X}_{s,T}^{(i)}, i \in \llbracket 1, N \rrbracket \right\}$ pour $s \leq t-1$
- A chaque étape peut calculer $\hat{\alpha}_t$ à partir de $\hat{\alpha}_{t-1}$

La construction du prédicteur à partir de ces stratégies est fournie par l'algorithme suivant :

Algorithme 2.3.1 : Agrégation des prédicteurs

Data : $\eta > 0$, strategy (1 ou 2), les prédicteurs $\hat{X}_{t,T}^{(i)}$, les observations $X_{t,T}$

Result : $\hat{X}_{t,T} = \hat{X}_{t,T}^{[\hat{\alpha}_t]}$ pour $t = 1, \dots, T$

Initialization :

```

     $\hat{\alpha}_1 = [1/N, \dots, 1/N];$ 
     $\hat{X}_{1,T} = \sum_{i=1}^N \hat{\alpha}_{i,1} \hat{X}_{1,T}^{(i)};$ 
for  $t = 2, \dots, T$  do
    for  $i = 1, \dots, N$  do
        switch strategy do
            case 1 do
                 $v_{i,t} = \hat{\alpha}_{i,t-1} \exp \left( -2\eta (\hat{X}_{t-1,T} - X_{t-1,T}) \hat{X}_{t-1,T}^{(i)} \right)$ 
            case 2 do
                 $v_{i,t} = \hat{\alpha}_{i,t-1} \exp \left( -\eta (\hat{X}_{t-1,T} - X_{t-1,T})^2 \right)$ 
             $\hat{\alpha}_t = \left( \frac{v_{i,t}}{\sum_{k=1}^N v_{k,t}} \right)_{i=1, \dots, N}$ 
     $\hat{X}_{t,T} = \sum_{i=1}^N \hat{\alpha}_{i,t} \hat{X}_{t,T}^{(i)}$ 
```

Le choix de la stratégie pour mettre à jour les coefficients α^i a une importance sur le prédicteur que l'on souhaite atteindre. En effet la borne de l'erreur n'est pas la même dans les deux cas. Les preuves de ces bornes se trouvent dans [?].

Stratégie 1 : Gradient de l'erreur quadratique La stratégie fait intervenir le gradient de l'erreur quadratique et permet d'approcher la meilleure combinaison convexe dans le simplexe S_N . [?] donne la relation suivante :

$$\frac{1}{T} \sum_{t=1}^T \mathbb{E}[(\hat{X}_t - X_t)^2] \leq \inf_{\alpha \in S_N} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[(\hat{X}_t^{[\alpha]} - X_t)^2] + \frac{\log(N)}{T\eta} + \beta_\eta \quad (2.7)$$

Où β_η contrôle les variations du modèle des prédicteurs choisis et $\hat{X}_t^{[\alpha]}$ est le prédicteur par agrégation obtenu lorsque l'on choisit la combinaison convexe α à l'instant t .

Donc pour $T \rightarrow \infty$, on a bien que le prédicteur obtenu par la stratégie 1 converge bien en moyenne vers le meilleur des prédicteurs par agrégation (on a en fait $\hat{\alpha} \xrightarrow{T \rightarrow \infty} \alpha$).

Stratégie 2 : Erreur quadratique Cette stratégie permet d'obtenir elle aussi un prédicteur par agrégation mais cette fois on ne peut espérer faire mieux que le meilleur des estimateurs donnés en entrée de l'algorithme. [?] donne cette fois la borne :

$$\frac{1}{T} \sum_{t=1}^T \mathbb{E}[(\hat{X}_t - X_t)^2] \leq \min_{1 \leq i \leq N} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[(\hat{X}_t^{(i)} - X_t)^2] + \frac{\log(N)}{T\eta} + \beta_\eta \quad (2.8)$$

Où $X_t^{(i)}$ est le i -ème prédicteur à l'instant t . Cette fois pour $T \rightarrow \infty$, on ne va donc ne trouver que le meilleur des prédicteurs dont on dispose.

Remarque. Le choix de η aussi appelé "taux d'apprentissage" ou encore "learning rate" a à voir avec le bruit dont sont entachées les observations. D'après [?], on peut intuitivement se dire que le bruit dégrade le risque empirique et η permet de quantifier l'importance du risque empirique dans la décision de mise à jour des coefficients α^i . Donc plus le bruit est important et dégrade la qualité des observation et moins on doit se fier au risque empirique (ie : $\eta \rightarrow 0$).

Remarque. La stratégie 1 donne un meilleur résultat théorique que la stratégie 2 mais les conditions d'application de la première stratégie sont plus contraignante alors que la seconde est applicable à n'importe quel ensemble d'échantillons.

2.3.2 Comparaison des risques d'estimation et d'agrégation

On se propose de montrer dans cette sous-section que le risque empirique d'un prédicteur par agrégation (quelque soit la stratégie) est moindre que celui d'un estimateur obtenu classiquement par NLMS.

Proposition 2.3.1. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus $TVAR(d)$ avec d dans \mathbb{N}^* . Soit $(\theta_t)_{t \in \mathbb{Z}}$ la famille de filtres associé tels que $\forall t \in [0, T]$, $X_t = \sum_{k=1}^d \theta_k X_{t-k}$.

On pose $(\hat{X}_t)_{t \in \mathbb{Z}}$ et $(\hat{\theta}_t)_{t \in \mathbb{Z}}$ les estimateurs associés. Alors $\forall t$:

$$\mathbb{E}[(\hat{X}_t - X_t)^2] \leq \mathbb{E}[|\hat{\theta}_t - \theta_t|^2] \mathbb{E}[|X_{t-1}|^4]^{1/2} \quad (2.9)$$

Démonstration. Partons du risque empirique pour une observation et son estimation au temps t :

$$\mathbb{E}[(\hat{X}_t - X_t)^2] = \mathbb{E}[(\hat{\theta}_t - \theta)X_{t-1} - \sigma(t-1)]^2 \quad (2.10)$$

$$\mathbb{E}[(\hat{X}_t - X_t)^2] = \mathbb{E}[(\hat{\theta}_t - \theta)X_{t-1}]^2 + \sigma(t-1)^2 \quad (2.11)$$

On peut alors appliquer l'inégalité de Cauchy-Schwarz à cette expression :

$$\mathbb{E}[(\hat{X}_t - X_t)^2] \leq \mathbb{E}[|\hat{\theta}_t - \theta|^4]^{\frac{1}{2}} \mathbb{E}[|X_{t-1}|^4]^{\frac{1}{2}} \quad (2.12)$$

Or par l'inégalité de Holder, on peut montrer facilement que si f est mesurable et $f \in L^4$, alors $f \in L^2$. D'où l'expression voulue. \square

Chapitre 3

Implémentations

3.1 Etude spectrale sur un TVAR(1) et un TVAR(2) à partir des racines

3.1.1 Cas d'un TVAR(1)