

Résumé du cours de Séries temporelles S2
Université de Strasbourg
M1 DUAS et M1 Statistiques

Davide Giraud

25 avril 2024

Table des matières

1	Introduction	2
1.1	Définition d'une série temporelle	2
1.2	Modélisation d'une série temporelle	2
2	Etude de la partie déterministe de la série temporelle	4
2.1	Estimation paramétrique	4
2.2	Estimation par lissage par moyenne mobile	5
2.3	Estimation par lissage exponentiel	8
3	Processus stochastiques à temps discret	10
3.1	Introduction aux processus stochastiques	10
3.2	Exemples de processus stochastiques	12
3.3	Espérance linéaire et innovation	13
3.4	Processus défini par une équation de récurrence	15
3.5	Processus stationnaires au second ordre : innovation, auto-covariance, prédiction linéaire	16
3.6	Estimation de la moyenne et de la fonction d'auto-corrélation	19
3.7	Fonction génératrice de l'auto-covariance	20
4	Les processus moyennes mobiles	20
4.1	Définition et principales caractéristiques	20
4.2	Estimation d'un modèle moyenne mobile	21

5	Les processus autorégressifs	22
5.1	Définition et principales caractéristiques	22
5.2	Estimation d'un modèle autorégressif	24
5.3	Prédiction d'un processus AR (p)	24
6	Processus ARMA	25
6.1	Définition	25
6.2	Représentation d'un processus ARMA (p, q) en MA (∞)	26
6.3	Représentation d'un ARMA (p, q) en AR (∞)	26
6.4	Innovation d'un processus ARMA	27

1 Introduction

1.1 Définition d'une série temporelle

On observe une série temporelle si on observe une suite $(x_t)_{t=1}^T$ de réels où t représente la date d'une mesure d'une quantité d'intérêt Q et x_t est sa valeur à l'instant t . Par exemple, Q peut être la température, le prix d'un bien, le taux d'humidité, etc...

Une série temporelle $(x_t)_{t=1}^T$ est représentée par l'ensemble des points $\{(t, x_t), t \in \llbracket 1, T \rrbracket\}$ reliés entre eux, où $\llbracket a, b \rrbracket = \{k \in \mathbb{N}, a \leq k \leq b\}$. La plupart du temps, ce type de graphique permet de mettre en évidence les composantes fondamentales d'une série temporelle :

- la tendance, qui représente l'évolution sur le long terme de la série,
- la composante saisonnière, qui induit des creux et sommets sur le graphique et qui correspond à un phénomène périodique (par exemple, la température relevée quotidiennement),
- la composante d'erreur (sans laquelle la série temporelle serait déterministe) qui induit des fluctuations aléatoires,
- la composante accidentelle, qui représente des phénomènes ponctuels mais ayant des conséquences importantes (crach boursier, catastrophe naturelle, etc..). Le traitement de celle-ci dépasse le cadre de cette unité d'enseignement.

L'objectif de ce cours est de présenter des modélisations possibles des séries temporelles afin de faire de la prévision. Plus exactement, si on dispose d'observations $(x_t)_{t=1}^T$ d'une série temporelle, on cherche à prédire la valeur de la série à l'horizon $h \in \mathbb{N}^*$, c'est-à-dire la valeur de x_{T+h} .

1.2 Modélisation d'une série temporelle

Une série temporelle $(x_t)_{t=1}^T$ est la réalisation aux instants $(i_t)_{t=1}^T$ d'un processus stochastique $(X_i)_{i \in E}$, où, pour tout $i \in E$, $X_i: \Omega \rightarrow \mathbb{R}$ est une variable aléatoire réelle définie sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. Si $E = \mathbb{R}$, on parlera de processus à temps continu et si $E = \mathbb{Z}$, de processus à temps discret. Ici, on s'intéressera uniquement au cas discret.

La difficulté réside dans le fait que les variables aléatoires $(X_t)_{t \in \mathbb{Z}}$ sont dépendantes. On impose un modèle de la forme

$$X_t = f(t, \varepsilon_t), t \in \mathbb{Z}, \quad (1.2.1)$$

où ε_t est une variable aléatoire centrée qui représente la composante d'erreur d'une série temporelle et $f: \mathbb{Z} \times \mathbb{R} \rightarrow \mathbb{R}$ est une fonction inconnue. Cependant, ce modèle est trop général d'un point de vue pratique.

1.2.1 Modèle additif

Définition 1.1. Soit $p \in \mathbb{N}^*$. Une série temporelle $(X_t)_{t \in \mathbb{Z}}$ est issue d'un modèle additif de période p s'il existe deux fonctions déterministes inconnues $T: \mathbb{Z} \rightarrow \mathbb{R}$ et $S(\cdot; p): \mathbb{Z} \rightarrow \mathbb{R}$ telles que

$$X_t = T(t) + S(t; p) + \varepsilon_t, \quad (1.2.2)$$

où

- $S(\cdot, p)$ est périodique de période $p \in \mathbb{N}^*$, i.e., pour tout $t \in \mathbb{Z}$, $S(t + p; p) = S(t; p)$,
- ε_t est une variable aléatoire.

La fonction $T(\cdot)$ représente la tendance et la fonction $S(\cdot; p)$ représente la composante saisonnière.

Définition 1.2. Les valeurs $S_i^{(p)} = S(i; p)$, $i \in \llbracket 1, p \rrbracket$, sont appelés coefficients saisonniers.

On a donc, pour $t \in \mathbb{Z}$, $S(t; p) = S_{j(t,p)}^{(p)}$ avec $j(t, p) \in \llbracket 1, p \rrbracket$ et $j(t, p) \equiv t \pmod{p}$.

On supposera souvent que les effets saisonniers se compensent sur une période, c'est-à-dire $\sum_{i=1}^p S_i^{(p)} = 0$.

Exemple 1.3 (Modèle de Buys-Ballot).

$$X_t = \beta_0 + \beta_1 t + S_{j(t,p)}^{(p)} + \varepsilon_t, \quad (1.2.3)$$

où $j(t, p)$ est défini comme précédemment et $\sum_{i=1}^p S_i^{(p)} = 0$.

1.2.2 Modèle multiplicatif

Définition 1.4. Soit $p \in \mathbb{N}^*$. Une série temporelle $(X_t)_{t \in \mathbb{Z}}$ est issue d'un modèle multiplicatif de période p s'il existe deux fonctions déterministes inconnues $T: \mathbb{Z} \rightarrow \mathbb{R}$ et $S(\cdot; p): \mathbb{Z} \rightarrow \mathbb{R}$ telles que

$$X_t = T(t) \times S(t; p) \times (1 + \varepsilon_t), \quad (1.2.4)$$

où

- $S(\cdot, p)$ est périodique de période $p \in \mathbb{N}^*$, i.e., pour tout $t \in \mathbb{Z}$, $S(t + p; p) = S(t; p)$,
- ε_t est une variable aléatoire telle que pour tout $t \in \mathbb{Z}$, $\mathbb{P}(\varepsilon_t > -1) = 1$.

La compensation des effets saisonniers se traduit par $\prod_{i=1}^p S_i^{(p)} = 1$.

- $\mathbb{E} \left[\widehat{\theta}_T^{(\text{MC})} \right] = \theta : \widehat{\theta}_T^{(\text{MC})}$ est sans biais.
- $\widehat{\theta}_T^{(\text{MC})}$ est linéaire car il est de la forme $\widehat{\theta}_T^{(\text{MC})} = L\mathbb{X}_t$ avec $L = (A^\top A)^{-1} A^\top$.
- $\widehat{\theta}_T^{(\text{MC})}$ est de variance minimale parmi les estimateurs sans biais et linéaires.

De plus, si on suppose que $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_T)$, alors $\widehat{\theta}_T^{(\text{MC})}$ est de variance minimale parmi tous les estimateurs sans biais.

La méthode des moindres carrés s'applique à des modèles plus généraux de la forme

$$X_t = g(t; \beta) + S_{j(t,p)}^{(p)} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad (2.1.4)$$

où $g(\cdot; \beta)$ est connue et le vecteur β inconnu.

2.2 Estimation par lissage par moyenne mobile

Une moyenne mobile est un opérateur qui s'applique sur les suites (indexées par \mathbb{Z}), en particulier, sur les séries temporelles.

Définition 2.1. Soit $(x_t)_{t \in \mathbb{Z}}$ une suite de réels. L'opérateur de retard B (Backward shift) est défini par

$$B((x_t)_{t \in \mathbb{Z}}) = (x_{t-1})_{t \in \mathbb{Z}}. \quad (2.2.1)$$

Sa réciproque B^{-1} est l'opérateur avance noté F (Forward shift) et est défini par

$$F((x_t)_{t \in \mathbb{Z}}) = (x_{t+1})_{t \in \mathbb{Z}}. \quad (2.2.2)$$

Autrement dit, le terme d'indice t de la suite $B((x_t)_{t \in \mathbb{Z}})$ est x_{t-1} et le terme d'indice t de la suite $F((x_t)_{t \in \mathbb{Z}})$ est x_{t+1} .

Les opérateurs B et F sont linéaires. Pour $k \in \mathbb{N}$, on note $B^k = \underbrace{B \circ \dots \circ B}_{k \text{ fois}}$, $F^k = \underbrace{F \circ \dots \circ F}_{k \text{ fois}}$

et $B^{-k} := F^k$, $F^{-k} = B^k$. De plus, on note I l'opérateur identité sur les suites indexées par \mathbb{Z} .

Définition 2.2. La moyenne mobile d'ordres $(m_-, m_+) \in \mathbb{N}^2$ et de paramètre $\theta = (\theta_i)_{i=-m_-}^{m_+} \in \mathbb{R}^{m_-+m_++1}$ est l'opérateur linéaire défini par

$$M_{(m_-, m_+, \theta)} = \sum_{i=-m_-}^{m_+} \theta_i B^{-i} = \sum_{i=-m_-}^{m_+} \theta_i F^i. \quad (2.2.3)$$

Étant donnée une série chronologique $(X_t)_{t \in \mathbb{Z}}$, on observe seulement une réalisation du vecteur aléatoire (X_1, \dots, X_T) . Ainsi, on ne peut observer $M_{(m_-, m_+, \theta)}((X_t)_{t \in \mathbb{Z}})$ que pour les instants $t \in \llbracket m_- + 1, T - m_+ \rrbracket$.

Définition 2.3. Une moyenne mobile $M_{(m_-, m_+, \theta)}$ est dite centrée si $m_- = m_+ = m$.

Définition 2.4. Une moyenne mobile $M_{(m_-, m_+, \theta)}$ est dite symétrique si elle est centrée et si le paramètre θ est tel que $\theta_{-i} = \theta_i$ pour tout $i \in \llbracket 1, m \rrbracket$.

Définition 2.5. Le polynôme caractéristique d'une moyenne mobile $M_{(m_-, m_+, \theta)}$ est le polynôme $P(\cdot, \theta)$ d'ordre $m_- + m_+$ tel que $M_{(m_-, m_+, \theta)} = B^{m_-} P(F, \theta)$, c'est-à-dire

$$P(z, \theta) = \theta_{-m_-} + \theta_{-m_-+1}z + \cdots + \theta_{m_+}z^{m_-+m_+} = \sum_{i=0}^{m_-+m_+} \theta_{-i-m_-} z^i. \quad (2.2.4)$$

On fera le léger abus de notation suivant : on note, pour $t \in \mathbb{Z}$, $M_{(m_-, m_+, \theta)} x_t$ l'élément d'indice t de la suite $M_{(m_-, m_+, \theta)}((x_t)_{t \in \mathbb{Z}})$.

On cherche une moyenne mobile $M_{(m_-, m_+, \theta)}$ qui préserve la tendance et annule la composante saisonnière, c'est-à-dire vérifiant

$$M_{(m_-, m_+, \theta)}((T(t)_{t \in \mathbb{Z}})) = (T(t))_{t \in \mathbb{Z}} \text{ et } M_{(m_-, m_+, \theta)}((S(t; p)_{t \in \mathbb{Z}})) = 0. \quad (2.2.5)$$

Définition 2.6. L'espace des invariants d'une moyenne mobile $M_{(m_-, m_+, \theta)}$ est l'ensemble des suites $(x_t)_{t \in \mathbb{Z}}$ vérifiant

$$\forall t \in \mathbb{Z}, \quad M_{(m_-, m_+, \theta)} x_t = x_t. \quad (2.2.6)$$

Le noyau d'une moyenne mobile $M_{(m_-, m_+, \theta)}$ est l'ensemble des suites $(x_t)_{t \in \mathbb{Z}}$ vérifiant

$$\forall t \in \mathbb{Z}, \quad M_{(m_-, m_+, \theta)} x_t = 0. \quad (2.2.7)$$

Proposition 2.7. Soit $q \in \mathbb{N}$. Si la valeur 1 est une racine d'ordre $q+1$ du polynôme $Q(x, \theta) = P(x, \theta) - x^{m_-}$ alors les polynômes d'ordre au plus $q \in \mathbb{N}$ appartiennent à l'espace des invariants d'une moyenne mobile $M_{(m_-, m_+, \theta)}$.

Proposition 2.8. Soit $(S(t))_{t \in \mathbb{Z}}$ une suite périodique de période p , i.e., $S(t) = S_{j(t,p)}^{(p)}$. Soit $M_{(m_-, m_+, \theta)}$ une moyenne mobile admettant $P(\cdot, \theta)$ comme polynôme caractéristique.

1. S'il existe un polynôme $Q(\cdot, \theta)$ tel que

$$P(x, \theta) = Q(x, \theta) (1 - x^p), \quad (2.2.8)$$

alors $(S(t))_{t \in \mathbb{Z}}$ appartient au noyau de la moyenne mobile $M_{(m_-, m_+, \theta)}$.

2. Si $\sum_{i=1}^p S_i^{(p)} = 0$ et qu'il existe un polynôme $Q(\cdot, \theta)$ tel que

$$P(x, \theta) = Q(x, \theta) \sum_{i=0}^{p-1} x^i, \quad (2.2.9)$$

alors $(S(t))_{t \in \mathbb{Z}}$ appartient au noyau de la moyenne mobile $M_{(m_-, m_+, \theta)}$.

On donne à présent un exemple de moyenne mobile qui conserve la tendance et annule les variations saisonnières dans le cas où la tendance est un polynôme de degré 1 et la somme des coefficients saisonniers s'annule.

Définition 2.9. Soit $p \in \mathbb{N}^*$. La moyenne mobile arithmétique $M_p^{(A)}$ est la moyenne mobile centrée et symétrique d'ordres $m_- + m_+ =: m_p^{(A)}$ et de paramètre $\theta_p^{(A)}$, où

$$m_p^{(A)} = \begin{cases} \frac{p}{2} \text{ et } \theta_p^{(A)} = \left(\frac{1}{2p}, \frac{1}{p}, \dots, \frac{1}{p}, \frac{1}{2p} \right) \in \mathbb{R}^{p+1}, & \text{si } p \text{ est pair} \\ \frac{p-1}{2} \text{ et } \theta_p^{(A)} = \left(\frac{1}{p}, \dots, \frac{1}{p} \right) \in \mathbb{R}^p, & \text{si } p \text{ est impair.} \end{cases} \quad (2.2.10)$$

Proposition 2.10. Soit $(\varepsilon_t)_{t \in \mathbb{Z}}$ une suite de variables aléatoires centrées telles que pour tout $(t, t') \in \mathbb{Z}^2$, $\mathbb{E}[\varepsilon_t \varepsilon_{t'}] = \sigma^2 \delta_{t,t'} < \infty$. Soit $M_{(m_-, m_+, \theta)}$ une moyenne mobile avec $\theta \in \mathbb{R}^{m_- + m_+ + 1}$. Pour tout $t \in \mathbb{Z}$, $\mathbb{E}[M\varepsilon_t] = 0$ et $\mathbb{E}[(M\varepsilon_t)^2] = \sigma^2 \theta^\top \theta$. Si de plus $\sum_{i=-m_-}^{m_+} \theta_i = 1$, alors

$$\mathbb{E}[(M\varepsilon_t)^2] \geq \sigma^2 / (m_- + m_+ + 1). \quad (2.2.11)$$

2.2.1 Estimation de la tendance et de la moyenne saisonnière

On considère le modèle additif

$$X_t = T(t) + S(t; p) + \varepsilon_t. \quad (2.2.12)$$

On cherche à estimer, à partir d'observations $(x_t)_{t=1}^T$, la tendance $T(\cdot)$ et les p coefficients saisonniers dont la somme fait 0.

Étape 1 : choix de la moyenne mobile.

On applique à $(x_t)_{t=1}^T$ une moyenne mobile M d'ordres (m_-, m_+) telle que

$$MT(t) = T(t) \quad \text{tendance conservée} \quad (2.2.13)$$

$$MS(t; p) = 0 \quad \text{variation saisonnière supprimée} \quad (2.2.14)$$

Il faut avoir formulé des hypothèses sur S et T , par exemple, $T(t) = a + bt$ et S périodique de période 12, et dans ce cas on prendrait la moyenne arithmétique.

Pour $t \in \llbracket m_- + 1, T - m_+ \rrbracket$, on note $x_t^* = Mx_t$.

Étape 2 : estimation des composantes saisonnières

La série corrigée de la tendance est donnée par

$$\tilde{S}(t; p) = x_t - x_t^*, \quad t \in \llbracket m_- + 1, T - m_+ \rrbracket. \quad (2.2.15)$$

On peut prendre (si $T - m_+ - m_- \geq p$)

$$\tilde{S}_k^{(p)} = \sum_{t=m_-+1}^{m_+} \tilde{S}(t; p) \mathbf{1}_{\{j(t,p)=k\}} / \sum_{t=m_-+1}^{m_+} \mathbf{1}_{\{j(t,p)=k\}}. \quad (2.2.16)$$

Afin d'avoir la compensation sur une période, on prend

$$\hat{S}_k^{(p)} = \tilde{S}_k^{(p)} - \frac{1}{p} \sum_{j=1}^p \tilde{S}_j^{(p)}, \quad k \in \llbracket 1, p \rrbracket. \quad (2.2.17)$$

Étape 3 : estimation de la tendance

La série corrigée des valeurs saisonnières (CVS) est définie pour $t \in \llbracket 1, T \rrbracket$ par

$$x_t^{\text{CVS}} = x_t - \hat{S}_{j(t,p)}^{(p)}. \quad (2.2.18)$$

Pour estimer la tendance, on impose une forme paramétrique de la forme $T(t) = g(t, \beta)$, où $g(\cdot, \cdot)$ est connue et $\beta \in \mathbb{R}^q$ inconnu. Il suffit donc d'estimer β , par exemple avec les moindres carrés :

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^q}{\operatorname{argmin}} \sum_{t=1}^T (x_t^{\text{CVS}} - g(t, \beta))^2 \quad (2.2.19)$$

et on estime la tendance par $\widehat{T}(t) = g(t, \widehat{\beta})$.

On peut prédire la valeur de la série à l'horizon $h \in \mathbb{N}^*$ par

$$\widehat{x}_{T+h} = \widehat{T}(T+h) + \widehat{S}_{j(T+h,p)}^{(p)} \quad (2.2.20)$$

mais on ne peut en général rien dire sur le comportement asymptotique de ces estimateurs. On peut tout de même estimer les valeurs $\{\varepsilon_1, \dots, \varepsilon_T\}$ de l'erreur par

$$\widehat{\varepsilon}_t = x_t - \left(\widehat{T}(t) + \widehat{S}(t, p) \right), \quad t \in \llbracket 1, T \rrbracket. \quad (2.2.21)$$

En étudiant les résidus (cf. Section 3), on peut savoir si ceux-ci contiennent encore une partie saisonnière ou tendancielle, auquel cas on peut remettre le modèle additif en question.

2.3 Estimation par lissage exponentiel

Définition 2.11. Soit $\{x_t, t \in \mathbb{Z}, t \leq T\}$ une série bornée, i.e.,

$$\sup_{t \in \mathbb{Z}: t \leq T} |x_t| < \infty. \quad (2.3.1)$$

La prévision à l'instant $T+1$ par la méthode du lissage exponentiel simple (LES) est donnée par

$$x_{T+1}^{\text{LES}}(\gamma) = (1 - \gamma) \sum_{k=0}^{\infty} \gamma^k x_{T-k}, \quad (2.3.2)$$

où $\gamma \in]0, 1[$ est un paramètre de lissage.

Formule de mise à jour

$$x_{t+1}^{\text{LES}}(\gamma) = \gamma x_t^{\text{LES}}(\gamma) + (1 - \gamma) x_t, \quad t \leq T. \quad (2.3.3)$$

Si $\gamma = 0$, $x_{t+1}^{\text{LES}}(\gamma) = 0$: seul le dernier instant compte. Si $\gamma = 1$, $x_{t+1}^{\text{LES}}(\gamma) = x_t^{\text{LES}}(\gamma)$ (la prévision ne dépend pas du temps).

Obstacle : $x_{T+1}^{\text{LES}}(\gamma)$ fait appel à une infinité d'observations, ce dont on ne dispose pas en pratique. Pour remédier à ce problème, on procède de la manière suivante.

- On dispose d'observations $\{x_1, \dots, x_T\}$.
- On initialise la formule de mise à jour par $x_1^{\text{LES}}(\gamma) = x_0$.
- On applique ensuite pour $t \in \llbracket 1, T \rrbracket$ la formule de mise à jour

$$x_{t+1}^{\text{LES}}(\gamma) = \gamma x_t^{\text{LES}}(\gamma) + (1 - \gamma) x_t, \quad t \leq T. \quad (2.3.4)$$

La série $\{x_t^{\text{LES}}(\gamma), t \in \llbracket 1, T+1 \rrbracket\}$, est appelée série lissée. On peut l'utiliser pour

- travailler avec une série plus propre que celle de départ
- estimer une valeur manquante dans la série initiale.

Comment choisir γ ?

$$\widehat{\gamma} = \operatorname{argmin}_{\gamma \in]0,1[} \sum_{t=1}^{\infty} (x_t^{\text{LES}}(\gamma) - x_t)^2. \quad (2.3.5)$$

Proposition 2.12. *Soit $\{x_t, t \leq T\}$ une série bornée et soit $\gamma \in]0,1[$. Alors*

$$x_{T+1}^{\text{LES}}(\gamma) = \operatorname{argmin}_{b \in \mathbb{R}} \sum_{k=0}^{\infty} \gamma^k (x_{T-k} - b)^2. \quad (2.3.6)$$

$x_{T+1}^{\text{LES}}(\gamma)$ est la meilleure prévision au sens des moindres carrés de la série constante.

Le LES suppose l'absence de composante tendancielle dans la série.

Définition 2.13. *Soit $\{x_t, t \leq T\}$ une série bornée et soit $\gamma \in]0,1[$. La prévision de cette série à l'horizon $h \in \mathbb{N}^*$ par la méthode du lissage exponentiel double (LED) est donnée par :*

$$x_{T+h}^{(\text{LED})}(\gamma) = \widehat{a}_T(\gamma) h + \widehat{b}_T, \quad (2.3.7)$$

avec

$$\left(\widehat{a}_T, \widehat{b}_T \right) = \operatorname{argmin}_{(a,b) \in \mathbb{R}^2} \sum_{k=0}^{\infty} \gamma^k (x_{T-k} + ak - b)^2. \quad (2.3.8)$$

Idée : on ajuste à la série de départ $\{x_t, t \leq T\}$ la série $\{a(T-t) + b, t \leq T\}$ à l'aide de la méthode des moindres carrés pondérés.

Proposition 2.14. *On pose*

$$L_{T+1}^{(1)}(\gamma) = (1 - \gamma) \sum_{k=0}^{\infty} \gamma^k x_{T-k}, \quad (2.3.9)$$

$$L_{T+1}^{(2)}(\gamma) = (1 - \gamma) \sum_{k=0}^{\infty} \gamma^k L_{T+1-k}^{(1)}. \quad (2.3.10)$$

Les paramètres du lissage exponentiel double sont donnés par

$$\widehat{a}_T(\gamma) = \frac{1 - \gamma}{\gamma} \left(L_{T+1}^{(1)}(\gamma) - L_{T+1}^{(2)}(\gamma) \right) \quad (2.3.11)$$

$$\widehat{b}_T(\gamma) = 2 \left(L_{T+1}^{(1)}(\gamma) - L_{T+1}^{(2)}(\gamma) \right) \quad (2.3.12)$$

On effectue deux lissages exponentiels successifs :

- le premier sur la série $\{x_t, t \leq T\}$,
- le second sur la série lissée $\left\{ L_t^{(1)}(\gamma), t \leq T + 1 \right\}$.

Proposition 2.15. *On a également les formules de mise à jour suivantes.*

$$\widehat{b}_t(\gamma) = (1 - \gamma^2) x_t + \gamma^2 \left(\widehat{a}_{t-1}(\gamma) - \widehat{b}_{t-1}(\gamma) \right) \quad (2.3.13)$$

$$\widehat{a}_t(\gamma) = \frac{2\gamma}{1 + \gamma} \widehat{a}_{t-1}(\gamma) + \frac{1 - \gamma}{1 + \gamma} \left(\widehat{b}_t(\gamma) - \widehat{b}_{t-1}(\gamma) \right). \quad (2.3.14)$$

3 Processus stochastiques à temps discret

Dans la section précédente, nous avons traité la partie déterministe d'une série temporelle de la forme $X_t = f(t) + \varepsilon_t$, où la partie déterministe (tendance et composante saisonnière) est représentée par $f(\cdot)$ et ε_t la partie aléatoire. Nous n'avons pas fait d'hypothèse sur cette dernière pour estimer f . En notant $\widehat{f}_T(\cdot)$ l'estimation de $f(\cdot)$ obtenue, on dispose ainsi d'estimations

$$\widehat{\varepsilon}_t := X_t - \widehat{f}_T(t), \quad t \in \llbracket 1, T \rrbracket, \quad (3.0.1)$$

des variables aléatoires $\varepsilon_1, \dots, \varepsilon_T$. La collection des $\widehat{\varepsilon}_t$ est appelée résidus.

Deux cas de figure se présentent.

1. Des tests statistiques laissent à penser que les résidus sont issus de variables aléatoires non corrélées. Dans ce cas, $\widehat{f}_T(\cdot)$ est une bonne estimation de la tendance.
2. Si on rejette à l'aide de ces test l'hypothèse que les résidus proviennent de réalisations de variables aléatoires non-corrélées. De l'information est encore contenue dans les résidus, pour lesquels une étude plus approfondie est nécessaire.

3.1 Introduction aux processus stochastiques

Nous allons étudier les termes aléatoires d'une série temporelle. Ils seront notés, pour faciliter les notations, $(X_t)_{t \in \mathbb{Z}}$ au lieu de $(\varepsilon_t)_{t \in \mathbb{Z}}$. Nous supposons que les processus étudiés sont de carré intégrable et nous ferons des rappels sur les espaces \mathbb{L}^2 .

Définition 3.1. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé.

1. On note $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ l'espace des variables aléatoires $X: (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telles que

$$\int_{\Omega} X(\omega)^2 d\mathbb{P}(\omega) = \mathbb{E}[X^2] < \infty. \quad (3.1.1)$$

2. On note $\mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ l'ensemble des classes d'équivalence pour l'égalité presque sûre des variables aléatoires de $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$.

Proposition 3.2. L'ensemble $\mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ est un espace vectoriel sur \mathbb{R} . De plus, l'application $\langle \cdot, \cdot \rangle$ définie pour $X, Y \in \mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ par

$$\langle X, Y \rangle := \int_{\Omega} X(\omega) Y(\omega) d\mathbb{P}(\omega) = \mathbb{E}[XY] \quad (3.1.2)$$

est une forme hermitienne définie positive, donc un produit scalaire.

Norme induite :

$$\|X\|_2 = \sqrt{\langle X, X \rangle} = \sqrt{\mathbb{E}[X^2]}. \quad (3.1.3)$$

Proposition 3.3 (Inégalité de Cauchy-Schwarz). Soient $X, Y \in \mathbb{L}^2$. Alors

$$|\langle X, Y \rangle| \leq \|X\|_2 \|Y\|_2, \quad (3.1.4)$$

ou, de manière équivalente,

$$(\mathbb{E}[XY])^2 \leq \mathbb{E}[X^2] \mathbb{E}[Y^2]. \quad (3.1.5)$$

Définition 3.4. Soit $(X_n)_{n \geq 1}$ une suite d'éléments de \mathbb{L}^2 et soit $X \in \mathbb{L}^2$. On dit que $(X_n)_{n \geq 1}$ converge vers X dans \mathbb{L}^2 si

$$\lim_{n \rightarrow \infty} \|X_n - X\|_2 = 0 \quad (3.1.6)$$

ou de manière équivalente, si

$$\lim_{n \rightarrow \infty} \mathbb{E} [(X_n - X)^2] = 0. \quad (3.1.7)$$

Proposition 3.5. Soient $(X_n)_{n \geq 1}$ et $(Y_n)_{n \geq 1}$ des suites d'éléments de \mathbb{L}^2 , convergent respectivement vers $X \in \mathbb{L}^2$ et $Y \in \mathbb{L}^2$. Alors $\lim_{n \rightarrow \infty} \langle X_n, Y_n \rangle = \langle X, Y \rangle$.

Définition 3.6. On dit que $X, Y \in \mathbb{L}^2$ sont orthogonales si $\langle X, Y \rangle = 0$, ou de manière équivalente, $\mathbb{E}[XY] = 0$.

Théorème 3.7. L'espace $\mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de Hilbert.

Théorème 3.8 (Théorème de la projection orthogonale). Soit \mathbb{H} un espace de Hilbert et F un sous-ensemble convexe fermé. Pour tout $x \in \mathbb{H}$, il existe un unique $p_F(x) \in F$ tel que

$$\|x - p_F(x)\|_2 = \inf_{y \in F} \|x - y\|_2. \quad (3.1.8)$$

Si F est de plus un sous-espace vectoriel fermé, le point $p_F(x)$ est l'unique point de F tel que

$$x - p_F(x) \in F^\perp := \{u \in \mathbb{H}, \forall v \in F, \langle u, v \rangle = 0\}. \quad (3.1.9)$$

Proposition 3.9. Soit \mathbb{H} un espace de Hilbert et soit F un sous-espace vectoriel fermé.

1. L'application $p_F: \mathbb{H} \rightarrow F$ est linéaire.
2. Si $x \in F^\perp$, alors $p_F(x) = 0$.
3. Si F est la somme de deux sous-espaces fermés orthogonaux F_1 et F_2 , alors $P_F(x) = P_{F_1}(x) + P_{F_2}(x)$.
4. Si F_1 et F_2 sont des sous-espaces vectoriels fermés tels que $F_1 \subset F_2$, alors pour tout $x \in \mathbb{H}$,

$$P_{F_1}(P_{F_2}(x)) = P_{F_2}(P_{F_1}(x)) = P_{F_1}(x). \quad (3.1.10)$$

Définition 3.10. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et soit $E \subset \mathbb{R}$. Pour tout $t \in E$, soit $X_t: (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une variable aléatoire. La famille $(X_t)_{t \in E}$ est appelée processus stochastique d'espace d'états E .

1. Si E est dénombrable, on dit que le processus $(X_t)_{t \in E}$ est à temps discret.
2. Si E n'est pas dénombrable, on dit que le processus $(X_t)_{t \in E}$ est à temps continu.

Définition 3.11. Soit $(X_t)_{t \in E}$ un processus stochastique défini sur $(\Omega, \mathcal{F}, \mathbb{P})$. Les lois finidimensionnelles de ce processus sont les mesure μ_{t_1, \dots, t_k} sur $\mathcal{B}(\mathbb{R}^k)$ définies pour $k \in \mathbb{N}^*$, $t_1, \dots, t_k \in E$ et $B_1, \dots, B_k \in \mathcal{B}(\mathbb{R})$ par

$$\mu_{t_1, \dots, t_k}(B_1 \times \dots \times B_k) = \mathbb{P}(\omega, \forall \ell \in \llbracket 1, k \rrbracket, X_{t_\ell}(\omega) \in B_\ell). \quad (3.1.11)$$

Proposition 3.12. *Si les processus $(X_t)_{t \in E}$ et $(Y_t)_{t \in E}$ ont les mêmes lois fini-dimensionnelles, alors ils ont la même loi.*

Définition 3.13. *On note*

$$\mathcal{M} := \{\mu_{t_1, \dots, t_k}, k \in \mathbb{N}^*, t_1, \dots, t_k \in E\}, \quad (3.1.12)$$

où pour chaque t_1, \dots, t_k , μ_{t_1, \dots, t_k} est une mesure sur $\mathcal{B}(\mathbb{R}^k)$. La famille de mesures \mathcal{M} est consistante si pour tout $k \geq 1$ et tous $B_1, \dots, B_k \in \mathcal{B}(\mathbb{R})$

1. l'égalité

$$\mu_{t_1, \dots, t_k}(B_1 \times \dots \times B_k) = \mu_{t_{\sigma(1)}, \dots, t_{\sigma(k)}}(B_{\sigma(1)}, \dots, B_{\sigma(k)}) \quad (3.1.13)$$

a lieu pour toute permutation $\sigma: \llbracket 1, k \rrbracket \rightarrow \llbracket 1, k \rrbracket$ et

2. pour tout $m \geq k$ et tous t_1, \dots, t_m ,

$$\mu_{t_1, \dots, t_k, t_{k+1}, \dots, t_m}(B_1 \times \dots \times B_k \times \mathbb{R} \times \dots \times \mathbb{R}) = \mu_{t_1, \dots, t_k}(B_1 \times \dots \times B_k). \quad (3.1.14)$$

Théorème 3.14. *Si la famille \mathcal{M} est consistante, alors il existe un processus stochastique $(X_t)_{t \in E}$ défini sur $(\Omega, \mathcal{F}, \mathbb{P})$ tel que pour tous $k \in \mathbb{N}^*$, $t_1, \dots, t_k \in E$ et $B_1, \dots, B_k \in \mathcal{B}(\mathbb{R})$ par*

$$\mu_{t_1, \dots, t_k}(B_1 \times \dots \times B_k) = \mathbb{P}(\omega, \forall \ell \in \llbracket 1, k \rrbracket, X_{t_\ell}(\omega) \in B_\ell). \quad (3.1.15)$$

Définition 3.15. *On dit qu'un processus $(X_t)_{t \in \mathbb{Z}}$ est ergodique en moyenne quadratique si pour chaque t , X_t est de carré intégrable et $\mathbb{E}[X_t] = \mathbb{E}[X_0]$ et si pour chaque $t_0 \in \mathbb{Z}$,*

$$\lim_{T \rightarrow \infty} \mathbb{E} \left[\left(\frac{1}{T} \sum_{t=t_0}^{T+t_0-1} X_t - \mathbb{E}[X_0] \right)^2 \right] = 0. \quad (3.1.16)$$

3.2 Exemples de processus stochastiques

Définition 3.16. *Un processus stochastique $(\eta_t)_{t \in \mathbb{Z}}$ est appelé bruit blanc fort si la collection de variables aléatoires $\{\eta_t, t \in \mathbb{Z}\}$ est indépendante et pour tout $t \in \mathbb{Z}$, $\mathbb{E}[\eta_t^2] = \sigma^2$ et $\mathbb{E}[\eta_t] = 0$.*

Définition 3.17. *Un processus stochastique $(\eta_t)_{t \in \mathbb{Z}}$ est appelé bruit blanc faible si pour tout $t \in \mathbb{Z}$, $\mathbb{E}[\eta_t^2] = \sigma^2$ et $\mathbb{E}[\eta_t] = 0$, et pour tous $s, t \in \mathbb{Z}$ tels que $s \neq t$, $\text{Cov}(\eta_s, \eta_t) = 0$.*

Dans la suite, « bruit blanc » fera référence au bruit blanc faible.

Définition 3.18. *On dit que le processus stochastique $(X_t)_{t \in \mathbb{Z}}$ est stationnaire au second ordre si pour tout $t \in \mathbb{Z}$, $\mathbb{E}[X_t^2]$ est finie, $\mathbb{E}[X_t] = \mathbb{E}[X_0]$ et pour tous $h, t \in \mathbb{Z}$,*

$$\text{Cov}(X_t, X_{t+h}) = \text{Cov}(X_0, X_h) =: \gamma(h). \quad (3.2.1)$$

Définition 3.19. *La fonction $\gamma(\cdot) : \mathbb{Z} \rightarrow \mathbb{R}$ définie par (3.2.1) est appelée fonction d'auto-covariance.*

Définition 3.20. On dit que $(X_t)_{t \in \mathbb{Z}}$ est un processus stationnaire au sens strict si pour tous $h, k \in \mathbb{N}^*$, tous $t_1, \dots, t_k \in \mathbb{Z}$, les vecteurs $(X_{t_1}, \dots, X_{t_k})$ et $(X_{t_1+h}, \dots, X_{t_k+h})$ ont la même loi.

Définition 3.21. La marche aléatoire est un processus $(X_t)_t$ défini par $X_t = 0$ pour $t \leq 0$ et $X_t = \sum_{s=1}^t \varepsilon_s$, où $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc faible.

Définition 3.22. Un processus $(X_t)_{t \in \mathbb{Z}}$ est gaussien si pour tout k , tous $t_1, \dots, t_k \in \mathbb{Z}$ et tous $c_1, \dots, c_k \in \mathbb{R}$, la variable aléatoire $\sum_{\ell=1}^k c_\ell X_{t_\ell}$ est de loi normale (de variance potentiellement nulle).

Définition 3.23. Un processus $(B_n)_{n \in \mathbb{N}}$ est un mouvement brownien standard si

1. $B_0 = 0$ presque sûrement,
2. pour tous $s, t \in \mathbb{N}$ tels que $t > s$, $B_t - B_s$ est de loi normale centrée et de variance $t - s$ et
3. pour tous $s, t \in \mathbb{N}$ tels que $t > s$, $B_t - B_s$ est indépendant de $(B_u)_{u=1}^s$.

Proposition 3.24. Le processus $(B_n)_{n \in \mathbb{N}}$ est gaussien et vérifie $\text{Cov}(B_s, B_t) = \min\{s, t\}$.

Définition 3.25. On dit qu'une suite de réels $(\Phi_t)_{t \in \mathbb{Z}}$ est absolument sommable si $\sum_{t \in \mathbb{Z}} |\Phi_t| < \infty$.

Proposition 3.26. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stochastique défini sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. On suppose que $\sup_{t \in \mathbb{Z}} \mathbb{E}[|X_t|] < \infty$. Soit $(\Phi_t)_{t \in \mathbb{Z}}$ une suite absolument sommable. Alors pour chaque $t \in \mathbb{Z}$, la suite de processus $(Y_{n,t})_{n \geq 1}$ définie par

$$Y_{n,t} = \sum_{i=-n}^n \phi_i B^i X_t = \sum_{i=-n}^n \phi_i X_{t-i} \quad (3.2.2)$$

converge presque sûrement, quand n tend vers l'infini, vers

$$Y_t := \sum_{i \in \mathbb{Z}} \phi_i X_{t-i}. \quad (3.2.3)$$

Si on suppose de plus que $\sup_{t \in \mathbb{Z}} \mathbb{E}[X_t^2] < \infty$, alors la convergence a lieu dans \mathbb{L}^2 .

Définition 3.27. Le processus $(Y_t)_{t \in \mathbb{Z}}$ défini par (3.2.3) est appelé moyenne mobile infinie.

3.3 Espérance linéaire et innovation

Définition 3.28. Soit I un ensemble au plus dénombrable et soit $(X_i)_{i \in I}$ une famille de variables aléatoires telles que pour tout $i \in I$, $X_i \in \mathbb{L}^2$. On note $\overline{\text{span}}(X_i, i \in I)$ le plus petit (pour l'inclusion) sous-espace vectoriel de $\mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ contenant toutes les combinaisons linéaires finies de variables aléatoires prises dans l'ensemble $\{X_i, i \in I\}$.

Définition 3.29. L'espérance linéaire d'une variable aléatoire $Y \in \mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ sur l'ensemble $(X_i)_{i \in I}$ est notée $\mathbb{E}L(Y | (X_i)_{i \in I})$ est la projection orthogonale de Y sur le sous-espace vectoriel fermé $\overline{\text{span}}(X_i, i \in I)$. On a donc

$$\|Y - \mathbb{E}L(Y | (X_i)_{i \in I})\|_2 = \inf \{\|Y - Z\|_2, Z \in \overline{\text{span}}(X_i, i \in I)\}. \quad (3.3.1)$$

Si $I = (i_\ell)_{\ell \geq 1}$ est infini dénombrable, alors

$$\mathbb{E} L(Y | (X_i)_{i \in I}) = \lim_{N \rightarrow \infty} \mathbb{E} L(Y | (X_{i_\ell})_{\ell=1}^N), \quad (3.3.2)$$

où la limite est prise dans \mathbb{L}^2 . Par le Théorème de la projection orthogonale, la variable aléatoire U définie par

$$U = Y - \mathbb{E} L(Y | (X_i)_{i \in I}) \quad (3.3.3)$$

appartient à l'orthogonal de $\overline{\text{span}}(X_i, i \in I)$.

S'il existe $i \in I$ tel que X_i est constante, alors U est orthogonale à une constante et donc nécessairement centrée.

Soient Y et $X_i, i \in I$ des variables aléatoires de carré intégrable. On note $Y' := Y - \mathbb{E}[Y]$ et $X'_i = X_i - \mathbb{E}[X_i]$. On a alors

$$\mathbb{E} L(Y | 1, (Y_i)_{i \geq 1}) = \mathbb{E}[Y] + \mathbb{E} L(Y' | (X'_i)_{i \in I}) \quad (3.3.4)$$

Par conséquent, si les variables aléatoires Y et X_i sont centrées, on a

$$\mathbb{E} L(Y | 1, (Y_i)_{i \geq 1}) = \mathbb{E} L(Y | (X_i)_{i \in I}). \quad (3.3.5)$$

Pour éviter l'introduction d'une constante, on supposera donc les variables aléatoires centrées.

Il ne faut pas confondre l'espérance linéaire avec l'espérance conditionnelle. Si on pose $F_1 = \overline{\text{span}}(X_i, i \in I)$ et $F_2 = \mathbb{L}^2(\Omega, \sigma(Y_i, i \in I), \mathbb{P})$ alors $F_1 \subset F_2$ et l'inclusion est stricte en général. Dans le cas simple où I ne contient qu'un élément noté $X \in \mathbb{L}^4$, $\mathbb{E} L(X^2 | X) = cX$ pour une certaine constante c , alors que $\mathbb{E}[X^2 | X] = X^2$.

En revanche,

Proposition 3.30. *Si $I = \llbracket 1, n \rrbracket$ est fini, et (Y, X_1, \dots, X_n) est gaussien centré, alors*

$$\mathbb{E} L(Y | (X_i)_{i=1}^n) = \mathbb{E}[Y | (X_i)_{i=1}^n]. \quad (3.3.6)$$

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus défini sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et tel que pour tout $t \in \mathbb{Z}$, $X_t \in \mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. On pose

$$\mathcal{H}_t^X = \overline{\text{span}}(X_s, s \in \mathbb{Z}, s \leq t). \quad (3.3.7)$$

L'espace \mathcal{H}_t^X est appelé histoire du processus $(X_t)_{t \in \mathbb{Z}}$ jusqu'à l'instant t .

Remarquons que $\mathcal{H}_t^X \subset \mathcal{H}_{t+1}^X$. On note

$$\mathcal{H}_{-\infty}^X = \bigcap_{t \in \mathbb{Z}} \mathcal{H}_t^X \quad (3.3.8)$$

le passé lointain du processus $(X_t)_{t \in \mathbb{Z}}$.

Définition 3.31. *Soit $(X_t)_{t \in \mathbb{Z}}$ un processus défini sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et tel que pour tout $t \in \mathbb{Z}$, $X_t \in \mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ et $\mathbb{E}[X_t] = 0$. L'innovation de ce processus est le processus $(\varepsilon_t)_{t \in \mathbb{Z}}$ donné par*

$$\varepsilon_t = X_t - \mathbb{E} L(X_t | \mathcal{H}_{t-1}^X). \quad (3.3.9)$$

3.4 Processus défini par une équation de récurrence

On suppose que $(X_t)_{t \in \mathbb{Z}}$ vérifie une relation de récurrence de la forme

$$P(B)X_t = Y_t, \quad (3.4.1)$$

où B est l'opérateur de retard et P un polynôme à coefficients réels. La première question qui se pose est l'existence d'un tel processus. De plus, on aimerait avoir une expression de X_t en fonction de $(Y_t)_{t \in \mathbb{Z}}$.

Définition 3.32. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et

$$V = \left\{ (X_t)_{t \in \mathbb{Z}}, X_t: \Omega \rightarrow \mathbb{R}, \sup_{t \in \mathbb{Z}} \mathbb{E} [|X_t|] < \infty \right\}, \quad (3.4.2)$$

l'espace des processus à temps discret bornés dans \mathbb{L}^1 .

Lemme 3.33. On pose

$$\| (X_t)_{t \in \mathbb{Z}} \|_V := \sup_{t \in \mathbb{Z}} \mathbb{E} [|X_t|]. \quad (3.4.3)$$

Alors V muni de $\|\cdot\|_V$ est un espace de Banach.

Lemme 3.34. Soit $B: V \rightarrow V$ défini par $B((X_t)_{t \in \mathbb{Z}}) = (X_{t-1})_{t \in \mathbb{Z}}$. Alors B est linéaire et continu. De plus, $\|B\|_{\mathcal{L}(V)} := \sup_{\|(X_t)_{t \in \mathbb{Z}}\|_V=1} \|B((X_t)_{t \in \mathbb{Z}})\|_V = 1$

Lemme 3.35. Soit $\lambda \in \mathbb{C}$ tel que $|\lambda| \neq 1$. Soit $P(B) = I - \lambda B: V \rightarrow V$. Alors $P(B)$ est inversible, d'inverse $(P(B))^{-1}$ défini de la manière suivante.

- Si $|\lambda| < 1$,

$$(P(B))^{-1} = \sum_{j=0}^{\infty} \lambda^j B^j, \quad (3.4.4)$$

- si $|\lambda| > 1$,

$$(P(B))^{-1} = - \sum_{j=-\infty}^{-1} \lambda^j B^j. \quad (3.4.5)$$

Lemme 3.36. Soit $\lambda \in \mathbb{C}$ tel que $|\lambda| \neq 1$. Soit $P(B) = (I - \lambda B)(I - \bar{\lambda} B): V \rightarrow V$. Alors $P(B)$ est inversible, d'inverse $(P(B))^{-1}$ défini de la manière suivante.

- Si $|\lambda| < 1$,

$$(P(B))^{-1} = \sum_{j=0}^{\infty} \sum_{\ell=0}^j \bar{\lambda}^\ell \lambda^{\ell-j} B^j, \quad (3.4.6)$$

- si $|\lambda| > 1$,

$$(P(B))^{-1} = - \sum_{j=-\infty}^0 \sum_{k=j+1}^{-1} \bar{\lambda}^j \lambda^{k-j} B^j. \quad (3.4.7)$$

Proposition 3.37. Soit P un polynôme à coefficients réels tel que $P(0) = 1$ et vérifiant $P(z) \neq 0$ pour tout $z \in \mathbb{C}$ tel que $|z| = 1$. Alors $P(B) : V \rightarrow V$ est inversible. Plus précisément, il existe une famille de réels absolument sommable $(a_j)_{j \in \mathbb{Z}}$ telle que

$$(P(B))^{-1} = \sum_{j \in \mathbb{Z}} a_j B^j. \quad (3.4.8)$$

Si on suppose de plus que $P(\cdot)$ n'a aucune racine dans le disque unité fermé, alors $P^{-1}(\cdot)$ admet la représentation

$$(P(B))^{-1} = \sum_{j=0}^{\infty} a_j B^j. \quad (3.4.9)$$

Proposition 3.38. Soit $(Y_t)_{t \in \mathbb{Z}}$ un processus dont les moments d'ordre 1 sont uniformément bornés, c'est-à-dire

$$\sup_{t \in \mathbb{Z}} \mathbb{E} [|Y_t|] < \infty. \quad (3.4.10)$$

1. On suppose que $P(\cdot)$ est un polynôme n'ayant aucune racine sur le cercle unité de \mathbb{C} . Alors l'équation $P(B) X_t = Y_t$ admet une unique solution $(X_t)_{t \in \mathbb{Z}}$ dont l'expression est donnée par

$$X_t = (P(B))^{-1} Y_t = \sum_{j \in \mathbb{Z}} a_j Y_{t-j}, \quad t \in \mathbb{Z}, \quad (3.4.11)$$

où les coefficients $(a_j)_{j \in \mathbb{Z}}$ sont absolument sommables.

2. On suppose que $P(\cdot)$ est un polynôme n'ayant aucune racine sur le disque unité de \mathbb{C} . Alors l'équation $P(B) X_t = Y_t$ admet une unique solution $(X_t)_{t \in \mathbb{Z}}$ dont l'expression est donnée par

$$X_t = (P(B))^{-1} Y_t = \sum_{j=0}^{\infty} a_j Y_{t-j}, \quad t \in \mathbb{Z}, \quad (3.4.12)$$

où les coefficients $(a_j)_{j \geq 0}$ sont absolument sommables.

Définition 3.39. Dans le cas 1, le processus $(X_t)_{t \in \mathbb{Z}}$ est inversible et dans le cas 2, le processus $(X_t)_{t \in \mathbb{Z}}$ est causal.

3.5 Processus stationnaires au second ordre : innovation, auto-covariance, prédiction linéaire

Proposition 3.40. Soit $(Y_t)_{t \in \mathbb{Z}}$ un processus stationnaire au second ordre, centré, de fonction d'auto-covariance $\gamma_Y(\cdot)$ et soit $(a_j)_{j \in \mathbb{Z}}$ une famille de réels absolument sommable. Le processus $(X_t)_{t \in \mathbb{Z}}$ défini par

$$Y_t = \sum_{j \in \mathbb{Z}} a_j X_{t-j} \quad (3.5.1)$$

est stationnaire, de fonction d'auto-covariance

$$\gamma_Y(h) = \sum_{i,j \in \mathbb{Z}} a_i a_{j+h} \gamma_X(i-j). \quad (3.5.2)$$

Définition 3.41. Une fonction $K: \mathbb{Z} \rightarrow \mathbb{R}$ est de type positif si pour tout entier $n \geq 1$, tous $a_1, \dots, a_n \in \mathbb{R}$ et tous $t_1, \dots, t_n \in \mathbb{Z}$,

$$\sum_{k,\ell=1}^n a_k a_\ell K(t_k - t_\ell) \geq 0. \quad (3.5.3)$$

Proposition 3.42. Une fonction $K: \mathbb{Z} \rightarrow \mathbb{R}$ est la fonction d'auto-covariance d'un processus stationnaire au second ordre si et seulement si elle est paire et de type positif.

Définition 3.43. La fonction d'auto-corrélation d'un processus stationnaire au second ordre $(X_t)_{t \in \mathbb{Z}}$ de variance σ^2 est la fonction $\rho: \mathbb{Z} \rightarrow \mathbb{R}$ définie par

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\text{Cov}(X_t, X_{t+h})}{\sigma^2}. \quad (3.5.4)$$

La quantité $\rho(h)$ est le coefficient de corrélation linéaire entre X_t et X_{t+h} .

On remarque que $\rho(-h) = \rho(h)$ et par l'inégalité de Cauchy-Schwarz, $|\rho(h)| \leq 1$.

On cherche à mesurer la dépendance directe entre X_t et X_{t+h} .

Définition 3.44. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire au second ordre. Pour $(s, t) \in \mathbb{Z}^2$ tel que $s \leq t$, on note $H_s^t = \{\sum_{i=s}^t c_i X_i, (c_i)_{s \leq i \leq t} \in \mathbb{R}\}$. Pour $h \geq 2$, on note

$$U_h := X_t - \mathbb{E}L(X_t | (X_i)_{t-h+1 \leq i \leq t-1}), \quad V_h := X_{t-h} - \mathbb{E}L(X_{t-h} | (X_i)_{t-h+1 \leq i \leq t-1}). \quad (3.5.5)$$

La fonction d'auto-corrélation partielle du processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ est définie pour $h \geq 2$ par

$$\tau(h) = \frac{\text{Cov}(U_h, V_h)}{(\text{Var}(U_h) \text{Var}(V_h))^{1/2}}. \quad (3.5.6)$$

On peut définir $\tau(h)$ pour $h \leq -2$ par $\tau(h) = \tau(-h)$. On pose également

$$\tau(0) = 0 \quad \tau(1) = \tau(-1) = \frac{\gamma(1)}{\gamma(0)}. \quad (3.5.7)$$

Proposition 3.45. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire au second ordre centré. Alors pour tout $t \in \mathbb{Z}$ et tout $h \geq 1$,

$$\mathbb{E}L(X_t | (X_i)_{t-h \leq i \leq t-1}) = \sum_{\ell=1}^h \tau(\ell) X_{t-\ell}. \quad (3.5.8)$$

Pour $h \in \mathbb{N}^*$ et $t \in \mathbb{Z}$, on note $\Gamma^{(h)}$ la matrice de covariance du vecteur $\mathbb{X} = (X_{t-1}, \dots, X_{t-h})^\top$. Autrement dit, $\Gamma_{i,j}^{(h)} = \gamma(|i-j|) / \gamma(0)$. On note également par $\tilde{\rho}_h = (\rho(1), \dots, \rho(h))^\top$ le vecteur des h premières valeurs de la fonction d'auto-corrélation.

Proposition 3.46. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire au second ordre centré. Soit $\Lambda = (\lambda_1, \dots, \lambda_h)$ tel que $\mathbb{E}L(X_t | H_{t-h}^{t-1}) = \Lambda^\top \mathbb{X}$. Alors $\tilde{\rho}_h = \Gamma^{(h)} \Lambda$.

On cherche à présent à exprimer la meilleure prédiction linéaire à l'horizon $h \in \mathbb{N}^*$, donnée par la projection orthogonale de X_{T+h} sur l'espace vectoriel engendré par X_1, \dots, X_T , noté H_T .

Proposition 3.47. *Soit $h \in \mathbb{N}^*$ et soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire au second ordre et centré de fonction d'auto-corrélation $\rho(\cdot)$. Soit Γ_T la matrice de corrélation du vecteur aléatoire $\mathbb{X} = (X_1, \dots, X_T)^\top$. Soit $\tilde{\rho}_h := (\rho(T+h-1), \dots, \rho(h))^\top$. Si Γ_T est inversible, alors les coefficients $\tilde{\Psi}_T = (\Psi_{1,h}, \dots, \Psi_{T,h})^\top$ de la régression linéaire*

$$\mathbb{E} L(X_{T+h} | H_T) = \sum_{\ell=1}^T \Psi_{\ell,h} X_\ell \quad (3.5.9)$$

sont donnés par $\tilde{\Psi}_T = \Gamma_T^{-1} \tilde{\rho}_T$. De plus,

$$\mathbb{E} [(X_{T+h} - \mathbb{E} L(X_{T+h} | H_T))^2] = \gamma(0) \left(1 - \tilde{\Psi}_T^\top \tilde{\rho}_T\right). \quad (3.5.10)$$

L'espace vectoriel des combinaisons linéaires des variables aléatoires X_1, \dots, X_t est le même que l'espace vectoriel constitué des combinaisons linéaires des variables aléatoires $X_i - \mathbb{E} L(X_i | H_{i-1})$, avec $H_{i-1} = \text{vect}(X_1, \dots, X_{i-1})$ et $\mathbb{E} L(X_1 | H_0) = \mathbb{E}[X_1]$. Par conséquent, on peut écrire

$$\mathbb{E} [X_{t+1} | H_t] = \sum_{i=1}^t \theta_i^{(t)} (X_{t+1-i} - \mathbb{E} L(X_{t+1-i} | H_{-it})). \quad (3.5.11)$$

L'algorithme des innovations permet de calculer récursivement les coefficients $\theta_i^{(t)}$ ainsi que le terme d'erreur

$$v_t = \mathbb{E} [(X_{t+1} - \mathbb{E} [X_{t+1} | H_t])^2]. \quad (3.5.12)$$

Proposition 3.48. *Soit $(X_t)_{t \in \mathbb{Z}}$ un processus centré défini sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et tel que $\mathbb{E}[X_t^2] < \infty$ pour tout $t \in \mathbb{Z}$. On suppose que pour tout $t \geq 1$, la matrice de covariance du vecteur (X_1, \dots, X_t) est inversible. Alors les coefficients $(\theta_i^{(t)})_{i \in [1,t]}$ dans (3.5.11) et les erreurs moyennes v_t dans (3.5.12) par*

$$v_0 = \mathbb{E}[X_1^2] \quad (3.5.13)$$

$$\theta_t^{(t)} = v_0^{-1} \mathbb{E}[X_{t+1} X_1] \quad (3.5.14)$$

$$\theta_{t-i}^{(t)} = v_i^{-1} \left(\mathbb{E}[X_{t+1} X_{i+1}] - \sum_{j=0}^{i-1} \theta_{j-i}^{(i)} \theta_{t-i}^{(t)} v_j \right) \quad (3.5.15)$$

$$v_t = \mathbb{E}[X_{t+1}^2] - \sum_{j=0}^{t-1} \left(\theta_{t-j}^{(t)} \right)^2 v_j. \quad (3.5.16)$$

En notant $S_{i,j} = \mathbb{E}[X_i X_j]$, on voit que l'on peut calculer récursivement les v_t et $\theta_{t-i}^{(t)}$. En effet, $v_0 = S_{1,1}$, $\theta_1^{(1)} = S_{1,2}/S_{1,1}$ et $v_1 = S_{2,2} - \left(\theta_1^{(1)}\right)^2 S_{1,1}$ puis

$$\theta_2^{(2)} = S_{1,3}/S_{1,1}, \quad \theta_1^{(2)} = \frac{1}{v_1} \left(S_{2,3} - \theta_1^{(1)} \theta_2^{(2)} S_{1,2} \right) \quad (3.5.17)$$

et

$$v_2 = S_{3,3} - \left(\theta_2^{(2)}\right)^2 v_0 - \left(\theta_1^{(2)}\right)^2 v_1. \quad (3.5.18)$$

On peut également mettre en place l'algorithme des innovations pour faire de la prédiction à l'horizon $h \geq 2$, c'est-à-dire déterminer $\mathbb{E}L(X_{T+h} | H_T)$ au lieu de $\mathbb{E}L(X_{T+1} | H_T)$. Par la Proposition 3.9,

$$\mathbb{E}L(X_{T+h} | H_T) = \mathbb{E}L(\mathbb{E}L(X_{T+h} | H_{T+h-1}) | H_T). \quad (3.5.19)$$

De plus, en utilisant (3.5.11) avec t remplacé par $T+h-1$, on obtient

$$\mathbb{E}L(X_{T+h} | H_T) = \sum_{u=1}^{T+h-1} \theta_u^{(T+h-1)} \mathbb{E}L(X_{T+h-u} - \mathbb{E}L(X_{T+h-u} | H_{T+h-u-1}) | H_T). \quad (3.5.20)$$

Si $u \leq h-1$, $X_{T+h-u} - \mathbb{E}L(X_{T+h-u} | H_{T+h-u-1})$ est orthogonale à $H_{T+h-u-1}$ donc à H_T et si $u \geq h$, alors $X_{T+h-u} - \mathbb{E}L(X_{T+h-u} | H_{T+h-u-1})$ est un élément de H_T , d'où

$$\mathbb{E}L(X_{T+h} | H_T) = \sum_{u=h}^{T+h-1} \theta_u^{(T+h-1)} (X_{T+h-u} - \mathbb{E}L(X_{T+h-u} | H_{T+h-u-1})). \quad (3.5.21)$$

3.6 Estimation de la moyenne et de la fonction d'auto-corrélation

Étant données des réalisations x_1, \dots, x_T de variables aléatoires X_1, \dots, X_T , on considère l'estimateur empirique de la moyenne

$$\hat{\mu}_T = \frac{1}{T} \sum_{t=1}^T X_t. \quad (3.6.1)$$

Proposition 3.49. *Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire au second ordre. Alors $\hat{\mu}_T$ est un estimateur sans biais de μ et*

$$\text{Var}(\hat{\mu}_T) = \frac{1}{T} \sum_{h=-(T-1)}^{T-1} \left(1 - \frac{|h|}{T}\right) \gamma_X(h). \quad (3.6.2)$$

Proposition 3.50. *Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire au second ordre tel que $(\gamma_X(h))_{h \in \mathbb{Z}}$ soit absolument sommable. Alors*

$$\lim_{T \rightarrow \infty} T \mathbb{E}[(\hat{\mu}_T - \mu)^2] = \sum_{h \in \mathbb{Z}} \gamma_X(h). \quad (3.6.3)$$

Ceci implique qu'un processus stationnaire au second ordre pour lequel $(\gamma_X(h))_{h \in \mathbb{Z}}$ est absolument sommable est ergodique en moyenne quadratique.

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire au second ordre. Alors pour chaque $i \in \mathbb{Z}$, la variable aléatoire $(X_i - \mu)(X_{i+h} - \mu)$ a pour espérance $\gamma_X(h)$. En se basant sur cette remarque, on considère l'estimateur de $\gamma(h)$ pour h tel que $|h| < T$, défini par

$$\hat{\gamma}_T(h) = \frac{1}{T} \sum_{t=1}^{T-|h|} (X_t - \hat{\mu}_T)(X_{t+|h|} - \hat{\mu}_T) \quad (3.6.4)$$

(on notera que la normalisation est T et non $T - |h|$). Pour la fonction de corrélation, on prend l'estimateur

$$\widehat{\rho}_T(h) = \frac{\widehat{\gamma}_T(h)}{\widehat{\gamma}_T(0)}. \quad (3.6.5)$$

On définit pour $p \in \llbracket 1, T \rrbracket$ l'estimateur $\widehat{\Gamma}_p$ de la matrice de corrélation du vecteur $\mathbb{X} = (X_1, \dots, X_p)^\top$, défini par

$$\left[\widehat{\Gamma}_p \right]_{i,j} = \widehat{\rho}_T(|i - j|) = \frac{\widehat{\gamma}_T(|i - j|)}{\widehat{\gamma}_T(0)}. \quad (3.6.6)$$

Proposition 3.51. *Si $\widehat{\gamma}_T(0) \neq 0$, alors la matrice $\widehat{\Gamma}_p$ est définie positive.*

Pour $h \in \llbracket 1, T - 1 \rrbracket$, on estime $\tau(h)$ par $\widehat{\tau}_T(h)$, la h^e composante du vecteur

$$\widehat{\Gamma}_h^{-1} (\widehat{\rho}_T(1), \dots, \widehat{\rho}_T(h))^\top.$$

3.7 Fonction génératrice de l'auto-covariance

Définition 3.52. *Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire au second ordre, de fonction d'auto-covariance $\gamma_X(\cdot)$. On suppose qu'il existe δ tel que pour tout $z \in C_\delta = \{z \in \mathbb{C}, 1 - \delta < |z| < 1 + \delta\}$,*

$$\sum_{h \in \mathbb{Z}} |\gamma_X(h)| |z|^h < \infty. \quad (3.7.1)$$

La fonction génératrice de l'auto-covariance est donnée par

$$G_X(z) = \sum_{h \in \mathbb{Z}} \gamma_X(h) z^h. \quad (3.7.2)$$

Si $(X_t)_{t \in \mathbb{Z}}$ est un bruit blanc faible, alors $G_X(z) = \sigma^2$ pour tout $z \in \mathbb{C}$.

Proposition 3.53. *Soit $(\eta_t)_{t \in \mathbb{Z}}$ un bruit blanc faible de variance σ^2 et soit $(a_i)_{i \in \mathbb{Z}}$ une famille de réels telle qu'il existe $\delta > 0$ pour lequel la série $\sum_{i \in \mathbb{Z}} |a_i| |z|^i$ converge pour tout $z \in C_\delta$. On pose $\Phi(z) = \sum_{i \in \mathbb{Z}} a_i z^i, z \in C_\delta$. Soit $(X_t)_{t \in \mathbb{Z}}$ le processus défini par $X_t = \Phi(B) \eta_t$.*

Alors la fonction génératrice de l'auto-covariance de $(X_t)_{t \in \mathbb{Z}}$ est donnée par $G_X(z) = \sigma^2 \Phi(z) \Phi(1/z)$, où z est tel que $z, 1/z \in C_\delta$.

4 Les processus moyennes mobiles

4.1 Définition et principales caractéristiques

Définition 4.1. *Soit $q \in \mathbb{N}^*$. On appelle moyenne mobile d'ordre q (noté MA(q)) tout processus admettant la représentation*

$$X_t = Q(B) \eta_t = \eta_t - \sum_{k=1}^q \theta_k \eta_{t-k}, \quad (4.1.1)$$

où $Q(\cdot)$ est un polynôme à coefficients réels d'ordre q défini pour $z \in \mathbb{C}$ par $Q(z) = 1 - \sum_{k=1}^q \theta_k z^k$, $\theta_q \neq 0$ et $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc faible de variance σ^2 .

Proposition 4.2. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus MA(q) de polynôme $Q(\cdot)$. On suppose que toutes les racines de $Q(\cdot)$ sont de module strictement supérieur à 1. Alors le bruit blanc d'innovation $\varepsilon_t = X_t - \mathbb{E}L(X_t | \mathcal{H}_{t-1}^X)$ est égal à η_t .

Proposition 4.3. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus MA(q) de polynôme $Q(\cdot)$. On suppose que $Q(\cdot)$ n'admet aucune racine de module 1. Alors il existe un polynôme $\tilde{Q}(\cdot)$ dont toutes les racines sont de module strictement supérieur à 1 et tel que $X_t = \tilde{Q}(B)\varepsilon_t$, où $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc faible.

Définition 4.4. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus MA(q) défini par $X_t = Q(B)\eta_t$, où $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc. Si toutes les racines de Q sont de module différent de 1, le polynôme $\tilde{Q}(\cdot)$ comme dans la Proposition 4.3 est appelé polynôme canonique.

Proposition 4.5. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus MA(q) défini par $X_t = Q(B)\eta_t$, où $Q(z) = 1 - \sum_{\ell=1}^q \theta_\ell z^\ell$ et $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc faible. Alors la covariance de $(X_t)_{t \in \mathbb{Z}}$ est donnée par

$$\gamma(h) = \mathbb{E}[\eta_0^2] \sum_{i=0}^{q-h} \theta_i \theta_{i+h} \mathbf{1}_{\{0 \leq h \leq q\}}, \quad (4.1.2)$$

où $\theta_0 = -1$.

Remarquons que si $h > q$, alors $\gamma(h) = 0$.

Proposition 4.6. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire au second ordre. On suppose qu'il existe $q \geq 1$ tel que si $h > q$, alors $\gamma_X(h) = 0$. Alors $(X_t)_{t \in \mathbb{Z}}$ est un processus MA(q).

Corollaire 4.7. Soient $(X_t)_{t \in \mathbb{Z}}$ et $(Y_t)_{t \in \mathbb{Z}}$ des processus stationnaires au second ordre et centrés. On suppose que $(X_t)_{t \in \mathbb{Z}}$ est un processus MA(q_1), que $(Y_t)_{t \in \mathbb{Z}}$ est un processus MA(q_2) et que pour tous $s, t \in \mathbb{Z}$, $\mathbb{E}[X_s Y_t] = 0$. Alors le processus $(Z_t)_{t \in \mathbb{Z}}$ défini par $Z_t = X_t + Y_t$ est un processus MA($\max\{q_1, q_2\}$).

4.2 Estimation d'un modèle moyenne mobile

Dans la pratique, on ne connaît pas la fonction d'auto-covariance ni celle d'auto-corrélation. On l'estime par

$$\hat{\rho}_T(h) = \frac{\sum_{t=1}^{T-|h|} (X_t - \hat{\mu}_T)(X_{t+|h|} - \hat{\mu}_T)}{\sum_{t=1}^T (X_t - \hat{\mu}_T)^2}, \quad (4.2.1)$$

où

$$\hat{\mu}_T = \frac{1}{T} \sum_{t=1}^T X_t. \quad (4.2.2)$$

Proposition 4.8. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus MA(q), c'est-à-dire que $X_t = \eta_t - \sum_{\ell=1}^q \theta_\ell \eta_{t-\ell}$, où $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc fort. Pour tout $h > q$, la suite $\left(\sqrt{T} \widehat{\rho}_T(h)\right)_{T \geq |h|}$ converge en loi vers une loi normale centrée de variance

$$\sigma^2(q) = \sum_{k=1}^{\infty} (\rho(k-h))^2 = \sum_{k=h-q}^{h+q} (\rho(k-h))^2 = 1 + 2\rho(1)^2 + \dots + 2\rho(q)^2. \quad (4.2.3)$$

Ceci permet de construire les intervalles de confiance suivants : pour tout $h > q$, $\mathbb{P}(\widehat{\rho}_T(h) \in I_{q,\alpha}) = 1 - \alpha$, avec

$$I_{q,\alpha} = \left[-\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \left(\frac{1 + 2 \sum_{k=1}^q \rho^2(k)}{T}\right)^{1/2}, \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \left(\frac{1 + 2 \sum_{k=1}^q \rho^2(k)}{T}\right)^{1/2} \right], \quad (4.2.4)$$

où Φ désigne la fonction de répartition d'une loi normale centrée réduite et Φ^{-1} sa réciproque.

En pratique, on cherche à savoir si des observations X_1, \dots, X_T sont issues d'un processus MA(q). On se fixe une erreur (typiquement, $\alpha = 0.05$). On calcule l'intervalle de confiance estimé

$$\widehat{I}_{1,\alpha} = \left[-\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \left(\frac{1 + 2\widehat{\rho}_T^2(1)}{T}\right)^{1/2}, \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \left(\frac{1 + 2\widehat{\rho}_T^2(1)}{T}\right)^{1/2} \right]. \quad (4.2.5)$$

Si pour tout $h > 1$, on trouve que $\widehat{\rho}_T(h) \in \widehat{I}_{1,\alpha}$, on accepte que les observations provienne d'un processus MA(1). Sinon, on recommence avec $q = 2$. En posant

$$\widehat{I}_{q,\alpha} = \left[-\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \left(\frac{1 + 2 \sum_{k=1}^q \widehat{\rho}_T^2(k)}{T}\right)^{1/2}, \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \left(\frac{1 + 2 \sum_{k=1}^q \widehat{\rho}_T^2(k)}{T}\right)^{1/2} \right],$$

l'ordre du processus est donné par

$$\widehat{q} = \inf \left\{ q \geq 1, \widehat{\rho}_T(h) \in \widehat{I}_{q,\alpha} \text{ pour tout } h > q \right\}. \quad (4.2.6)$$

Si on trouve des estimations en dehors des intervalles de confiance (en pratique pour $h > 20$), on rejette l'hypothèse d'un modèle MA.

Pour estimer les coefficients, on fait appel à l'algorithme des innovations.

5 Les processus autorégressifs

5.1 Définition et principales caractéristiques

Définition 5.1. Soit $p \in \mathbb{N}^*$. On dit que $(X_t)_{t \in \mathbb{Z}}$ est un processus auto-régressif d'ordre p (noté AR(p)) si $(X_t)_{t \in \mathbb{Z}}$ est stationnaire au second ordre et est solution d'une équation de récurrence de la forme

$$P(B) X_t = X_t - \sum_{k=1}^p \Phi_k X_{t-k} = \eta_t, \quad (5.1.1)$$

où $P(\cdot)$ est un polynôme à coefficients réels et de degré p et $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc faible de variance σ^2 .

On notera qu'un processus vérifiant l'équation de récurrence $P(B)X_t = \eta_t$ n'est pas nécessairement stationnaire.

Par la Proposition 3.38 :

- Si $P(\cdot)$ n'a aucune racine sur le cercle unité, alors $(X_t)_{t \in \mathbb{Z}}$ est inversible et s'écrit sous la forme

$$X_t = \sum_{i \in \mathbb{Z}} a_i \eta_{t-i}, \quad (5.1.2)$$

où $(a_i)_{i \in \mathbb{Z}}$ est absolument sommable.

- Si $P(\cdot)$ n'a aucune racine sur le disque unité fermé, alors $(X_t)_{t \in \mathbb{Z}}$ est causal et s'écrit sous la forme

$$X_t = \sum_{i=0}^{\infty} a_i \eta_{t-i}, \quad (5.1.3)$$

où $(a_i)_{i \geq 0}$ est absolument sommable.

Remarquons que si $(X_t)_{t \in \mathbb{Z}}$ est inversible, alors ce processus est nécessairement stationnaire au second ordre et centré.

Proposition 5.2. *Soit $(X_t)_{t \in \mathbb{Z}}$ un processus AR(p) défini par l'équation $P(B)X_t = \eta_t$. Si $(X_t)_{t \in \mathbb{Z}}$ est causal, alors son bruit blanc d'innovation est le processus $(\eta_t)_{t \in \mathbb{Z}}$.*

Comme pour les processus MA, on peut écrire un processus AR en fonction de son bruit blanc d'innovation.

Proposition 5.3. *Soit $(X_t)_{t \in \mathbb{Z}}$ un processus AR(p) défini par $P(B)X_t = \eta_t$. On suppose que $P(\cdot)$ n'a aucune racine sur le cercle unité. Alors il existe un polynôme $\tilde{P}(\cdot)$ n'ayant aucune racine dans le disque unité fermé et tel que $\tilde{P}(B)X_t = \varepsilon_t$, où $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc.*

Définition 5.4. *L'écriture $\tilde{P}(B)X_t = \varepsilon_t$, où $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc et les racines de \tilde{P} sont toutes de module strictement supérieur à 1 est appelée forme canonique de $(X_t)_{t \in \mathbb{Z}}$.*

Proposition 5.5. *Soit $(X_t)_{t \in \mathbb{Z}}$ un processus AR(p) défini par $P(B)X_t = \eta_t$, où $P(\cdot)$ a toutes ses racines en dehors du disque unité.*

1. La fonction d'auto-covariance est donnée pour tout $h \in \mathbb{N}^*$ par

$$\gamma_X(h) = \mathbb{E}[\eta_0^2] \sum_{i=0}^{\infty} a_i a_{i+h}. \quad (5.1.4)$$

Pour $h \in \mathbb{N}$, on a

$$\gamma_X(h) = \sum_{i=1}^p \Phi_i \gamma(i-h). \quad (5.1.5)$$

2. La fonction d'auto-corrélation partielle $\tau(\cdot)$ est telle que

$$\tau(p) = \Phi_p \text{ et } \tau(h) = 0 \text{ pour } h > p. \quad (5.1.6)$$

Proposition 5.6 (Équations de Yule-Walker). Soit $(X_t)_{t \in \mathbb{Z}}$ un processus AR(p) défini par $P(B)X_t = \eta_t$, où $P(\cdot)$ n'a aucune racine dans le disque unité. On a

$$\mathbb{E}[\eta_0^2] = \gamma(0) - \sum_{i=1}^p \Phi_i \gamma(i), \quad (5.1.7)$$

et pour $h \geq p$, $(\rho(1), \dots, \rho(h))^\top = \Gamma_h(\Phi_1, \dots, \Phi_h)^\top$, où $\Phi_i = 0$ si $i > p$ et Γ_h est la matrice de corrélation du vecteur $(X_1, \dots, X_h)^\top$.

5.2 Estimation d'un modèle autorégressif

On cherche à estimer les coefficients d'un processus auto-régressif. Tout d'abord, il faut estimer l'ordre. Le point clé est que dans l'équation

$$\mathbb{E}L(X_t | X_{t-1}, \dots, X_{t-h}) = \sum_{j=1}^h \Phi_j X_{t-j}, \quad (5.2.1)$$

si $(X_t)_{t \in \mathbb{Z}}$ est un processus AR(p), alors $\Phi_j = 0$ pour $j > p$. Or on sait que les coefficients Φ_j peuvent s'exprimer à l'aide de la fonction d'auto-corrélation partielle.

On pose

$$\hat{\rho}_T(h) = \sum_{t=1}^{T-|h|} (X_t - \hat{\mu}_T)(X_{t+|h|} - \hat{\mu}_T) / \sum_{t=1}^T (X_t - \hat{\mu}_T)^2, \quad (5.2.2)$$

où

$$\hat{\mu}_T = \frac{1}{T} \sum_{t=1}^T X_t. \quad (5.2.3)$$

Un estimateur de $\Phi^{(h)} = (\Phi_1, \dots, \Phi_h)^\top$ est donné par $\hat{\Phi}_T^{(h)} = (\hat{\Phi}_1, \dots, \hat{\Phi}_h)^\top = \hat{\Gamma}_h^{-1}(\hat{\rho}_T(1), \dots, \hat{\rho}_T(h))^\top$.

Proposition 5.7. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus AR(p) causal défini par $P(B)X_t = \eta_t$, où $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc fort. Alors pour tout $h > p$, la suite $(\sqrt{T}\hat{\Phi}_T^{(h)})_{T \geq 1}$ converge vers une loi normale centrée réduite.

On pose alors

$$\hat{p} = \inf \left\{ r \geq 1, \forall h \geq r, \left| \hat{\Phi}_h \right| \leq \Phi^{-1}(1 - \alpha/2) / \sqrt{T} \right\}, \quad (5.2.4)$$

où $\Phi^{-1}(u)$ désigne le quantile d'une loi normale centrée réduite. On prend en pratique $\alpha = 0.05$.

Pour estimer les paramètres, on utilise les équations de Yule-Walker.

5.3 Prédiction d'un processus AR(p)

Proposition 5.8. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus AR(p) sous sa forme canonique. La meilleure prédiction de X_{T+h} à partir de X_1, \dots, X_T est donnée par

$$\mathbb{E}L(X_{T+h} | X_1, \dots, X_T) = \sum_{i=1}^p \Phi_i \mathbb{E}L(X_{T+h-i} | X_1, \dots, X_T). \quad (5.3.1)$$

On peut estimer à l'horizon h de manière récursive. On estime $\mathbb{E}L(X_{T+1} | X_1, \dots, X_T)$ par $\widehat{X}_{T+1} = \sum_{i=1}^p \widehat{\Phi}_i X_{T+1-i}$. On se base sur cette estimation de X_{T+1} pour en fournir une de $\mathbb{E}L(X_{T+2} | X_1, \dots, X_T)$:

$$\widehat{X}_{T+2} = \widehat{\Phi}_1 \widehat{X}_{T+1} + \sum_{i=2}^p \widehat{\Phi}_i X_{T+1-i}. \quad (5.3.2)$$

On peut également estimer les résidus. Pour un AR(p) causal :

$$\eta_t = X_t - \mathbb{E}L(X_t | \mathcal{H}_{t-1}^X) = X_t - \sum_{i=1}^p \Phi_i X_{t-i} \quad (5.3.3)$$

et les résidus sont estimés par

$$\widehat{\eta}_t = X_t - \sum_{i=1}^p \widehat{\Phi}_i X_{t-i}, \quad t \in \llbracket p+1, T \rrbracket. \quad (5.3.4)$$

En cas d'adéquation au modèle AR(p), les observations $\{\widehat{\eta}_t, t \in \llbracket p+1, T \rrbracket\}$ doivent être issues d'un bruit blanc faible.

6 Processus ARMA

6.1 Définition

Définition 6.1. Soient $(p, q) \in \mathbb{N}$. Un processus autorégressif moyenne mobile d'ordres (p, q) (noté ARMA(p, q)) est un processus $(X_t)_{t \in \mathbb{Z}}$ stationnaire au second ordre, centré, solution d'une équation de récurrence de la forme

$$P(B)X_t = Q(B)\eta_t, \quad t \in \mathbb{Z}, \quad (6.1.1)$$

où $P: z \mapsto 1 - \sum_{i=1}^p \Phi_i z^i$, $\Phi_p \neq 0$, $Q: z \mapsto 1 - \sum_{k=1}^q \theta_k z^k$, $\theta_q \neq 0$ et $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc faible de variance σ^2 .

On cherche dans un premier temps à simplifier l'équation de récurrence (6.1.1). On suppose que les polynômes $P(\cdot)$ et $Q(\cdot)$ ont une racine commune μ de module différent de 1. On peut alors exprimer $P(\cdot)$ sous la forme $P(z) = (1 - z/\mu)P_1(z)$ et $Q(z) = (1 - z/\mu)Q_1(z)$, où $P_1(0) = Q_1(0) = 1$. Ainsi, $(X_t)_{t \in \mathbb{Z}}$ vérifie l'équation de récurrence

$$\left(I - \frac{1}{\mu}B\right)P_1(B)X_t = \left(I - \frac{1}{\mu}B\right)Q_1(B)\eta_t \quad (6.1.2)$$

Comme l'opérateur $I - \mu^{-1}B: V \rightarrow V$ est inversible, $(X_t)_{t \in \mathbb{Z}}$ est solution de l'équation de récurrence

$$P_1(B)X_t = Q_1(B)\eta_t. \quad (6.1.3)$$

On peut ainsi « éliminer » les racines communes de $P(\cdot)$ et $Q(\cdot)$, ce qui nous mène à la définition suivante.

Définition 6.2. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus ARMA(p, q) solution de l'équation de récurrence (6.1.1). On dit que le processus $(X_t)_{t \in \mathbb{Z}}$ est donné dans sa représentation minimale si les racines communes de $P(\cdot)$ et $Q(\cdot)$ sont toutes de module différent de 1.

6.2 Représentation d'un processus ARMA (p, q) en MA (∞)

Proposition 6.3. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus ARMA (p, q) défini par (6.1.1).

1. Si $P(\cdot)$ n'a aucune racine sur le cercle unité de \mathbb{C} , alors le processus $(X_t)_{t \in \mathbb{Z}}$ est inversible et s'écrit sous la forme

$$X_t = \sum_{i \in \mathbb{Z}} \Phi_i^* \eta_{t-i}, \quad t \in \mathbb{Z}, \quad (6.2.1)$$

où $(\Phi_i^*)_{i \in \mathbb{Z}}$ est absolument sommable.

2. Si $P(\cdot)$ n'a aucune racine sur le disque unité fermé de \mathbb{C} , alors le processus $(X_t)_{t \in \mathbb{Z}}$ est causal et s'écrit sous la forme

$$X_t = \sum_{i=0}^{\infty} \Phi_i^* \eta_{t-i}, \quad t \in \mathbb{Z}, \quad (6.2.2)$$

où $(\Phi_i^*)_{i=0}^{\infty}$ est absolument sommable. De plus,

$$\phi_0 = 1 \text{ et } \Phi_i^* = -\theta_i + \sum_{k=1}^i \Phi_k \Phi_{i-k}^*. \quad (6.2.3)$$

6.3 Représentation d'un ARMA (p, q) en AR (∞)

Proposition 6.4. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus ARMA (p, q) défini par (6.1.1).

1. Si $Q(\cdot)$ n'a aucune racine sur le cercle unité, alors $(\eta_t)_{t \in \mathbb{Z}}$ est inversible et s'écrit sous la forme

$$\eta_t = \sum_{i \in \mathbb{Z}} \theta_i^* X_{t-i}, \quad t \in \mathbb{Z}, \quad (6.3.1)$$

où $(\theta_i^*)_{i \in \mathbb{Z}}$ est absolument sommable.

2. Si $Q(\cdot)$ n'a aucune racine sur le cercle unité, alors $(\eta_t)_{t \in \mathbb{Z}}$ est causal et s'écrit sous la forme

$$\eta_t = \sum_{i=0}^{\infty} \theta_i^* X_{t-i}, \quad t \in \mathbb{Z}, \quad (6.3.2)$$

où $(\theta_i^*)_{i=0}^{\infty}$ est absolument sommable. De plus,

$$\theta_0^* = 1 \text{ et } \theta_i^* = -\Phi_i + \sum_{k=1}^i \theta_k \theta_{i-k}^*. \quad (6.3.3)$$

La démonstration serait un copié-collé de celle de la Proposition 6.3 en échangeant les rôles de $P(\cdot)$ et $Q(\cdot)$ ainsi que ceux de $(\eta_t)_{t \in \mathbb{Z}}$ et $(X_t)_{t \in \mathbb{Z}}$.

6.4 Innovation d'un processus ARMA

Comme ce fut le cas pour les processus MA et AR, le processus $(\eta_t)_{t \in \mathbb{Z}}$ n'est pas nécessairement le bruit blanc d'innovation du processus $(X_t)_{t \in \mathbb{Z}}$.

Proposition 6.5. *Soit $(X_t)_{t \in \mathbb{Z}}$ un processus ARMA (p, q) défini par (6.1.1). Si toutes les racines de $P(\cdot)$ ainsi que toutes celles de $Q(\cdot)$ se trouvent à l'extérieur du disque unité fermé de \mathbb{C} , alors $(X_t)_{t \in \mathbb{Z}}$ est un processus ARMA (p, q) de bruit blanc d'innovation $(\eta_t)_{t \in \mathbb{Z}}$.*