INSA, 4gmm,
Année 2016/2017.

Introduction à l'Étude des Séries Temporelles

Jean-Yves Dauxois

Ce polycopié est construit en vue d'une Progression en Groupe (PEG). Il permet une formation en auto-apprentissage. Les preuves et les exemples sont présentés sous forme d'exercices. Il est plus aisé et conseillé de travailler en groupe de 3-4 étudiants pour favoriser également un apprentissage par les pairs.

Cette version du polycopié du cours de Séries Temporelles est encore récente et toujours "en construction". Elle est nécessairement non complete et surtout loin d'être sans coquilles ni sans possibilités d'amélioration. Les lecteurs attentifs et attentionnés auront la gentillesse de me faire remonter des remarques constructives, tant sur le fond que sur la forme. Merci d'avance !

Table des matières

CHAPITRE 1

Introduction et Analyse descriptive

1. Introduction

L'objectif de ce cours est de modéliser l'évolution dans le temps, ici supposé discret, d'un phénomène aléatoire. On parle de série temporelle.

1.1. Exemples de séries temporelles.

Les séries temporelles sont présentes dans de nombreux domaines d'application. Par exemple, en démographie, on peut tracer l'évolution de la taille d'une population comme celle des USA entre 1790 et 1990, représentée dans la Figure 1.

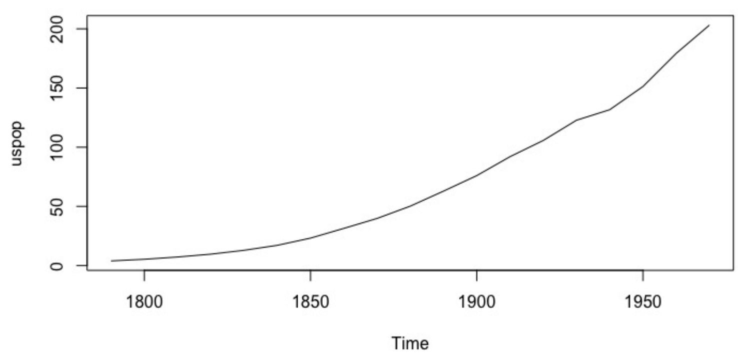

FIGURE 1. Taille de la population (en million d'habitants) des USA entre les années 1790 et 1990

En écologie, une série temporelle souvent citée en exemple est celle du nombre de lynx capturés au Canada de 1821 à 1934 et dont la représentation est donnée par la Figure 2.

En économie également, les séries temporelles sont très utilisées. Un exemple est donné par la série (*AirPassengers*) du nombre de passagers dans les lignes aériennes entre les années 1949 et 1960 et est représentée dans la Figure 3. On peut transformer cette dernière série en considérant le logarithme népérien du nombre de passagers aériens. On obtient la série représentée dans la Figure 4.

Les exemples ne manquent pas. En écologie, on pourrait aussi s'intéresser à l'évolution de la température (moyenne) en un lieu donné sur un siècle, à la concentration en Ozone dans l'Atmosphère d'une grande ville. En finance, le cours d'une action est évidemment suivi avec intérêt, tout autant que le tracé de l'électrocardiogramme d'un

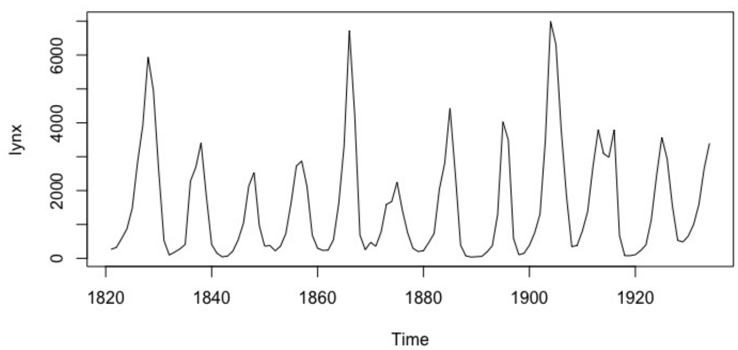

FIGURE 2. Nombre de lynx capturés au Canada de 1821 à 1934

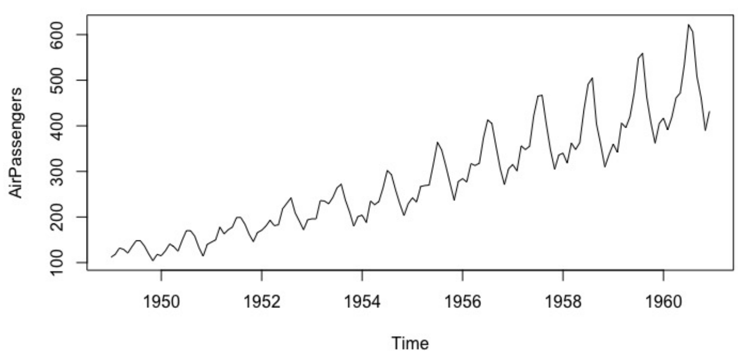

FIGURE 3. Nombre mensuel de passagers (en milliers) entre les années 1949 et 1960 dans les lignes aériennes.

patient par un médecin... L'intérêt peut aussi se porter sur la variation du nombre de connexions à internet via un serveur, comme représenté dans la Figure 5.

Dans la suite, on va supposer que la série observée est une suite de variables aléatoires (v.a.). Comme il s'agit d'évaluer l'évolution d'un phénomène au cours du temps, l'ordre est important. On note Y_t la valeur observée au temps t, le processus $(Y_t)_{t\in T}$ est alors appelée série temporelle. Souvent l'espace des temps T est $\mathbb{N}$, voire $\mathbb{Z}$. C'est en raison du caractère discret du temps que l'on parle de série temporelle.

1.2. Quelles questions sont posées au statisticien qui veut étudier une série temporelle ?

Le statisticien est d'abord confronté au problème de la **modélisation** de le série temporelle. Il souhaite trouver, ajuster, un modèle qui décrive ou représente au mieux les données. Se poseront alors nécessairement les questions d'estimation des paramètres, de tests d'adéquation à un modèle choisi comme souvent en Statistique. Mais attention,

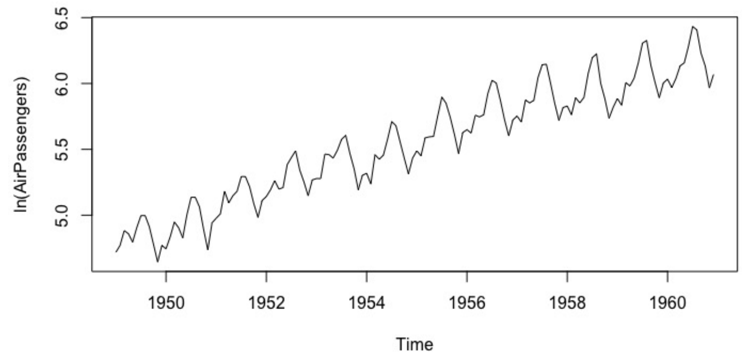

FIGURE 4. Logarithme du nombre mensuel de passagers (en milliers) entre les années 1949 et 1960 dans les lignes aériennes.

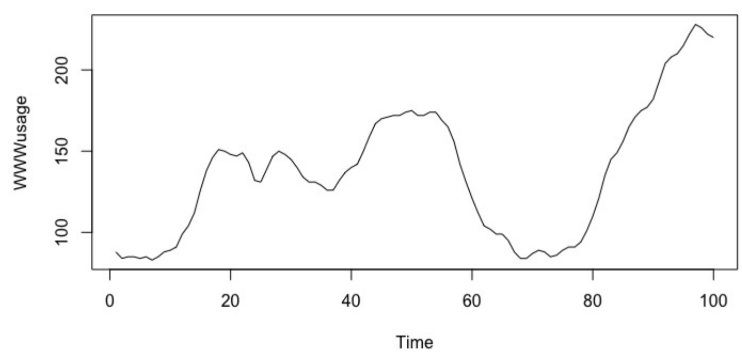

FIGURE 5. Nombre d'usagers connectés à internet via un serveur toutes les minutes.

un ajustement exact aux données n'est généralement pas recherché. Le statisticien craint souvent le sur-ajustement ! L'objectif est plutôt d'extraire la structure du signal et enlever la partie bruitage, les erreurs de mesure ou encore les fluctuations aléatoires non explicables. En particulier, le statisticien s'attachera à déceler la présence d'une tendance, d'une saisonnalité dans les données comme nous le détaillerons plus bas. Si l'on reprend la série temporelle du nombre de passagers aériens (*Airpassengers*), le statisticien peut estimer une saisonnalité et en déduire ce que l'on appelle la série corrigée des variations saisonnières, comme le montre la Figure 6.

Mais souvent l'intérêt n'est pas uniquement d'expliquer ou modéliser la structure de la série temporelle. Souvent on sollicitera le statisticien pour une mission de **prévision** les valeurs futures de la série. On en comprend l'intérêt en finance pour le cours d'un action, en santé publique pour la concentration d'Ozone dans l'atmosphère par exemple. Pour bien des raisons, la prévision de la température n'est pas inutile non plus...

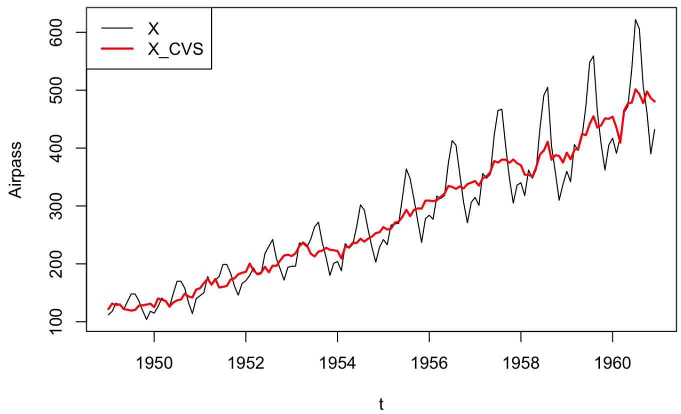

FIGURE 6. Série AirPassengers corrigée des variations saisonnières.

Ainsi, sur la base des observations $Y_1, \ldots, Y_T$ de la série aux T premiers instants, le statisticien aura à prévoir la valeur suivante Y_{T+1} ou plus éloignée dans le temps Y_{T+h}. Un exemple de prévision des valeurs futures de la série Airpassengers est donné dans la Figure 7. Le statisticien sera très vite confronté aux questions de la qualité des prédictions, matérialisée par l'erreur de prédiction, à celle de la taille des intervalles de prédiction, etc...

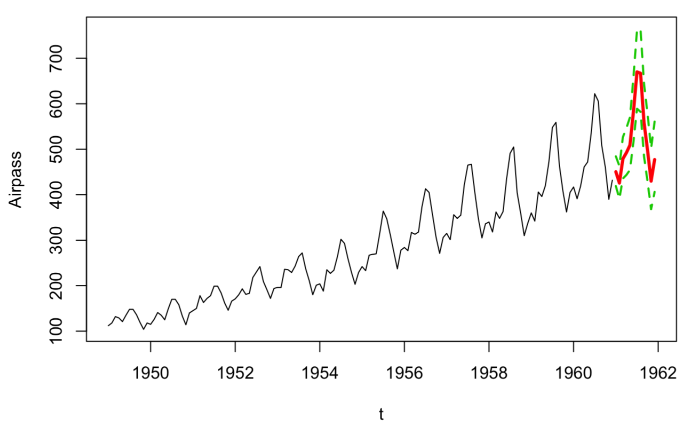

FIGURE 7. Prévision sur l'année suivante pour la série AirPassengers ainsi que son intervalle de confiance.

2. Décomposition d'une série temporelle

2.1. Quelques modéles de séries temporelles avec tendance et/ou saisonnalité.

EXERCICE 1.1. *Reprendre les séries temporelles illustrées dans les figures 1 à 5 et les comparer. Quelle structure pourrait-on déceler sur chacune d'entre elles ? On pourra considérer des notions de tendance, de saisonnalité, en réfléchissant aux définitions que l'on pourrait donner à ces termes. On pourra aussi les comparer à la série d'un bruit blanc gaussien donnée dans la Figure 8.*

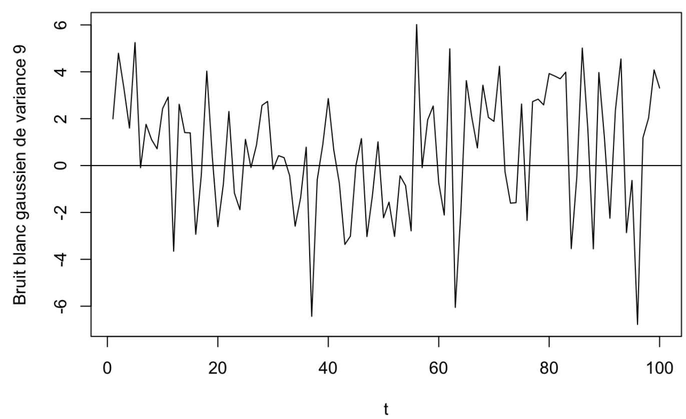

FIGURE 8. Représentation d'un bruit blanc gaussien.

2.2. Tendance et saisonnalité d'une série temporelle.

Comme on vient de le constater sur quelques exemples, on peut souvent décomposer une série temporelle en trois termes sous la forme :

$$Y_t = m_t + s_t + X_t,$$

où

- t est l'indice du temps, à valeurs dans T sous ensemble de $\mathbb{N}$ ou $\mathbb{Z}$,
- m_t est une fonction déterministe à variation que l'on espère lente (appelée **tendance**), qui capte les variations de niveau et que l'on espère assez lisse (variations à long terme),
- s_t est une fonction déterministe périodique (appelée **saisonnalité**) de période r telle que
$$\sum_{i=1}^{r} s_{t+i} = 0, \forall t \in T$$
- X_t est un **bruit** aléatoire **stationnaire** (terme restant à définir !). On l'appelle parfois résidu.

Remarquons que si la saisonnalité et les variations semblent croître, il est parfois possible d'atténuer ce phénomène en tentant une transformation des données. C'est en particulier ce que nous avons réalisé et constaté sur les séries illustrées par les Figures 3 et 4 .

L'hypothèse de somme nulle de la saisonnalité sur la longueur de la période n'est pas contraignante, puisque cette somme (fixe !) peut sinon être ajoutée à la tendance.

Dans la suite, on suppose que cette modélisation est appropriée (quitte à avoir fait une transformation des données au préalable). L'objectif de ce cours est d'apprendre à modéliser et estimer les composantes m_t, s_t et X_t puis de pouvoir faire des prévisions sur les valeurs futures de la série temporelle initiale Y_t.

2.3. Une approche générale de la modélisation des séries temporelles.

La méthode générale pour étudier une série temporelle est la suivante :

1) Dans un premier temps on trace la série des données et on repère ses principales caractéristiques. On regarde en particulier si on décèle

- une tendance
- une composante saisonnière
- une ou des ruptures dans le comportement de la série
- une ou des observations aberrantes.

2) Dans un deuxième temps, on s'attache à estimer ou enlever la tendance et la composante saisonnière (m_t et s_t) pour obtenir une série X_t de résidus stationnaires. Pour cela on peut utiliser plusieurs techniques : transformation des données, estimer les tendances et composantes saisonnières puis les enlever des données, différencier la série (considérer par exemple la série $Y_t - Y_{t-d}$).

3) Choisir un modèle de série stationnaire pour les résidus

4) Prévoir les valeurs futures de la série en prévoyant d'abord celles des résidus puis "remonter" jusqu'à la série initiale en les transformations inverses.

3. Estimation et élimination d'une tendance en absence de saisonnalité

En l'absence de saisonnalité, le modèle précédent s'écrit :

$$Y_t = m_t + X_t, \forall t \in T$$

où (X_t) est, sans perte de généralité, un processus centré. En effet, si $\mathbb{E}X_t \neq 0$, on remplace m_t et X_t par $m_t + \mathbb{E}X_t$ et $X_t - \mathbb{E}X_t$ respectivement.

On suppose que l'on observe le processus sur les instants de temps $t = 1, \ldots, \tau$.

3.1. Estimation de la tendance (en l'absence de saisonnalité).

3.1.1. *Estimation par Moindre carrés.*

Estimation.

On suppose que la tendance est une combinaison linéaire de fonctions temporelles, connues et déterministes :

$$m_t = \sum_{i=0}^{n} \alpha_i m_t^i.$$

Seuls les coefficients α_i, pour $i = 1, \ldots, n$, sont inconnus et doivent donc être estimés.

EXEMPLE 1.1. *On peut considérer des fonctions tendances m_t de la forme :*

- *Tendance linéaire* $m_t = \alpha_0 + \alpha_1 t$ *;*
- *Tendance quadratique* $m_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2$ *;*
- *Tendance polynomiale* $m_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_n t^n$ *;*
- *Tendance exponentielle* $m_t = c_0 + c_1 \alpha^t$ *;*
- *Tendance de Gompertz* $m_t = \exp(c_0 + c_1 \alpha^t)$ *;*
- *Tendance Logistique* $m_t = 1/(c_0 + c_1 \alpha^t)$ *;*
- *ou bien des mélanges de ces types de fonctions.*

L'idée est alors d'estimer les paramètres inconnus α_i, pour $i = 1, \ldots, n$ par moindres carrés, i.e.

$$(\hat{\alpha}_0, \ldots, \hat{\alpha}_n) = \arg \min_{(\alpha_0, \ldots, \alpha_n) \in \mathbb{R}^n} \sum_{t=1}^{\tau} (y_t - m_t)^2$$

Un cas relativement simple est celui où la tendance est polynomiale ou plus généralement sous forme d'une combinaison linéaire de fonctions entièrement connues, ce qui n'est pas le cas avec les tendances exponentielles, Gompertz et Logistique. On tombe en effet dans le cadre d'un modèle de régression linéaire. En effet, en introduisant

$$\mathbb{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_\tau \end{pmatrix}, \varepsilon = \begin{pmatrix} X_1 \\ \vdots \\ X_\tau \end{pmatrix}, \beta = \begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_n \end{pmatrix} \text{ et } \mathbb{X} = \begin{pmatrix} m_1^0 & \cdots & m_1^n \\ \vdots & \ddots & \vdots \\ m_\tau^0 & \cdots & m_\tau^n \end{pmatrix},$$

on peut écrire

$$\mathbb{Y} = \mathbb{X}\beta + \varepsilon.$$

D'après les résultats sur le cours de régression linéaire, on sait que l'estimateur des moindres carrés de β est donné par :

$$\hat{\beta} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{Y}.$$

Données corrigées.

Une fois que l'on a estimé les paramètres inconnus du modèle, on obtient naturellement les données corrigées de la tendance via

$$\hat{Y}_t^{\mathrm{CT}} = Y_t - \hat{m}_t,$$

où

$$\hat{m}_t = \sum_{i=0}^{n} \hat{\alpha}_i m_t^i.$$

3.1.2. *Estimation par filtrage de moyenne mobile.*

Définition 1.1. *L'opérateur* ***retard*** B *sur une série temporelle est défini par :*

$$BX_t = X_{t-1}, \ \forall t \in T.$$

On note de manière naturelle :

$$B^i X_t = X_{t-i}, \ \forall t \in T \text{ et } \forall i \in \mathbb{N}^*$$

L'opérateur ***avance*** F *sur une série temporelle est défini par :*

$$FX_t = X_{t+1}, \ \forall t \in \mathbb{Z}.$$

On note aussi

$$F^i X_t = X_{t+i},\ \forall t \in \mathbb{Z} \text{ et } \forall i \in \mathbb{N}^*$$

et $B^{-i}=F^i$*, pour tout* i*.*

DÉFINITION 1.2. *Une **moyenne mobile** est un opérateur linéaire (ou encore filtre linéaire) de forme générale*

$$M = \sum_{i=-m_1}^{m_2} \theta_i B^i,$$

où $(m_1, m_2) \in \mathbb{N}^* \times \mathbb{N}^*$ *et* $\theta_i \in \mathbb{R}$ *pour tout* i*. L'**ordre** de la moyenne mobile est l'entier* $m_1 + m_2 + 1$*.*

La moyenne mobile est dite

- ***normalisée*** *si*
$$\sum_{i=-m_1}^{m_2} \theta_i = 1;$$
- ***centrée*** *si* $m_1 = m_2$ *;*
- ***symétrique*** *si* $m_1 = m_2 = m$ *et* $\theta_i = \theta_{-i}$*, pour* $i = 1, \ldots, n$*.*

EXERCICE 1.2.

- *Donner l'expression de l'application de la moyenne mobile* M *à la série temporelle* $(X_t)_{t\in T}$*.*
- *Pourquoi dit-on que la moyenne mobile est un opérateur linéaire ?*

Ces moyennes mobiles sont parfois appelées filtres passe bas car elles enlèvent à une série $(X_t)_{t\in T}$ ses fluctuations rapides (dites encores hautes fréquences). Il ne reste plus qu'un terme de tendance à variation relativement lente. Ces moyennes mobiles peuvent aussi être vues comme une estimation non paramétrique de la tendance par moyennes locales.

EXERCICE 1.3.

Soit q *un entier non nul. Considérons la série*

$$M_{2q+1} Y_t = \frac{1}{2q+1} \sum_{j=-q}^{q} Y_{t-j}.$$

- *Préciser sa nature.*
- *Supposons qu'une série temporelle* $(Y_t)_{t\in T}$ *s'écrive sous la forme :* $Y_t = m_t + X_t$*, où* m_t *est la tendance et* X_t *un processus centré. Développer* $M_{2q+1}Y_t$ *en séparant les effets de la moyenne mobile sur la tendance et sur le processus centré.*
- *On suppose que la tendance est linéaire, i.e. de la forme* $m_t = a + bt$*. Que vaut* $M_{2q+1}m_t$ *pour les valeurs* $q = 1$ *et* $q = 2$ *?*
- *Que s'attend-on à avoir pour l'autre partie, c'est à dire pour* $M_{2q+1}X_t$*, au moins quand la valeur de* q *est assez grande ?*
- *Finalement, quel est l'intérêt de considérer une telle moyenne mobile ?*

3.1.3. *Estimation par lissage exponentiel.*

DÉFINITION 1.3. *Le* ***lissage exponentiel de paramètre*** α *consiste à estimer la tendance via la formule récursive :*

$$\begin{aligned} \hat{m}_t &= \alpha Y_t + (1-\alpha)\hat{m}_{t-1}, \text{ pour } t = 2, \ldots, \tau \\ \hat{m}_1 &= Y_1, \end{aligned}$$

où $\alpha \in [0,1]$.

EXERCICE 1.4.

- *Montrer que le lissage exponentiel est une moyenne mobile, préciser sa nature et ses coefficients.*
- *Voir ce lissage comme une moyenne pondérée des valeurs précédentes. Cela correspond-il vraiment (en tout point) avec la définition d'un moyenne mobile vue dans la Définition 1.2 ?*
- *Que dire de l'évolution des poids en fonction du passé considéré ? Que se passe-t-il quand le coefficient* α *est proche de 1 et à l'inverse, quand il est proche de 0 ?*

Le lissage exponentiel est parfois utilisé pour estimer/prévoir une valeur non observée de la série temporelle. Ainsi, supposons que l'on observe la série temporelle sur les instants $\{1, \ldots, \tau\}$. Une prévision à l'horizon h de la série est donnée par :

$$\hat{Y}_{\tau+h} = \hat{Y}_{\tau+1} = \sum_{k=0}^{\tau-2} \alpha(1-\alpha)^k Y_{\tau-k} + (1-\alpha)^{\tau-1} Y_1.$$

La formule de mise jour permet de voir qu'une observation supplémentaire de la série ne nécessite pas de recalculer entièrement la prévision.

3.2. Élimination de la tendance par différenciation.

DÉFINITION 1.4. *L'opérateur* ***différenciation (à l'ordre 1)*** ∇ *d'une série temporelle est défini comme opérateur qui associe à une série* (Y_t) *une série* $(\tilde{Y}_t)$ *déterminée par*

$$\tilde{Y}_t = \nabla Y_t = Y_t - Y_{t-1} = Y_t - BY_t = (I - B)Y_t, \ \forall t \in T.$$

Naturellement, la différenciation ***à l'ordre*** k *de la série* (Y_t) *est la série* $(\tilde{Y}_t)$ *définie par :*

$$\tilde{Y}_t = \nabla^k Y_t = (I - B)^k Y_t, \ \forall t \in T.$$

EXERCICE 1.5.

- *Considérons une série temporelle avec tendance linéaire de la forme :*

$$m_t = a + bt.$$

  Que se passe-t-il si l'on applique une différenciation à l'ordre 1 sur cette série ?
- *Si la tendance est maintenant polynomiale d'ordre* k*, que feriez-vous pour l'annuler ? Justifier votre réponse.*

Ainsi, la différenciation permet d'éliminer les tendances polynomiales et donc pratiquement toutes les tendances car elles peuvent très souvent être approchées par des polynômes. Mais insistons bien sur le fait que cette technique ne permet pas un estimation de la tendance mais seulement son élimination...

4. Estimation et élimination de la tendance et de la composante saisonnière

Reprenons une série temporelle standard où tendance et saisonnalité sont présentes :

$$Y_t = m_t + s_t + X_t, \text{ pour } t \in T$$

avec $\mathbb{E}X_t = 0$ et la saisonnalité est telle que $s_{t+r} = s_t$ et $\sum_{j=1}^r s_j = 0$.

4.1. Estimation.

4.1.1. *Estimation par Moindres carrés.*

Estimation.

On peut reprendre la méthode des moindres carrés en supposant cette fois-ci que la tendance mais aussi la composante saisonnière sont des combinaisons linéaires de fonctions connues, i.e. supposer qu'il existe des fonctions m_t^i, pour $i = 0, \ldots, n$ et s_t^j, pour $j = 1, \ldots, p$ telles que :

$$Y_t = \sum_{i=0}^n \alpha_i m_t^i + \sum_{j=1}^p \beta_j s_t^j + X_t.$$

Seuls les coefficients α_i, pour $i = 1, \ldots, n$, et β_j, pour $j = 1, \ldots, p$, sont inconnus et doivent donc être estimés. Pour modéliser la partie tendance, on peut piocher à nouveau dans les fonctions listées dans l'Exemple 1.1. Pour la partie saisonnalité, on peut considérer les cas suivants.

EXEMPLE 1.2. *Quelques exemples classiques de fonctions saisonnalité sont les suivants.*

- *On peut utiliser des indicatrices, comme par exemple dans le cas d'une saisonnalité trimestrielle avec*

$$\forall j \in \{1, \ldots, 4\} : s_t^j = \begin{cases} 1 & \textit{si le trimestre à l'instant } t \textit{ est } j \\ 0 & \textit{sinon} \end{cases}.$$

- *On peut aussi utiliser une combinaison de fonctions sinusoïdales.*
- *Les différentes fonctions de base de la saisonnalité s_t^j, pour $j = 1, \ldots, p$ peuvent avoir des périodes différentes.*

EXERCICE 1.6. *Écrire l'expression de l'estimation par Moindres Carrés Ordinaires du paramètre mutlidimensionnel $(\alpha_0, \ldots, \alpha_n, \beta_1, \ldots, \beta_p)$ du modèle en prenant soin de bien donner l'expression des matrices utilisées.*

Données ajustées.
Les données ajustées sont alors

$$\hat{Y}_t = \hat{m}_t + \hat{s}_t = \sum_{i=0}^{n} \hat{\alpha}_i m_t^i + \sum_{j=1}^{p} \hat{\beta}_j s_t^j, \forall t \in T.$$

On appelle **série corrigée des variations saisonnières** la série :

$$\hat{Y}_t^{\mathrm{CVS}} = Y_t - \hat{s}_t.$$

C'est ce que l'on a obtenu pour la série *AirPassengers* dans la Figure 7.

4.1.2. *Estimation par Moyenne Mobile.*

Méthodologie générale.
Nous avons vu l'intérêt qu'il peut y avoir à utiliser l'opérateur moyenne mobile pour estimer la tendance d'une série temporelle. Nous allons voir que, s'il est convenablement choisi, son usage nous permet également d'estimer la saisonnalité.

On suppose toujours que l'on dispose d'une série temporelle avec tendance et saisonnalité

$$Y_t = m_t + s_t + X_t, \text{ pour } t \in T,$$

et l'on souhaite identifier cette décomposition. L'idée est alors de parvenir à trouver une moyenne mobile M qui vérifie les propriétés suivantes :

- la moyenne mobile laisse invariante la tendance, i.e. $Mm_t = m_t$;
- la moyenne mobile absorbe la saisonnalité, i.e. $Ms_t = 0$;
- la moyenne mobile réduit la variance du processus observé, i.e. MY_t a une variance plus faible que Y_t. Le coefficient MY_t/Y_t est appelé **pouvoir de réduction** de la variance.

Dans ce cas MY_t donnera une estimation de la tendance comme on l'a déjà vu et il restera à travailler sur la série corrigée de la tendance, i.e. sur $\hat{Y}_t^{\mathrm{CT}} = Y_t - MY_t$ pour estimer la saisonnalité. On va distinguer les cas suivant la parité de la périodicité de la saisonnalité.

Cas d'une périodicité impaire.
Considérons dans un premier temps une série temporelle pour laquelle on sait que la saisonnalité est de période impaire, i.e. $r = 2q + 1$. Supposons que l'on ait observé la série aux instant $t = 1, \ldots, \tau$.

EXERCICE 1.7. *Considérons la moyenne mobile symétrique, normalisée et d'ordre $r = 2q + 1$:*

$$M_{2q+1} = \frac{1}{2q+1} \sum_{j=-q}^{q} B^j.$$

- *Se convaincre que l'application de cette moyenne mobile sur la série initiale (Y_t) nous donne directement une estimation de la tendance et annule la saisonnalité. On supposera pour cela que la tendance est assez lisse et que q est assez grand.*

- *On note* $\hat{m}_t$ *l'estimation de la tendance de la série temporelle* $(Y_t)_{t\in T}$ *ainsi obtenue,*

$$\hat{m}_t = M_{2q+1}Y_t = \frac{1}{2q+1}\sum_{j=-q}^{q} Y_{t-j},$$

et on considère alors la série corrigée de la tendance

$$\hat{Y}_t^{CT} = Y_t - \hat{m}_t,$$

sur laquelle on va tâcher d'estimer la saisonnalité.

Pour cela, pour $k = 1, \ldots, r$, *on calcule les* ω_k *:*

$$\omega_k = \frac{1}{Card\{j : q < k + jr \leq \tau - q\}} \sum_{\{j:q<k+jr\leq\tau-q\}} \hat{Y}_{k+jr}^{CT}.$$

Que représentent ces termes ? Sont-ils de somme nulle ?

- *Que proposeriez-vous comme estimation de la composante saisonnière basée sur ces coefficients* ω_k *?*

Cas d'une périodicité paire.

On suppose maintenant que la périodicité est paire, i.e. $r = 2q$.

EXERCICE 1.8. *On peut alors utiliser une approche similaire en considérant la moyenne mobile symétrique normalisée :*

$$M_rY_t = \frac{1}{r}\left(Y_{t-q+\frac{1}{2}} + \ldots + Y_{t-\frac{1}{2}} + Y_{t+\frac{1}{2}} + \cdots + Y_{t+q-\frac{1}{2}}\right),$$

où

$$Y_{t-\frac{1}{2}} = \frac{1}{2}\left(Y_{t-1} + Y_t\right)$$

pour estimer la tendance et éliminer la saisonnalité.

- *Montrer que l'on peut réécrire* MY_t *sous la forme :*

$$M_rY_t = \frac{1}{2r}\left(Y_{t-q} + 2Y_{t-q+1} + \cdots + 2Y_t + \cdots + 2Y_{t+q-1} + Y_{t+q}\right).$$

- *Vérifier que la saisonnalité de périodicité* $r = 2q$ *est annulée par une telle moyenne mobile.*
- *Que feriez-vous ensuite pour estimer la saisonnalité ?*

Données ajustées.

Les données ajustées sont alors, pour $t \in T$:

$$\hat{Y}_t = \hat{m}_t + \hat{s}_t$$

et les données corrigées des variations saisonnières :

$$\hat{Y}_t^{\text{CVS}} = Y_t - \hat{s}_t.$$

Souvent on procède à une ré-estimation de la tendance en considérant les données corrigées des variations saisonnières et en appliquant une des techniques vues précédemment. On obtient alors $\tilde{m}_t$. La série des bruits est alors :

$$X_t = Y_t - \tilde{m}_t - \hat{s}_t, \forall t \in T.$$

4.1.3. *Elimination de la saisonnalité par différenciation.*

Pour supprimer une saisonnalité de périodicité r, on peut utiliser une méthode de différenciation similaire à celle utilisée pour éliminer la tendance.

Exercice 1.9.

- *Quel opérateur de différenciation appliqueriez-vous à la série initiale* (Y_t) *pour obtenir éliminer la saisonnalité supposée de période* r *?*
- *Donner l'expression de la série obtenue, notée* $(\tilde{Y}_t)$, *en fonction de la tendance initiale* (m_t) *et du processus* (X_t). *Préciser pour quelles valeurs de* t, *on peut considérer cette série désaisonnalisée.*
- *Comment feriez vous pour estimer ou éliminer la tendance résiduelle ?*

5. Travaux Dirigés

Exercice TD 1.1 *(Solutions d'équations de récurrence linéaire)*

On cherche à résoudre une équation de récurrence linéaire à coefficients réels, précisément à trouver les suites $(X_t)_{t\in\mathbb{Z}}$ satisfaisant une équation de la forme :

$$\alpha_0 X_t + \alpha_1 X_{t+1} + \ldots + \alpha_n X_{t+n} = 0, \tag{1}$$

pour tout $t \in \mathbb{Z}$ et où $\alpha_0, \ldots, \alpha_n$ sont des réels. Cette équation est dite d'ordre n. On note $\mathcal{S}_n$ l'ensemble des solutions dans $\mathbb{R}$ de cette équation de récurrence linéaire.

(1) Montrer que $\mathcal{S}_n$ est un sous-espace vectoriel $\mathbb{R}^{\mathbb{Z}}$ des suites réelles d'indices dans $\mathbb{Z}$.

(2) Montrer qu'à chaque suite solution de l'équation (1) on associe de manière unique un n-uplet $x_0, x_1, \ldots, x_{n-1}$ de $\mathbb{R}^n$. On dit que les espaces $\mathcal{S}_n$ et $\mathbb{R}^n$ sont isomorphes. En déduire la dimension de l'espace $\mathcal{S}_n$.

(3) On va identifier une base de cet espace $\mathcal{S}_n$. Cherchons des solutions de la forme λ^t avec $\lambda \in \mathbb{C}$. Montrer qu'une telle valeur de λ doit nécessairement être racine du polynôme :

$$\alpha_0 + \alpha_1 z + \ldots + \alpha_n z^n. \tag{2}$$

(4) On va séparer les cas, suivant les racines trouvées. Supposons dans un premier temps que les n racines $\lambda_1, \ldots, \lambda_n$ de ce polynôme soient distinctes dans $\mathbb{C}$. On peut montrer que les suites

$$X_{it} = \lambda_i^t, i = 1, \ldots, n, t \in \mathbb{Z}$$

sont linéairement indépendantes.

(a) Montrer que toute solution de l'équation de récurrence linéaire 1 s'écrit donc :

$$X_t = \sum_{i=1}^n c_i X_{it} = \sum_{i=1}^n c_i \lambda_i^t.$$

(b) On va s'intéresser à l'allure des solutions. Supposons qu'une solution λ_j particulière soit réelle. En utilisant le logiciel de votre choix (R par exemple), tracer des exemples de séries en fonction de la valeur de λ_j. Quels types de courbes obtient-on ?

(c) *Exemple d'application.* Résoudre dans $\mathbb{R}$ l'équation de récurrence linéaire :

$$2X_{t+2} + 3X_{t+1} - 5X_t = 0.$$

(d) On suppose maintenant que la solution λ_j est complexe, écrite sous forme exponentielle :

$$\lambda_j = \rho_j e^{i\omega_j}.$$

Justifier que le complexe conjugué $\bar{\lambda}_j$ est également racine du polynôme donné en (2) et correspond donc à une autre racine $\lambda_{j'}$.

(e) On rappelle que l'on cherche des solutions dans $\mathbb{R}$ à l'équation de récurrence linéaire (1). Montrer qu'alors toutes combinaison linéaire des solutions X_{jt} et $X_{j't}$ associées aux racines conjuguées précédentes λ_j et $\bar{\lambda}_j$, à condition qu'elle soit réelle, peut s'écrire sous la forme :

$$X_t = \rho_j^t(c_{1j}\cos(\omega_j t) + c_{2j}\sin(\omega_j t)).$$

(f) En utilisant à nouveau le logiciel de votre choix, tracer des exemples de série X_t précédente en fonction de la valeur de λ_j (ou précisément de son module et son argument). Quels types de courbes obtient-on ?

(g) *Exemple d'application.* Résoudre dans $\mathbb{R}$ l'équation de récurrence linéaire :

$$X_{t+2} - 4X_{t+1} + 8X_t = 0.$$

(5) Supposons maintenant que toutes les racines du polynôme donné en (2) ne soient plus distinctes et notons α_j l'ordre de multiplicité de la racine λ_j. Toute solution de l'équation de récurrence linéaire 1 s'écrit alors (dans $\mathbb{C}$) :

$$X_t = \sum_{j=1}^{J}(c_{1j} + c_{2j}t + \cdots + c_{j\alpha_j}t^{\alpha_j - 1})\lambda_j^t,$$

où les coefficients $c_{1j}, \ldots, c_{j\alpha_j}$, pour $j = 1, \ldots, J$, sont dans $\mathbb{C}$ et J est le nombre de racines distinctes. En utilisant le même raisonnement que dans le cas précédent où toutes les racines étaient distinctes, écrire la forme générale d'une solution (dans $\mathbb{R}$) de l'équation de récurrence linéaire (1).

Exercice TD 1.2

Soient les moyennes mobiles $M_{(1)}$ et $M_{(2)}$ suivantes :

$$M_{(1)} = \frac{1}{2}B\left(I + F\right),$$
$$M_{(2)} = \frac{1}{4}B\left(I + F + F^2 + F^3\right).$$

(1) (a) Quelles sont les séries temporelles absorbées (annulées) par $M_{(1)}$?
(b) Quelles sont les séries temporelles invariantes par $M_{(1)}$?
(2) (a) Quelles sont les séries temporelles absorbées par $M_{(2)}$?
(b) Quelles sont les séries temporelles invariantes par $M_{(2)}$?
(3) Montrer que $M_{(1)}M_{(2)}$ laisse invariantes les tendances linéaires.
(4) Quelles sont les séries temporelles absorbées par $\left(M_{(2)}\right)^2$?

Exercice TD 1.3

Soient les moyennes mobiles suivantes :

$$M_{(1)} = \frac{1}{3}B^2\left(I + F + F^2\right),$$
$$M_{(2)} = 2M_{(1)} - \left(M_{(1)}\right)^2.$$

(1) Montrer que $M_{(2)}$ laisse invariantes les tendances linéaires.
(2) $M_{(2)}$ absorbe-t-elle les saisonnalités périodiques d'ordre 3 ?
(3) Quelles sont les séries temporelles absorbées par $M_{(2)}$ et $\left(M_{(2)}\right)^2$?
(4) Calculer le pouvoir de réduction de variance de $M_{(2)}$.

Exercice TD 1.4

Soit $(Y_t)_{t\in\{1,\ldots,\tau\}}$ une série temporelle dont la tendance est localement polynomiale de degré $p = 3$:

$$\forall h \in \{-m, \ldots, m\} : Y_{t+h} = a_0 + a_1 h + a_2 h^2 + a_3 h^3 + \varepsilon_{t+h},$$

où ε_t est une perturbation aléatoire d'espérance nulle.
On veut estimer les paramètres a_i pour $i \in \{0, 1, 2, 3\}$ du modèle par les moindres carrés ordinaires sur les $2m + 1$ points $Y_{t-m}, \ldots, Y_{t+m}$. On note $\widehat{a}_i$ cet estimateur de a_i.

(1) Déterminer le système d'équations définissant les $\widehat{a}_i$.
(2) En déduire que $\hat{a}_0$ est (avec $\hat{a}_2$) solution du système :

$$\left\{\begin{array}{rcl} \hat{a}_0 S_0 + \hat{a}_2 S_2 & = & \sum_{h=-m}^{h=m} Y_{t+h} \\ \hat{a}_0 S_2 + \hat{a}_2 S_4 & = & \sum_{h=-m}^{h=m} h^2 Y_{t+h} \end{array}\right.,$$

où

$$S_k = \sum_{h=-m}^{h=m} h^k$$

pour $k = 0, \ldots, 4$.

Expliciter l'estimateur $\widehat{a}_0$ comme une combinaison linéaire des valeurs

$$Y_{t-m}, \ldots, Y_{t+m}.$$

Montrer que cela définit une moyenne mobile M_{2m+1}^3 centrée symétrique d'ordre $2m+1$ appelée moyenne mobile cubique sur $2m+1$ points dont les coefficients sont

$$\theta_h = \frac{h^2 S_2 - S_4}{S_2^2 - S_0 S_4},$$

pour $h = -m, \ldots, -1, 0, 1, \ldots m$.

(3) Montrer que M_{2m+1}^3 laisse invariantes les tendances polynomiales de degré inférieur ou égal à trois.
(4) On considère maintenant la moyenne mobile cubique sur 7 points : M_7^3.
   (a) Expliciter les coefficients de M_7^3.
   (b) Vérifier que M_7^3 peut s'écrire :

$$M_7^3 = I - \frac{1}{21}(I - B)^4\left(2F^3 + 5F^2 + 2F\right).$$

(c) Calculer le pouvoir de réduction de variance de M_7^3.

Exercice TD 1.5

Les moyennes mobiles d'ordre $2m+1$ présentent l'inconvénient de ne pas permettre le calcul des m premiers termes et surtout des m derniers points. Cet exercice a pour but de montrer que dans le cas de moyennes mobiles définies à partir d'ajustements locaux de polynômes, il est aisé de trouver des formules d'extrapolation de la moyenne mobile pour ces points.

(1) Déterminer les coefficients de la moyenne mobile M_5^1 d'ordre 5 obtenue par ajustement local d'une droite $Y_t = at + b$ par la méthode des moindres carrés.

(2) Calculer les coefficients $\hat{a}$ et $\hat{b}$ pour $t = \tau - 2$, c'est à dire pour le dernier ajustement possible.

(3) En déduire les valeurs ajustées $\hat{Y}_{\tau_1}$ et $\hat{Y}_\tau$ en fonction des valeurs de $Y_\tau, Y_{\tau_1}, \ldots, Y_{\tau-4}$.

CHAPITRE 2

Modélisation aléatoire des séries temporelles

1. Processus stochastiques

1.1. Définition.

Une série temporelle, quand elle n'est pas déterministe, comme ce sera le cas dans toutes les situations que nous rencontrerons, est vue comme une réalisation d'un processus stochastique.

DÉFINITION 2.1. *On appelle* ***processus stochastique*** *toute famille de variables aléatoires (v.a.)* $(X_t)_{t\in T}$ *d'un espace probabilisé* $(\Omega, \mathcal{A}, P)$ *vers un espace probabilisable* $(E, \mathcal{E})$*, i.e.*

$$\forall t \in T : X_t \text{ est une v.a. de } (\Omega, \mathcal{A}, P) \text{ vers } (E, \mathcal{E}).$$

L'ensemble T *est appelé* ***espace des temps*** *et* E ***espace des états****. Chacun de ces espaces peut être discret ou continu.*

Pour $\omega \in \Omega$*, on appelle* ***trajectoire du processus*** *la fonction (déterministe) :*

$$t \mapsto X_t(\omega).$$

Dans ce cours, les séries temporelles seront à espace des temps discret, à savoir $T = \mathbb{N}$, ou pour des raisons mathématiques $T = \mathbb{Z}$. En général on supposera que l'on observe qu'une partie de la trajectoire de ce processus. Comme espace des états on considèrera plus généralement $E = \mathbb{R}$ et la tribu sur cet espace sera la tribu borellienne $\mathcal{E} = \mathcal{B}_{\mathbb{R}}$. Mais bien sûr le cas de dimensions supérieures pour les espaces des temps ou des états peuvent être considérés. Si T est de dimension supérieure ou égale à 2, on parle de processus spatiaux (Cf. cours de géostatistique de 5GMM-MMS). Si $E = \mathbb{R}^n$ on parle de séries temporelles multivariées.

1.2. Premiers exemples de séries temporelles.

1.2.1. *Bruits blancs.*

DÉFINITION 2.2. *Un* ***bruit blanc fort*** *est une suite de variables indépendantes et identiquement distribuées (i.i.d.)* $(X_t)_{t\in\mathbb{Z}}$ *centrées et de variance* σ^2*. On note* $(X_t) \sim bbF(0, \sigma^2)$*.*

Par définition il n'y a donc aucune dépendance entre les observations. Ainsi, par exemple, la connaissance de $X_1, \ldots, X_n$ n'informe en rien sur la valeur (future) de X_{n+h}.

Le processus binaire est un exemple de bruit blanc fort où la loi pour tout t est donnée par :

$$P(X_t = 1) = p \; ; \; P(X_t = -1) = 1 - p.$$

Définition 2.3. *On appelle* ***bruit blanc gaussien****, tout bruit blanc fort pour lequel la loi commune des v.a.r.* $(X_t)_{t\in\mathbb{Z}}$ *est une* $\mathcal{N}(0,\sigma^2)$.

Définition 2.4. *On appelle* ***bruit blanc faible****, toute suite de v.a.r.* $(X_t)_{t\in\mathbb{Z}}$*, centrées, non corrélées et de variance* σ^2*. On note :* $(X_t) \sim bb(0,\sigma^2)$.

En fait, en tant que tels, ces modèles n'ont aucun intérêt pour celui qui veut faire de la prévision. Mais ils seront utilisés pour construire des séries temporelles plus compliquées, donc plus intéressantes...

1.2.2. *Marche aléatoire.*

Définition 2.5. *Une* ***marche aléatoire*** $(S_t)_{t\in\mathbb{N}}$ *est obtenue en écrivant*

$$S_t = X_1 + \ldots + X_t$$

pour tout t *dans* $\mathbb{N}$ *et où* $(X_t)_{t\in\mathbb{N}}$ *est un bruit blanc fort. Si ce dernier est un processus binaire, alors la marche aléatoire* $(S_t)_{t\in\mathbb{N}}$ *est dite* ***symétrique simple****.*

1.2.3. *Processus gaussien.*

Définition 2.6. *Un processus* ***gaussien*** *à temps discret* $(X_t)_{t\in\mathbb{Z}}$ *est une série temporelle telle que la loi de n'importe quel vecteur extrait est gaussien, i.e.*

$$\forall n \in \mathbb{N}^*, \forall (t_1,\ldots,t_n) \in \mathbb{Z}^n : (X_{t_1},\ldots,X_{t_n}) \text{ est un vecteur gaussien.}$$

Un bruit blanc gaussien est une série temporelle gaussienne.

1.2.4. *Processus MA(1).*
Introduisons maintenant un premier exemple, simple mais fondamental, de séries temporelles.

Définition 2.7. *On appelle* ***processus moyenne mobile d'ordre 1*** *toute série temporelle* $(X_t)_{t\in\mathbb{Z}}$ *définie par :*

$$X_t = \varepsilon_t + \theta\varepsilon_{t-1},\ \forall t \in \mathbb{Z},$$

où $(\varepsilon_t) \sim bb(0,\sigma^2)$ *et* $\theta \in \mathbb{R}$*. On note :* $(X_t) \sim$*MA(1).*

1.2.5. *Processus AR(1).*
Un deuxième exemple, également simple mais fondamental, est le suivant.

Définition 2.8. *On appelle* ***processus autoregressif d'ordre 1*** *toute série temporelle* $(X_t)_{t\in\mathbb{Z}}$ *définie par :*

$$X_t = \varphi_0 + \varphi_1 X_{t-1} + \varepsilon_t,\ \forall t \in \mathbb{Z},$$

où $(\varepsilon_t) \sim bb(0,\sigma^2)$ *et* $\theta \in \mathbb{R}$*. On note :* $(X_t) \sim$*AR(1).*

2. Processus du second ordre

2.1. Espace L^2.

On donne ici brièvement quelques éléments importants sur l'espace L^2. Il ne devrait s'agir que de rappels...

DÉFINITION 2.9. *On dit qu'une v.a.r. est dans l'espace $L^2(\Omega, \mathcal{A}, P)$, si l'on a*

$$\mathbb{E}X^2 < +\infty.$$

Cet espace $L^2(\Omega, \mathcal{A}, P)$ est un espace de Hilbert muni du produit scalaire

$$\langle X, Y\rangle = \mathbb{E}(XY).$$

et de la norme

$$||X||_{L^2} = \sqrt{\mathbb{E}(X^2)}.$$

On a donc dans cet espace :

$$X \perp Y \iff \mathbb{E}(XY) = 0 \iff (\text{Cov}(X, Y) = 0 \text{ si les v.a.} X \text{ et } Y \text{sont centrées}).$$

On rappelle l'inégalité de Cauchy-Schwarz pour les v.a. de carré intégrables.

THÉORÈME 2.10. *Si X et Y sont deux v.a. dans $L^2(\Omega, \mathcal{A}, P)$ alors la v.a. XY est dans $L^1(\Omega, \mathcal{A}, P)$ et on a :*

$$||XY||_{L^1} = \mathbb{E}|XY| \leq ||X||_{L^2}||Y||_{L^2}.$$

Les projections dans l'espace L^2 sont facilitées par la propriété d'Hilbert de cet espace. Rappelons d'abord la définition d'un sous espace vectoriel fermé d'un espace de Hilbert (écrite dans le cas où l'espace de Hilbert est $L^2(\Omega, \mathcal{A}, P)$)

DÉFINITION 2.11. *Un sous-espace vectoriel $\mathcal{H}$ de l'espace $L^2(\Omega, \mathcal{A}, P)$ est dit fermé s'il contient toutes ses limites, i.e. si :*

$$\Big(X_n \in \mathcal{H},\ \textit{pour tout } n \in \mathbb{N},\ \textit{et } X_n \xrightarrow{L^2} X,\ \textit{quand } n \to +\infty\Big) \implies X \in \mathcal{H}.$$

Le théorème de projection qui suit est très connu et sera utilisé en prévision des séries temporelles.

THÉORÈME 2.12. *Soit X une v.a. de l'espace $L^2(\Omega, \mathcal{A}, P)$ et $\mathcal{H}$ un sous espace vectoriel fermé de cet espace $L^2(\Omega, \mathcal{A}, P)$. Il existe alors une unique v.a. $\hat{X}$ dans $\mathcal{H}$ telle que :*

$$||X - \hat{X}||_{L^2} = \min_{Y \in \mathcal{H}} ||X - Y||_{L^2}.$$

*La v.a. $\hat{X}$ correspond à la **projection orthogonale** de X sur l'espace $\mathcal{H}$ et est donc telle que :*

$$\hat{X} \in \mathcal{H} \textit{ et } X - \hat{X} \perp \mathcal{H}.$$

2.2. Processus du second ordre.

DÉFINITION 2.13. *Une série temporelle* $(X_t)_{t\in T}$ *est dite du* ***second ordre*** *si la v.a.* X_t *est dans* $L^2(\Omega, \mathcal{A}, P)$*, pour tout* t *dans* T*.*

On rappelle également la définition de convergence dans L^2. Nous l'utiliserons ici pour justifier la convergence de sommes d'un nombre infini de v.a. dans L^2.

DÉFINITION 2.14. *Soient* $(X_n)_{n\in\mathbb{N}}$ *une suite de v.a. dans* $L^2(\Omega, \mathcal{A}, P)$ *et* X *une v.a. dans* $L^2(\Omega, \mathcal{A}, P)$*. On dit que* $(X_n)_{n\in\mathbb{N}}$ *converge dans* L^2 *vers* X *si l'on a :*

$$\lim_{n\to+\infty} ||X_n - X||_{L^2} = 0.$$

La proposition suivante nous sera très utile dans la suite pour autoriser l'inversion entre les signe somme et le signe intégral (l'espérance mathématique).

PROPOSITION 2.15. *Si la série* $\sum_{i=0}^{n} X_i$ *converge dans* L^2 *(vers* $\sum_{i=0}^{+\infty} X_i$*), alors on a :*

$$\mathbb{E}\left(\sum_{i=0}^{+\infty} X_i\right) = \sum_{i=0}^{+\infty} \mathbb{E}(X_i).$$

EXERCICE 2.1. Preuve de cette proposition.
Notons $U_n = \sum_{i=0}^{n} X_i$ *et* $U = \sum_{i=0}^{+\infty} X_i$*.*

- *Montrer que l'on a :*

$$|\mathbb{E}(U_n) - \mathbb{E}(U)| \leq ||U_n - U||_{L^2}.$$

- *Conclure.* □

Remarque. La proposition précédente reste valable pour une série indexée dans les deux directions, i.e., sous réserve de la convergence de $\sum_{i=-m}^{n} X_i$ vers $\sum_{i=-\infty}^{+\infty} X_i$ quand n et m tendent vers $+\infty$, on a :

$$\mathbb{E}\left(\sum_{i=-\infty}^{+\infty} X_i\right) = \sum_{i=-\infty}^{+\infty} \mathbb{E}(X_i).$$

PROPOSITION 2.16. *Si les séries* $\sum_{i=0}^{n} X_i$ *et* $\sum_{j=0}^{n} Y_j$ *convergent dans* L^2 *(vers* $\sum_{i=0}^{+\infty} X_i$ *et* $\sum_{j=0}^{+\infty} Y_j$ *respectivement), alors on a :*

$$\mathbb{E}\left(\sum_{i=0}^{+\infty} X_i \cdot \sum_{j=0}^{+\infty} Y_j\right) = \sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty} \mathbb{E}(X_i Y_j).$$

EXERCICE 2.2. Preuve de cette proposition.
On note $U_n = \sum_{i=0}^{n} X_i$*,* $V_n = \sum_{i=0}^{n} Y_i$ *ainsi que* U *et* V *leurs limites respectives.*

- *Montrer que pour toutes v.a.* U_n, V_n, U *et* V *dans* L^2 *on a :*
$$\mathbb{E}(U_n V_m) - \mathbb{E}(UV) = \mathbb{E}(U_n(V_m - V)) + \mathbb{E}((U_n - U)V).$$
- *En déduire l'égalité de la proposition.*

L'extension aux séries bidimensionnelles reste vraie ici aussi. Enfin, il découle des deux précédentes proposition, le corollaire suivant.

COROLLAIRE 2.17. *Si les séries* $\sum_{i=0}^n X_i$ *et* $\sum_{j=0}^n Y_j$ *convergent dans* L^2 *(vers* $\sum_{i=0}^{+\infty} X_i$ *et* $\sum_{j=0}^{+\infty} Y_j$ *respectivement), alors on a :*
$$Cov\left(\sum_{i=0}^{+\infty} X_i, \sum_{j=0}^{+\infty} Y_j\right) = \sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty} Cov(X_i, Y_j).$$

EXERCICE 2.3. Preuve de ce corollaire.
En écrivant qu'une covariance est l'espérance du produit moins le produit des espérances, montrer ce résultat.

3. Processus stationnaires

On a vu dans le premier chapitre qu'une série temporelle peut souvent être décomposée sous la forme
$$Y_t = m_t + s_t + X_t,$$
où m_t est une fonction à variation lente appelée tendance et s_t une fonction périodique (de somme nulle) appelée saisonnalité. Le terme restant, le processus X_t, est donc supposé être "plus stable" dans un sens restant à définir et que nous appellerons stationnarité.

3.1. Stationnarité forte.

DÉFINITION 2.18. *Une série temporelle* $(X_t)_{t\in\mathbb{Z}}$ *est dite* ***stationnaire au sens fort*** *si la loi de tout vecteur* $(X_{t_1}, \ldots, X_{t_n})$ *est invariante par translation temporelle, i.e. si*
$$\mathcal{L}(X_{t_1}, \ldots, X_{t_n}) = \mathcal{L}(X_{t_1+h}, \ldots, X_{t_n+h}),\ \forall (t_1, \ldots, t_n) \in \mathbb{Z}^n \text{ et } h \in \mathbb{Z}.$$

EXERCICE 2.4.

- *Si la série temporelle* $(X_t)_{t\in\mathbb{Z}}$ *est stationnaire au sens fort, que dire de la loi de* X_t*, pour tout* t *?*
- *Le bruit blanc fort est-il fortement stationnaire ? Et le bruit blanc faible ?*

Cette hypothèse de stationnarité forte est très (trop) contraignante. Elle est souvent peu réaliste. C'est pourquoi on introduit l'hypothèse de stationnarité faible.

3.2. Stationnarité faible.

On va introduire une notion de stationnarité plus faible pour les processus du second ordre. Introduisons en premier lieu les fonctions moyennes et fonctions variances pour les processus du second ordre.

DÉFINITION 2.19. *Soit $(X_t)_{t\in\mathbb{Z}}$ une série temporelle du second ordre. La* ***fonction moyenne*** *de (X_t) est définie sur $\mathbb{Z}$ par :*

$$t \mapsto \mathbb{E}X_t.$$

La ***fonction covariance*** *est définie sur $\mathbb{Z}^2$ par :*

$$\begin{aligned} (s,t) \mapsto Cov(X_s, X_t) &= \mathbb{E}\left((X_s - \mathbb{E}X_s)(X_s - \mathbb{E}X_t)\right) \\ &= \langle X_s - \mathbb{E}X_s, X_t - \mathbb{E}X_t \rangle_{L^2} \\ &= \mathbb{E}X_s X_t - \mathbb{E}X_s \mathbb{E}X_t \end{aligned}$$

DÉFINITION 2.20. *Un processus du second ordre $(X_t)_{t\in\mathbb{Z}}$ est dit (faiblement)* ***stationnaire*** *si :*

- *sa fonction moyenne est constante, i.e.*

$$\mathbb{E}X_t = \mu_X, \ \forall t \in \mathbb{Z};$$

- *sa fonction covariance ne dépend que de la différence en temps, i.e.*

$$\begin{aligned} Cov(X_s, X_t) &= Cov(X_{s+u}, X_{t+u}), \forall (s,t,u) \in \mathbb{Z}^3 \\ &= Cov(X_0, X_{t-s}) = f(t-s) \end{aligned}$$

EXERCICE 2.5.

- *Quel(s) lien(s) d'implication existe-t-il entre stationnarité forte et stationnarité (faible) ? Démontrer ces résultats.*
- *Que dire de la fonction variance d'une série temporelle faiblement stationnaire ?*

EXERCICE 2.6.

- *Préciser la fonction moyenne et la fonction covariance d'un bruit blanc fort. Ce processus est-il stationnaire ? Et un bruit blanc faible ? Justifier les réponses.*
- *Préciser la fonction moyenne, la fonction variance et la fonction covariance d'une marche aléatoire telle que définie dans la Définition 2.5. Une marche aléatoire est-elle stationnaire ?*

3.3. Fonctions d'autocovariance et d'autocorrélation.

DÉFINITION 2.21. *Soit* (X_t) *une série temporelle faiblement stationnaire. Sa* ***fonction d'autocovariance*** $\gamma_X(\cdot)$ *est la fonction définie sur* $\mathbb{Z}$ *par :*

$$\begin{array}{rcl} \gamma_X : \mathbb{Z} & \to & \mathbb{R} \\ h & \mapsto & \gamma_X(h) = Cov(X_h, X_0) = Cov(X_{t+h}, X_t),\ \forall t. \end{array}$$

Le tracé de la fonction $h \mapsto \gamma_X(h)$ *pour* h *dans* $\mathbb{N}$ *est appelé* ***autocovariogramme****.*

La fonction ***d'autocorrélation*** $\rho_X(\cdot)$ *est définie par :*

$$\begin{array}{rcl} \rho_X : \mathbb{Z} & \to & [-1,1] \\ h & \mapsto & \rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = Corr(X_{t+h}, X_t) = \frac{Cov(X_{t+h},X_t)}{\sqrt{VarX_t VarX_{t+h}}},\ \forall t. \end{array}$$

Le tracé de la fonction $h \mapsto \rho_X(h)$ *pour* h *dans* $\mathbb{N}$ *est appelé* ***autocorrélogramme simple****.*

Ces deux fonctions mesurent le degré de dépendance entre les valeurs d'un série à des instants différents.

PROPOSITION 2.22. *La fonction d'autocovariance vérifie les propriétés suivantes :*

- $\gamma(0) \geq 0$ *;*
- $|\gamma(h)| \leq \gamma(0)$*, pour tout* h*, la fonction autocovariance est donc bornée ;*
- $\gamma(h) = \gamma(-h)$*, pour tout* h*, autrement dit l'autocovariance est une fonction paire.*

EXERCICE 2.7.
Démontrer cette proposition. Pour le deuxième point, on pourra utiliser une propriété bien connue de la corrélation entre deux v.a.

EXERCICE 2.8.

- *Calculer, pour un processus MA(1) tel que défini dans la Définition 2.7, la fonction moyenne, la fonction variance et la fonction covariance. Ce processus est-il stationnaire ? Calculer sa fonction d'autocorrélation.*
- *Considérons un processus stationnaire autorégressif d'ordre 1, AR(1) défini par :*
$$X_t = \phi X_{t-1} + \varepsilon_t,\ \forall t \in \mathbb{Z},$$
où $(\varepsilon_t) \sim bb(0, \sigma^2)$ *et* ϕ *est un réel tel que* $|\phi| < 1$*. On suppose de plus que le processus* (ε_t) *est tel que, pour tout* t*, la v.a.* ε_t *est décorrélée avec tous les* X_s *pour* $s < t$*, i.e.*
$$\begin{array}{rcl} Cov(\varepsilon_t, X_s) = 0,\ \forall s < t & \iff & \langle \varepsilon_t, X_s \rangle_{L^2} = 0,\ \forall s < t \\ & \iff & \varepsilon_t \perp \underline{H}_{t-1} = sp\{X_{t-1}, X_{t-2}, \ldots\}. \end{array}$$
Calculer la fonction moyenne, la fonction covariance et la fonction d'autocorrélation pour tout $h \in \mathbb{Z}$*.*

Nous allons maintenant introduire une condition nécessaire et suffisante pour qu'une fonction sur $\mathbb{N}$ soit une fonction d'autocovariance. Pour cela nous avons besoin de rappeler la définition d'une fonction semi-définie positive sur $\mathbb{N}$.

DÉFINITION 2.23. *Une fonction $K(\cdot)$ à valeurs réelles définie sur $\mathbb{N}$ est dite semi-définie positive si l'on a*

$$\sum_{i,j=1}^{n} a_i K(i-j) a_j \geq 0,$$

pour tout n et tout vecteur $(a_1, \ldots, a_n)$ de $\mathbb{R}^n$.

THÉORÈME 2.24. *Une fonction réelle définie sur $\mathbb{N}$ est une fonction d'autocovariance d'une série temporelle si, et seulement si, elle est paire et semi-définie positive.*

EXERCICE 2.9. *Montrer que la fonction d'autocovariance d'un processus stationnaire $(X_t)_{t\in\mathbb{Z}}$ est toujours semi-définie positive. Pour cela, on pourra calculer*

$$Var\left(\sum_{i=1}^{n} a_i X_{t-i}\right).$$

Remarque. La fonction d'autocorrélation a les mêmes propriétés que la fonction d'autocovariance et on a de plus $\rho(0) = 1$.

DÉFINITION 2.25. *Si $(X_t)_{t\in\mathbb{Z}}$ est un processus stationnaire du second ordre, on appelle **matrice d'autocorrélation** de $(X_t, \ldots, X_{t+h-1})$ pour h dans $\mathbb{N}^*$ la matrice :*

$$R_h = \begin{pmatrix} 1 & \rho(1) & \cdots & \rho(h-1) \\ \rho(1) & 1 & \ddots & \rho(h-2) \\ \vdots & \ddots & \ddots & \vdots \\ \rho(h-1) & \cdots & \rho(1) & 1 \end{pmatrix}.$$

Cette matrice d'autocorrélation est une matrice de Toeplitz (symétrique avec égalité des termes dioganoaux).

EXERCICE 2.10. *En considérant la matrice R_h par bloc, constater qu'elle peut s'écrire en fonction de R_{h-1}.*

La fonction d'autocorrélation vérifie la même propriété de semi-définie positivité que la fonction d'autocovariance. La proposition suivante donne en outre une condition équivalente en terme du déterminant des matrices R_h.

PROPOSITION 2.26. *La fonction d'autocorrélation est également une fonction semi-définie positive. D'après le critère de Sylvester, cette propriété est équivalente à la positivité de tous les déterminants des mineurs principaux, i.e.*

$$det\ R_h \geq 0,$$

pour tout h dans $\mathbb{Z}$.

Cette propriété fixe une infinité de contraintes sur les corrélations, du genre

- $\rho^2(1) \leq 1$, pour $\det(R_2) \geq 0$,

- $(1-\rho(2))(1+\rho(2)-2\rho^2(1)) \geq 0$, et donc $1+\rho(2)-2\rho^2(1) \geq 0$, pour $\det(R_3) \geq 0$,

et ainsi de suite...

Dans la modélisation ARMA des séries temporelles, on considèrera souvent des séries en B comme opérateur appliqué à une série stationnaire. On parlera de **filtrage linéaire**. Plus précisément, notons

$$\psi(B) = \sum_{i\in\mathbb{Z}} \psi_i B^i$$

la série en B de coefficients $(\psi_j)_{j\in\mathbb{Z}}$. Appliquer ce filtrage linéaire à une série temporelle $(X_t)_{t\in\mathbb{Z}}$ donne la série $(Y_t)_{t\in\mathbb{Z}}$, telle que :

$$Y_t = \sum_{i=-\infty}^{+\infty} \psi_i X_{t-i}$$

On reviendra ultérieurement sur cette notion de série en B afin de mieux la définir et établir des résultats la concernant. Mais nous allons dès maintenant montrer que, sous une condition de sommabilité de ses coefficients, le filtrage linéaire effectué par cette série en B donne une série temporelle qui est bien définie (la série bidimensionnelle converge bien) et hérite du processus initial la propriété de stationnarité faible.

Proposition 2.27. *Soit $(X_t)_{t\in\mathbb{Z}}$ une série temporelle stationnaire de moyenne μ_X et de fonction d'autocovariance $\gamma_X(\cdot)$. Si la suite de réels $(\psi_j)_{j\in\mathbb{Z}}$ est sommable, i.e. si :*

$$\sum_{i=-\infty}^{+\infty} |\psi_i| < +\infty,$$

alors la série $(Y_t)_{t\in\mathbb{Z}}$, obtenue par filtrage linéaire de la série initiale $(X_t)_{t\in\mathbb{Z}}$, et donc définie par

$$Y_t = \psi(B)X_t = \sum_{i=-\infty}^{+\infty} \psi_i X_{t-i},$$

existe et est également stationnaire. La moyenne et la fonction d'autocovariance de la série $(Y_t)_{t\in\mathbb{Z}}$ sont alors telles que :

$$\begin{aligned} \mu_Y &= \mu_X \sum_{i=-\infty}^{+\infty} \psi_i \\ \gamma_Y(h) &= \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} \psi_i \psi_j \gamma_X(h+i-j) \end{aligned}$$

Exercice 2.11. *Preuve de ce résultat.*

- *Montrer dans un premier temps que*

$$||X_{t-i}||_{L^2}^2 = \gamma_X(0) + \mu_X^2.$$

- *En déduire que si la suite de réels* $(\psi_j)_{j\in\mathbb{Z}}$ *est sommable, alors la série*

$$\sum_{i=-\infty}^{+\infty} \psi_i X_{t-i},$$

est normalement convergente.
- *En déduire que la série* $(Y_t)_{t\in\mathbb{Z}}$ *est bien définie (i.e. la série converge dans* L^2*) et qu'elle est du second ordre.*
- *Calculer la fonction moyenne de* $(Y_t)_{t\in\mathbb{Z}}$ *ainsi que sa fonction d'autocovariance.*
- *Prouver la stationnarité de cette série linéaire obtenue par filtrage.*

EXERCICE 2.12. *Filtrage linéaire d'un bruit blanc*
Un cas particulier important de la proposition précédente est obtenu dans le cas où le processus initial est un bruit blanc $bb(0,\sigma_\varepsilon^2)$ *que l'on note* $(\varepsilon_t)_{t\in\mathbb{Z}}$ *. Calculer la fonction moyenne et la fonction d'autocovariance du filtrage d'un tel bruit blanc, c'est à dire du processus* $(Y_t)_{t\in\mathbb{Z}}$ *défini par :*

$$Y_t = \sum_{i\in\mathbb{Z}} \psi_i \varepsilon_{t-i}.$$

EXERCICE 2.13. *Que se passe-t-il avec des sommes finies ? Le résultat est-il toujours vrai et est-ce plus simple à établir ?*

- *Considérer le processus* $(Y_t)_{t\in\mathbb{Z}}$*, défini par*

$$Y_t = \sum_{i=0}^{t} \varepsilon_{t-i}$$

où $(\varepsilon_t)_{t\in\mathbb{Z}}$ *est toujours un bruit blanc* $bb(0,\sigma_\varepsilon^2)$*. Que dire au sujet de l'existence d'un tel processus ? Est-il du second ordre ?*
- *Calculer la fonction moyenne et sa fonction covariance.*
- *Conclure.*

Ainsi, de manière paradoxale, les sommes finies peuvent nous poser plus de "problèmes" que les sommes infinies...

4. Densité spectrale

La densité spectrale est une fonction qui contient la même information que la fonction d'autocovariance, qui recense donc tous les liens de corrélations entre les valeurs de la série temporelle à des instants différents.

DÉFINITION 2.28. *Soit* $(X_t)_{t\in\mathbb{Z}}$ *un processus stationnaire de fonction d'autocovariance* $\gamma(\cdot)$*. On appelle* ***densité spectrale****, quand elle existe, la fonction* $f(\cdot)$ *définie sur* $\mathbb{R}$ *par :*

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

C'est donc la transformée de Fourier (discrète) de la fonction $\gamma(\cdot)$ *définie sur* $\mathbb{Z}$

La condition d'existence de la densité spectrale est assurée par la proposition suivante dont la preuve est évidente.

Proposition 2.29. *La densité spectrale d'un processus stationnaire $(X_t)_{t\in\mathbb{Z}}$ de fonction d'autocovariance $\gamma(\cdot)$ existe dès que l'on a :*

$$\sum_{h\in\mathbb{Z}} |\gamma(h)| < +\infty.$$

Proposition 2.30. *On suppose que*

$$\sum_{h\in\mathbb{Z}} |\gamma(h)| < +\infty.$$

La densité spectrale d'un processus stationnaire $(X_t)_{t\in\mathbb{Z}}$ est alors une fonction réelle, continue, positive, paire et 2π périodique.

Exercice 2.14. *Preuve de cette Proposition.*

- *En utilisant l'écriture du cosinus en fonction de l'exponentielle complexe ($\cos(x) = (e^{ix} + e^{-ix})/2$), montrer que l'on peut écrire :*

$$f(\omega) = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} \gamma(h)\cos(\omega h).$$

  Quelles propriétés peut-on en tirer sur la densité spectrale ?
- *Considérons, pour $N \in \mathbb{N}$, la fonction :*

$$f_N(\omega) = \frac{1}{2\pi N}\mathbb{E}\left(\left|\sum_{j=1}^{N}(X_j - \mu_X)e^{-i\omega j}\right|^2\right).$$

  Montrer que l'on a :

$$f_N(\omega) = \frac{1}{2\pi N}\sum_{j=1}^{N}\sum_{k=1}^{N}\gamma(j-k)e^{-i\omega(j-k)}.$$

- *Montrer que l'on peut écrire $f_N(\omega)$ sous la forme :*

$$f_N(\omega) = \frac{1}{2\pi N}\sum_{h:|h|<N}(N - |h|)\gamma(h)e^{-i\omega h}$$

  et finalement que :

$$f_N(\omega) = \frac{1}{2\pi}\sum_{h:|h|<N}\gamma(h)e^{-i\omega h} - \frac{1}{2\pi}\sum_{h:|h|<N}\frac{|h|}{N}\gamma(h)e^{-i\omega h}.$$

- *Déduire de l'hypothèse du théorème, la convergence et la limite du premier terme quand $N \to +\infty$.*
- *Établir la convergence vers 0 du module du second terme, en utilisant le résultat d'analyse suivant.*

LEMME 2.31. *Soit $(a_h)_{h\in\mathbb{Z}}$ une suite de réels positifs tels que*

$$\sum_{h\in\mathbb{Z}} a_h < +\infty.$$

On a alors

$$\lim_{N\to+\infty} \sum_{h:|h|<N} \frac{|h|}{N} a_h = 0.$$

- *Conclure sur la positivité de la densité spectrale.*

THÉORÈME 2.32. *Il est équivalent de connaître la fonction d'autocovariance ou la densité spectrale, quand elle existe, d'un processus stationnaire $(X_t)_{t\in\mathbb{Z}}$. On a en effet :*

$$\gamma(h) = \int_{-\pi}^{\pi} f(\omega)\cos(\omega h)d\omega = \int_{-\pi}^{\pi} f(\omega)e^{i\omega h}d\omega.$$

EXERCICE 2.15. *Preuve du Théorème.*

- *Montrer que l'on peut écrire :*

$$\int_{-\pi}^{\pi} f(\omega)\cos(\omega h)d\omega = \int_{-\pi}^{\pi} f(\omega)e^{i\omega h}d\omega$$

et finalement

$$\int_{-\pi}^{\pi} f(\omega)\cos(\omega h)d\omega = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}}\gamma(k)\int_{-\pi}^{\pi} e^{i\omega(h-k)}d\omega.$$

- *En calculant cette dernière intégrale, en déduire le résultat du théorème.*

EXERCICE 2.16. *Montrer que la densité spectrale d'un bruit blanc faible $(\varepsilon_t)_{t\in\mathbb{Z}} \sim bb(0,\sigma^2)$ est une constante dont on précisera la valeur.*

On peut montrer que, réciproquement, tout processus stationnaire de densité spectrale constante est un bruit blanc faible.

PROPOSITION 2.33. *Soit $(X_t)_{t\in\mathbb{Z}}$ un processus stationnaire dont la densité spectrale existe et est notée $f_X(\cdot)$. La série temporelle $(Y_t)_{t\in\mathbb{Z}}$, filtrage linéaire de la série initiale $(X_t)_{t\in\mathbb{Z}}$ et donc définie par :*

$$Y_t = \psi(B)X_t = \sum_{j=-\infty}^{+\infty} \psi_j X_{t-j},$$

avec

$$\sum_{i=-\infty}^{+\infty} |\psi_i| < +\infty,$$

est un processus stationnaire de densité spectrale :

$$f_Y(\omega) = f_X(\omega)\left|\sum_{j=-\infty}^{+\infty}\psi_j e^{-i\omega j}\right|^2.$$

Exercice 2.17. *Preuve de cette Proposition.*
Notons

$$M = \sum_{h\in\mathbb{Z}} |\gamma_X(h)|,$$

qui, par hypothèse d'existence de la densité spectrale, est fini.

- *En utilisant la Proposition 2.27, montrer que l'on a :*

$$\sum_{h\in\mathbb{Z}} |\gamma_Y(h)| \leq \sum_{j=-\infty}^{+\infty} |\psi_j| \cdot \sum_{k=-\infty}^{+\infty} |\psi_k| \cdot M.$$

  En déduire l'existence de la densité de la densité spectrale du processus $(Y_t)_{t\in\mathbb{Z}}$.
- *En utilisant à nouveau la Proposition 2.27, établir l'équation reliant les densités spectrales des processus* $(X_t)_{t\in\mathbb{Z}}$ *et* $(Y_t)_{t\in\mathbb{Z}}$.

Exercice 2.18. *Densité spectrale du filtrage linéaire d'un* $bb(0,\sigma^2)$
Soit $(\varepsilon_t)_{t\in\mathbb{Z}} \sim bb(0,\sigma^2)$ *et* $(X_t)_{t\in\mathbb{Z}}$ *son filtrage linéaire défini par :*

$$X_t = \sum_{j\in\mathbb{Z}} \psi_j \varepsilon_{t-j}$$

avec $\sum_{j\in\mathbb{Z}} |\psi_j| < +\infty$. *Donner la densité spectrale de ce processus.*

5. Travaux Dirigés

Exercice TD 2.1 *Stationnarité forte versus stationnarité faible*

Considérons Y une v.a. dans $L^2(\Omega, \mathcal{A}, P)$ de loi non symétrique telle que : $\mathbb{E}(Y) = 0$ et $\mathbb{E}(Y^2) = \sigma^2$. Définissions le processus $(X_t)_{t\in\mathbb{Z}}$ par :

$$X_t = (-1)^t Y.$$

(1) Montrer que $(X_t)_{t\in\mathbb{Z}}$ est un processus faiblement stationnaire.
(2) Est-il fortement stationnaire ?

Exercice TD 2.2 *Stationnarité*

Soit $(\varepsilon_t)_{t\in\mathbb{Z}}$ un bruit blanc de variance σ^2. Etudier la stationnarité des processus $(X_t)_{t\in\mathbb{Z}}$ suivants, et donner leur fonction d'autocovariance lorsque cela s'avère possible :

(1) $X_t = a + b\varepsilon_t + c\varepsilon_{t-1}$ où $(a, b, c) \in (\mathbb{R}^*)^3$;
(2) $X_t = \varepsilon_t\varepsilon_{t-1}$ (on supposera ici que le bruit blanc est un bbF$(0, \sigma^2)$ et possède des moments d'ordre 4) ;
(3) $X_t = \varepsilon_t \cos(\omega t) + \varepsilon_{t-1} \sin(\omega t)$ où $\omega \in [0, 2\pi[$.

Exercice TD 2.3 *Moyenne mobile finie unidimensionnelle*

Soit $(\varepsilon_t)_{t\in\mathbb{Z}}$ un bruit blanc de variance σ^2. Soit $(X_t)_{t\in\mathbb{Z}}$ un processus vérifiant :

$$X_t = \sum_{i=0}^{q} \theta_i \varepsilon_{t-i}$$

où $q \in \mathbb{N}^*$ et $(\theta_0, \theta_1, \ldots, \theta_q) \in \mathbb{R}^{q+1}$.

(1) Indiquer pourquoi le processus $(X_t)_{t\in\mathbb{Z}}$ est stationnaire.
(2) Calculer l'espérance de $(X_t)_{t\in\mathbb{Z}}$.
(3) Calculer la fonction d'autocovariance de $(X_t)_{t\in\mathbb{Z}}$.

Exercice TD 2.4 *Moyenne mobile infinie unidimensionnelle*

Soit $(\varepsilon_t)_{t\in\mathbb{Z}}$ un bruit blanc de variance σ^2. Soit $(X_t)_{t\in\mathbb{Z}}$ un processus défini par :

$$X_t = \sum_{i=0}^{+\infty} \theta^i \varepsilon_{t-i} \tag{3}$$

où θ est tel que $\theta \neq 0$ et $|\theta| < 1$.

(1) Vérifier que ce processus est bien défini dans L^2.
(2) Rappeler pourquoi ce processus est stationnaire.
(3) Calculer la fonction d'autocovariance et l'autocorrélogramme simple de $(X_t)_{t\in\mathbb{Z}}$?
(4) Calculer la densité spectrale de $(X_t)_{t\in\mathbb{Z}}$. La représenter graphiquement.

Exercice TD 2.5 *Processus autorégressif d'ordre 1, AR(1)*

Considérons une série temporelle $(X_t)_{t\in\mathbb{Z}}$ de type AR(1) définie par l'équation :

$$X_t = \phi X_{t-1} + \varepsilon_t, \ \forall t \in \mathbb{Z},$$

où $(\varepsilon_t) \sim bb(0, \sigma^2)$ et ϕ est un réel tel que $|\phi| < 1$. On suppose de plus que le processus (ε_t) est tel que, pour tout t, la v.a. ε_t est décorrélée avec tous les X_s pour $s < t$.

Calculer la fonction moyenne, la fonction d'autocovariance et la fonction d'autocorrélation de ce processus.

CHAPITRE 3

Statistique des processus stationnaires du second ordre

Nous avons jusqu'à présent introduit une étude descriptive des séries temporelles afin d'identifier ou supprimer les éventuelles tendance et saisonnalité. Puis nous avons introduit la modélisation probabiliste des séries temporelles en considérant les notions de stationnarité, de fonction d'autocovariance ou autocorrélation ainsi que la fonction densité spectrale. Avant d'introduire des modèles particuliers de processus stationnaires que sont les processus ARMA, nous envisageons dans ce chapitre l'inférence statistique que l'on peut faire des séries temporelles. Nous verrons comment estimer la moyenne d'une série temporelle stationnaire, sa fonction d'autocovariance ou d'autocorrélation. Nous étudierons aussi comment faire de la prévision pour les valeurs futures (donc non encore observées) de la série temporelle. Enfin nous verrons quelques tests statistiques qui permettent de se prononcer sur la stationnarité du processus étudié.

1. Estimation des moments

Supposons que l'on observe le processus stationnaire $(X_t)_{t\in\mathbb{Z}}$ sur les instants $t = 1, \dots, \tau$. Une des premières questions que l'on peut se poser est celle de l'estimation de la moyenne et des fonctions autocovariance et autocorrélation.

1.1. Estimation de la moyenne du processus stationnaire. Évidemment, l'estimation de la moyenne μ_X est facilement obtenue via l'estimateur de la moyenne empirique :

$$\hat{\mu}_X = \bar{X}_\tau = \frac{1}{\tau}\sum_{t=1}^{\tau} X_t.$$

Exercice 3.1.

- *Qu'est-ce qui justifie l'utilisation de cet estimateur de la fonction moyenne du processus considéré ?*
- *Que dire du biais de cet estimateur ?*

Le résultat suivant montre que l'estimateur considéré est L^2-consistant.

Théorème 3.1. *Si $(X_t)_{t\in\mathbb{Z}}$ est un processus stationnaire de moyenne μ_X et de fonction d'autocovariance $\gamma_X(\cdot)$ alors :*

- *si $\gamma_X(h) \to 0$ quand $h \to 0$, alors*
$$\mathrm{Var}(\bar{X}_\tau) \to 0, \text{ quand } \tau \to +\infty;$$
- *si de plus $\sum_{h\in\mathbb{Z}} |\gamma_X(h)| < +\infty$, alors*
$$\tau\mathrm{Var}(\bar{X}_\tau) \to \sum_{h\in\mathbb{Z}} \gamma_X(h), \text{ quand } \tau \to +\infty.$$

Exercice 3.2. *Preuve du Théorème.*

- *Montrer que*
$$\tau \mathrm{Var}(\bar{X}_\tau) = \frac{1}{\tau} \sum_{h:|h|<\tau} (\tau - |h|)\gamma_X(h),$$
puis que
$$\tau \mathrm{Var}(\bar{X}_\tau) \leq |\gamma_X(0)| + 2\sum_{h=1}^{\tau} |\gamma_X(h)|.$$
- *Utiliser alors le théorème de Césaro (pour les suites) pour conclure sur le premier point du Théorème.*
- *Utiliser le Théorème de convergence dominée sur la première égalité ci-dessus pour obtenir le second point.*

Le premier résultat prouve la convergence dans L^2 de l'estimateur vers la moyenne μ_X du processus. Le second donne la variance asymptotique de l'estimateur normalisé. Sous certaines hypothèses supplémentaires, il est possible d'obtenir un comportement asymptotiquement normal de cet estimateur de la moyenne μ_X.

Proposition 3.2. *Soit $(X_t)_{t\in\mathbb{Z}}$ un processus stationnaire défini, pour tout $t \in \mathbb{Z}$, par*
$$X_t = \mu_X + \sum_{j\in\mathbb{Z}} \psi_j \varepsilon_{t-j}$$
où $(\varepsilon_t)_{t\in\mathbb{Z}}$ est un bbF$(0,\sigma^2)$ tel que $\mathbb{E}(\varepsilon_t^4) < +\infty$ et où la suite des coefficients $(\psi_j)_{j\in\mathbb{Z}}$ est sommable et de somme non nulle, donc telle que :
$$\sum_{j\in\mathbb{Z}} |\psi_j| < +\infty \text{ et } \sum_{j\in\mathbb{Z}} \psi_j \neq 0.$$
On a alors, quand τ tend vers $+\infty$, la convergence :
$$\sqrt{\tau}\left(\bar{X}_\tau - \mu_X\right) \xrightarrow{\mathcal{L}} N(0, \sum_{h\in\mathbb{Z}} \gamma_X(h)).$$

Exercice 3.3.

Tirer de ce résultat la construction d'un intervalle de confiance pour la moyenne μ_X du processus stationnaire $(X_t)_{t\in\mathbb{Z}}$.

1.2. Estimation de la fonction d'auto-covariance.

Exercice 3.4.

- *Que proposeriez-vous comme estimateur de la fonction variance du processus stationnaire ? Sur quelle(s) propriété(s) vous basez-vous pour proposer un tel estimateur ?*
- *Que proposeriez-vous comme estimateur de la fonction d'auto-covariance ? Pour quelle(s) valeur(s) de h peut-on considérer cette estimation ? Est-elle raisonnable pour toutes celles-ci ?*
- *Donnez de la même manière un estimateur de la fonction d'auto-corrélation.*

En pratique on n'estime pas les fonctions d'auto-covariance et d'auto-corrélation pour des valeurs de h supérieures au quart du nombre τ de valeurs observées de la série temporelle. Néanmoins, ces estimateurs sont naturellement consistants quand $\tau \to +\infty$, par application de la loi des grands nombres. On a de plus le comportement asymptotique suivant.

PROPOSITION 3.3. *Soit $(X_t)_{t\in\mathbb{Z}}$ un processus stationnaire défini par*

$$X_t = \mu_X + \sum_{j\in\mathbb{Z}} \psi_j \varepsilon_{t-j}$$

où $(\varepsilon_t)_{t\in\mathbb{Z}}$ est un bbF$(0,\sigma^2)$ tel que $\mathbb{E}(\varepsilon_t^4) < +\infty$ et où la suite des coefficients $(\psi_j)_{j\in\mathbb{Z}}$ est sommable et de somme non nulle, donc telle que :

$$\sum_{j\in\mathbb{Z}} |\psi_j| < +\infty \text{ et } \sum_{j\in\mathbb{Z}} \psi_j \neq 0.$$

On a alors, quand τ tend vers $+\infty$ et pour n'importe quel k fixé, la convergence :

$$\sqrt{\tau} \begin{pmatrix} \hat{\rho}(1) - \rho(1) \\ \vdots \\ \hat{\rho}(k) - \rho(k) \end{pmatrix} \xrightarrow{\mathcal{L}} N(0, \Sigma_k),$$

où $\Sigma_k = (\Sigma_{ij})_{1\leq i,j\leq k}$ est la matrice de covariance asymptotique déterminée par

$$\Sigma_{ij} = \sum_{h\in\mathbb{Z}} \quad \{(\rho_X(h+i) + \rho_X(h-i) - 2\rho_X(i)\rho_X(h)) \\ \times (\rho_X(h+j) + \rho_X(h-j) - 2\rho_X(j)\rho_X(h))\}.$$

2. Prévision linéaire optimale

Comme nous l'avons dit en introduction de ce chapitre, les problèmes statistiques posés par les séries temporelles ne se limitent pas à celui de l'estimation des moyennes et fonctions d'autocovariance ou d'autocorrélation du processus. Un objectif important dans l'étude de telles séries est de pouvoir prédire les valeurs non encore observées de la série et cela à des horizons plus ou moins éloignés. Nous allons maintenant introduire la méthode la plus couramment utilisée à cet effet : la prévision linéaire optimale.

2.1. Espaces linéaires engendrés par un processus du second ordre.

L'approche que nous allons utiliser pour prédire les valeurs futures d'une série temporelle est grandement basée sur la propriété d'espace de Hilbert de l'espace $L^2(\Omega, \mathcal{A}, P)$. Considérons $(X_t)_{t\in\mathbb{Z}}$, un processus du second ordre.

DÉFINITION 3.4. *On appelle **espace vectoriel fermé engendré** par une famille $(X_t)_{t\in I}$, où $I \subset \mathbb{Z}$, le plus petit sous espace vectoriel fermé de $L^2(\Omega, \mathcal{A}, P)$ qui contient tous les X_t, pour t dans I. On le note $\overline{Vect}(X_t, t \in I)$.*

On a le cas particulier suivant.

PROPOSITION 3.5. *Le sous espace vectoriel fermé engendré par une famille finie $(X_t)_{t\in I}$, où $I \subset \mathbb{Z}$, est l'ensemble de toutes les combinaisons linéaires, i.e. l'ensemble des v.a. Y de la forme $Y = \sum_{i\in I} \alpha_i X_i$.*

Ainsi dans le cas d'un espace vectoriel fermé engendré par une famille finie $(X_t)_{t\in I}$ de v.a. de $L^2(\Omega, \mathcal{A}, P)$, où I est un sous ensemble fini de $\mathbb{Z}$, la projection d'une v.a. X sur $\mathcal{H} = \overline{\text{Vect}}(X_t, t \in I)$ est l'unique élément

$$\hat{X} = P_{\mathcal{H}} X = \sum_{i\in I} \alpha_i X_i,$$

tel que

$$\langle X - \hat{X}, Z\rangle = 0, \text{ pour tout } Z \in \mathcal{H}$$

ou encore tel que

$$\langle \hat{X}, X_i\rangle = \langle X, X_i\rangle, \text{ pour tout } i \in I.$$

On introduit les notations suivantes :

- $\mathcal{H}_1^\tau = \overline{\text{Vect}}(1, X_1, \ldots, X_\tau)$,
- $\mathcal{H}_{-\infty}^\tau = \overline{\text{Vect}}(1, X_t, t \leq \tau)$.

2.2. Régression linéaire et innovations.

Exercice 3.5.
On considère la question de la meilleure prévision que l'on puisse faire de $X_{\tau+1}$ (ou $X_{\tau+h}$) sur la base de l'observation de la série temporelle pour (par exemple) $t = 1, \ldots, \tau$. Cherchons en premier lieu à trouver $\hat{X}_{\tau+1} = g(X_1, \ldots, X_\tau)$, pour g une fonction mesurable de $\mathbb{R}^\tau$ vers $\mathbb{R}$, qui soit le plus proche de X au sens de la norme L^2. Que proposeriez-vous comme prévision optimale $\hat{X}_{\tau+1}$ de $X_{\tau+1}$? Justifier votre réponse.

Mais une telle estimation basée sur une espérance conditionnelle n'est, en règle générale, pas aisée à calculer. C'est pourquoi, nous nous contenterons de projeter sur l'espace vectoriel fermé $\mathcal{H}_1^\tau$, engendré par les v.a. $1, X_1, \ldots, X_\tau$, qui est inclus mais bien plus restreint que

$$\mathcal{M}(X_1, \ldots, X_\tau) = \{g(X_1, \ldots, X_\tau); \text{ avec } g \text{ fonction borélienne de } \mathbb{R}^\tau \text{ vers } \mathbb{R}\}.$$

En d'autres termes nous chercherons une v.a. $\hat{X}_{\tau+1}$ qui soit une combinaison linéaire des v.a. $1, X_1, \ldots, X_\tau$, plutôt qu'une fonction mesurable quelconque de ces variables.

Définition 3.6. *On appelle **régression linéaire** d'une v.a. Y de $L^2(\Omega, \mathcal{A}, P)$ sur*

$$\mathcal{H}_1^\tau = \overline{Vect}(1, X_1, \ldots, X_\tau),$$

la projection orthogonale, au sens de la norme L^2, de Y sur cet espace. On parle aussi de régression affine sur les v.a. $X_1, \ldots, X_\tau$. On la note $EL(Y|\mathcal{H}_1^\tau)$.

D'après ce que nous avons rappelé à la fin du paragraphe précédent, on caractérise cette régression linéaire de la manière suivante.

Proposition 3.7. *Soit Y dans $L^2(\Omega, \mathcal{A}, P)$. La régression linéaire $\hat{Y} = EL(Y|\mathcal{H}_1^\tau)$, de Y sur $\mathcal{H}_1^\tau$, est la v.a. $\hat{Y} = \alpha_0 + \sum_{t=1}^\tau \alpha_t X_t$ telle que :*

$$\begin{aligned} \mathbb{E}(\hat{Y}) &= \mathbb{E}(Y) \\ \mathbb{E}(\hat{Y} X_t) &= \mathbb{E}(Y X_t), \text{ pour } t = 1, \ldots, \tau. \end{aligned}$$

Exercice 3.6.
Démontrer cette Proposition.

Revenons à notre problème de prévision d'une valeur future d'une série temporelle en fonction de l'observation de son passé.

Définition 3.8. *Soit $(X_t)_{t\in\mathbb{Z}}$ un série temporelle stationnaire. La **prévision linéaire optimale** de $X_{\tau+1}$ sachant son passé observé est :*

- $\hat{X}_{\tau+1} = EL(X_{\tau+1}|\mathcal{H}_1^\tau)$, *dans le cas d'un passé fini,*
- $\hat{X}_{\tau+1} = EL(X_{\tau+1}|\mathcal{H}_{-\infty}^\tau)$, *dans le cas d'un passé infini.*

Définition 3.9. *Soit $(X_t)_{t\in\mathbb{Z}}$ un série temporelle stationnaire et, pour tout t dans $\mathbb{Z}$, la prévision linéaire optimale $\hat{X}_t = EL(X_t|\mathcal{H}_{-\infty}^{t-1})$ de X_t sachant le passé (infini) du processus. On appelle **processus des innovations** le processus $(\varepsilon_t)_{t\in\mathbb{Z}}$ des erreurs de prévision successives, i.e. défini par*

$$\varepsilon_t = X_t - \hat{X}_t,$$

pour t dans $\mathbb{Z}$.

On a le résultat suivant.

Proposition 3.10. *Le processus des innovations $(\varepsilon_t)_{t\in\mathbb{Z}}$ est un bruit blanc.*

Exercice 3.7.
Démonstration de cette Proposition.

- *En utilisant les propriétés caractéristiques d'une projection orthogonale, montrer que le processus des innovations est centré puis que les innovations sont orthogonales deux à deux.*
- *On peut montrer que tout élément de $\mathcal{H}_{-\infty}^{t-1}$ s'écrit sous la forme $a_0+\sum_{i=1}^{+\infty}\alpha_i X_{t-i}$. En déduire que*

$$\mathrm{Var}(\varepsilon_t) = \mathrm{Var}\left(-\sum_{i=0}^{+\infty}\alpha_i X_{t-i}\right),$$

en prenant $\alpha_0 = -1$.
- *En développant cette dernière expression, achever la démonstration en montrant que cette variance est constante en temps.*

2.3. Prévision linéaire optimale dans le cas d'un passé fini.

PROPOSITION 3.11. *Soit $(X_t)_{t\in\mathbb{Z}}$ un série temporelle stationnaire de moyenne μ_X et de fonction d'autocovariance $\gamma_X(\cdot)$. La prévision linéaire optimale de $X_{\tau+h}$, pour $h \in \mathbb{N}^*$, ayant observé le passé $X_1, \ldots, X_\tau$ du processus $(X_t)_{t\in\mathbb{Z}}$, est*

$$\hat{X}_\tau(h) = \hat{X}_{\tau+h} = \alpha_0 + \sum_{t=1}^{\tau} \alpha_t X_t,$$

où les coefficients $\alpha_0, \ldots, \alpha_\tau$ sont donnés par :

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_\tau \end{pmatrix} = \Gamma_{X,\tau}^{-1} \begin{pmatrix} \gamma_X(\tau+h-1) \\ \vdots \\ \gamma_X(h) \end{pmatrix}$$

et

$$\mu_X = \alpha_0 + \mu_X \sum_{t=1}^{\tau} \alpha_t,$$

où

$$\Gamma_{X,\tau} = \begin{pmatrix} \gamma_X(0) & \gamma_X(1) & \cdots & \gamma_X(\tau-1) \\ \gamma_X(1) & \gamma_X(0) & \ddots & \gamma_X(\tau-2) \\ \vdots & \ddots & \ddots & \vdots \\ \gamma_X(\tau-1) & \cdots & \gamma_X(1) & \gamma_X(0) \end{pmatrix}.$$

est la matrice de covariance du vecteur $(X_1, \ldots, X_\tau)$, matrice supposée inversible.

Remarque. Si le processus $(X_t)_{t\in\mathbb{Z}}$ est centré, alors le premier coefficient α_0 est nul et

$$\hat{X}_\tau(h) = \hat{X}_{\tau+h} = \sum_{t=1}^{\tau} \alpha_t X_t.$$

EXERCICE 3.8.
Démonstration de cette Proposition.

- *Utiliser la Proposition 3.7 pour établir d'une part l'équation avec la moyenne μ_X et, d'autre part, l'égalité :*

$$\alpha_0 \mathbb{E}(X_i) + \sum_{t=1}^{\tau} \alpha_t \mathbb{E}(X_t X_i) = \mathbb{E}(X_{\tau+h} X_i).$$

- *Utiliser ces deux équations pour montrer que*

$$\mathbb{E}(X_{\tau+h}X_i) - \mathbb{E}(X_{\tau+h})\mathbb{E}(X_i) = \sum_{t=1}^{\tau} \alpha_t \left(\mathbb{E}(X_t X_i) - \mu_X^2\right) = \sum_{t=1}^{\tau} \alpha_t Cov(X_t, X_i).$$

- *Conclure en admettant l'inversibilité de la matrice $\Gamma_{X,\tau}$.*

EXERCICE 3.9. ***Processus autorégressif d'ordre 1, AR(1)***

On rappelle qu'une série temporelle (X_t) *est un AR(1) si elle est stationnaire et vérifie l'équation :*

$$X_t = \phi X_{t-1} + \varepsilon_t, \ \forall t \in \mathbb{Z},$$

où $(\varepsilon_t) \sim bb(0, \sigma^2)$ *et* ϕ *est un réel tel que* $|\phi| < 1$*. On rappelle que les* (ε_t) *sont indépendants du passé de la série temporelle. On va appliquer la Proposition 3.11 pour trouver* $\hat{X}_\tau(1)$*, la prévision linéaire optimale de* $X_{\tau+1}$ *à l'aide du passé fini* $\mathcal{H}_1^\tau$*.*

- *En appliquant ce théorème au cas particulier considéré, on sait que la meilleure prévision linéaire de* $X_{\tau+1}$ *sur la base des observations de* $X_1, \ldots, X_\tau$ *est de la forme*
$$\hat{X}_\tau(1) = \hat{X}_{\tau+1} = \alpha_0 + \sum_{t=1}^{\tau} \alpha_t X_t.$$
Préciser pourquoi le terme α_0 *est nul.*
- *On sait que les coefficients* $(\alpha_1, \ldots, \alpha_\tau)$ *de la régression linéaire sont solutions du système :*
$$\Gamma_{X,\tau} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_\tau \end{pmatrix} = \begin{pmatrix} \gamma_X(\tau + h - 1) \\ \vdots \\ \gamma_X(h) \end{pmatrix}.$$
Donner l'expression des coefficients du système donnant $(\alpha_1, \ldots, \alpha_\tau)$ *dans ce cas précis.*
- *En déduire que* $(\alpha_1, \ldots, \alpha_\tau) = (0, \ldots, 0, \phi)$ *est une solution à cette équation.*
- *Conclure sur l'expression de la prévision linéaire optimale* $\hat{X}_\tau(1)$ *dans ce modèle.*

2.4. Evolution des prévisions linéaires optimales en fonction de la taille de la mémoire.

On considère un processus stationnaire $(X_t)_{t\in\mathbb{Z}}$ tel que ses matrices d'autocorrélation $R(h)$ sont inversibles pour tout h dans $\mathbb{N}$. On s'intéresse à la meilleure prévision linéaire optimale de X_t en fonction de l'observation des v.a. $X_{t-1}, \ldots, X_{t-k}$. L'entier k représente donc la taille de la mémoire que l'on prend en compte pour faire les prévisions.

On a vu dans la Proposition 3.11 que le terme constant n'est présent que dans le cas d'un processus non centré. Sinon ce terme constant est obtenu facilement en fonction de la moyenne du processus et des autres coefficients de la régression linéaire. Sans perte de généralité, **nous supposons donc ici que le processus est centré**. Dans cette même Proposition 3.11 on a vu comment obtenir les coefficients $\alpha_1(k), \ldots, \alpha_k(k)$ de la prévision linéaire optimale

$$\mathrm{EL}(X_t | \mathcal{H}_{t-k}^{t-1}) = \alpha_1(k) X_{t-1} + \cdots + \alpha_k(k) X_{t-k}$$

de X_t en fonction du passé observé $X_{t-1}, \ldots, X_{t-k}$. On fera attention ici à l'ordre inversé des coefficients par rapport à celui utilisé dans la Proposition 3.11.

EXERCICE 3.10. *Montrer que les coefficients sont donnés par l'équation :*

$$\begin{pmatrix} \alpha_1(k) \\ \vdots \\ \alpha_k(k) \end{pmatrix} = R(k)^{-1} \begin{pmatrix} \rho_X(1) \\ \vdots \\ \rho_X(k) \end{pmatrix}.$$

Souvent, dans la pratique, on observe de plus en plus de données avec le temps. On augmente ainsi la taille de la mémoire et il faut donc après chaque nouvelle observation recalculer les coefficients en inversant la matrice de corrélation. Nous allons chercher à obtenir une méthode itérative qui nous en dispense en permettant de déterminer les nouveaux coefficients (avec une mémoire de taille $k+1$) en fonctions des anciens (avec une mémoire de taille k).

Montrons en premier un lemme technique qui établit l'identité entre la régression sur le passé et la régression sur le futur.

LEMME 3.12. *Les coefficients de la régression de X_t sur le passé de taille de mémoire k sont les mêmes que ceux de la régression de X_t sur les k prochaines variables du processus, i.e.*

$$EL(X_t|\mathcal{H}_{t-k}^{t-1}) = \sum_{i=1}^{k} \alpha_i(k) X_{t-i} \implies EL(X_t|\mathcal{H}_{t+1}^{t+k}) = \sum_{i=1}^{k} \alpha_i(k) X_{t+i}$$

EXERCICE 3.11. *Démonstration de ce Lemme.*

En cherchant les coefficients $\beta_1(k), \ldots, \beta_k(k)$ de X_t sur ses k variables futures, par le même raisonnement que précédemment, on obtient l'équation :

$$\begin{pmatrix} \beta_1(k) \\ \vdots \\ \beta_k(k) \end{pmatrix} = R(k)^{-1} \begin{pmatrix} \rho_X(-1) \\ \vdots \\ \rho_X(-k) \end{pmatrix}.$$

- *Préciser pourquoi l'on retrouve exactement la matrice $R(k)$ dans ce cas.*
- *Déduire de l'égalité matricielle ci-dessus, que l'on a $\alpha_i(k) = \beta_i(k)$, pour $i = 1, \ldots, k$.*

La première étape dans l'établissement d'une équation récursive exprimant les coefficients pour une mémoire de taille k en fonction de ceux d'une mémoire $k-1$ est donnée par le lemme suivant.

LEMME 3.13. *On a la relation :*

$$\alpha_i(k) = \alpha_i(k-1) - \alpha_k(k)\alpha_{k-i}(k-1),$$

pour $i = 1, \ldots, k-1$.

EXERCICE 3.12. *Démonstration de ce Lemme.*

On note $P_{\mathcal{H}}(Y)$ la projection orthogonale d'une v.a. Y de L^2 sur le sous espace vectoriel fermé $\mathcal{H}$.

- *Justifier que l'on a :*
$$P_{\mathcal{H}_{t-k+1}^{t-1}}(X_t) = P_{\mathcal{H}_{t-k+1}^{t-1}} \circ P_{\mathcal{H}_{t-k}^{t-1}}(X_t).$$
- *En déduire que l'on a l'égalité :*
$$\begin{aligned}\alpha_1(k)X_{t-1} & + & \cdots + \alpha_{k-1}(k)X_{t-k+1} + \alpha_k(k)EL(X_{t-k}|\mathcal{H}_{t-k+1}^{t-1}) \\ & = & \alpha_1(k-1)X_{t-1} + \cdots + \alpha_{k-1}(k-1)X_{t-k+1}.\end{aligned}$$
- *Utiliser le Lemme 3.12 sur le terme $EL(X_{t-k}|\mathcal{H}_{t-k+1}^{t-1})$ pour conclure la preuve.*

De ce lemme que l'on vient de montrer, on tirera un algorithme récursif si l'on est capable d'exprimer également le dernier terme $\alpha_k(k)$ en fonction des $\alpha_i(k-1)$. C'est ce que donne le Lemme suivant.

Lemme 3.14. *On a la relation :*
$$\alpha_k(k) = \frac{\rho_X(k) - \sum_{i=1}^{k-1} \alpha_i(k-1)\rho_X(k-i)}{1 - \sum_{i=1}^{k-1} \alpha_i(k-1)\rho_X(i)}.$$

Exercice 3.13. *Démonstration de ce Lemme.*
On revient à l'équation matricielle :
$$\begin{pmatrix} \rho_X(1) \\ \vdots \\ \rho_X(k) \end{pmatrix} = R(k) \begin{pmatrix} \alpha_1(k) \\ \vdots \\ \alpha_k(k) \end{pmatrix}$$

- *De la dernière ligne, tirer l'égalité :*
$$\alpha_k(k) = \rho_X(k) - \sum_{i=1}^{k-1} \rho_X(k-i)\alpha_i(k).$$
- *En utilisant d'abord le Lemme 3.13 et en effectuant le changement d'indice $j = k - i$ dans l'une des sommes, achever la démonstration.*

Les formules, que l'on vient d'établir, de mise à jour des coefficients avec une augmentation de la taille de la mémoire sont connues sous le nom d'algorithme de Durbin-Levinson.

Proposition 3.15 (Algorithme de **Durbin-Levinson**).
Les coefficients de la régression linéaire $EL(X_t|\mathcal{H}_{t-k}^{t-1})$ pour une mémoire de taille k s'obtiennent en fonction de ceux de la régression linéaire $EL(X_t|\mathcal{H}_{t-k+1}^{t-1})$ pour une mémoire de taille $k-1$ grâce aux formules récursives :
$$\begin{aligned}\alpha_i(k) & = & \alpha_i(k-1) - \alpha_k(k)\alpha_{k-i}(k-1), \\ \alpha_k(k) & = & \frac{\rho_X(k) - \sum_{i=1}^{k-1} \alpha_i(k-1)\rho_X(k-i)}{1 - \sum_{i=1}^{k-1} \alpha_i(k-1)\rho_X(i)}, \\ \alpha_1(1) & = & \rho_X(1) = 1.\end{aligned}$$

3. Autocorrélations partielles

Le coefficient $\alpha_k(k)$ devant le X_{t-k}, obtenu quand on effectue la prévision linéaire optimale de X_t en fonction du passé fini $\mathcal{H}_{t-k}^{t-1}$ de la série temporelle $(X_t)_{t\in\mathbb{Z}}$ joue un rôle particulier dans l'étude des séries temporelles. Nous allons en effet voir qu'il est équivalent de connaître ces coefficients que les autocorrélations. Ils portent le nom d'autocorrélations partielles.

Proposition 3.16. *Le coefficient $\alpha_k(k)$ défini par*

$$EL(X_t|\mathcal{H}_{t-k}^{t-1}) = \alpha_1(k)X_{t-1} + \cdots + \alpha_k(k)X_{t-k}$$

est également le coefficient de corrélation entre les variables $X_t - EL(X_t|\mathcal{H}_{t-k+1}^{t-1})$ et $X_{t-k} - EL(X_{t-k}|\mathcal{H}_{t-k+1}^{t-1})$.

Exercice 3.14. *Démonstration de cette Proposition.*

- *Justifier l'égalité :*

$$EL(X_t|\mathcal{H}_{t-k+1}^{t-1}) = \alpha_1(k)X_{t-1} + \cdots + \alpha_{k-1}(k)X_{t-k+1} + \alpha_k(k)EL(X_{t-k}|\mathcal{H}_{t-k+1}^{t-1}).$$

- *En déduire que :*

$$\begin{aligned}&Cov\left(EL(X_t|\mathcal{H}_{t-k}^{t-1}) - EL(X_t|\mathcal{H}_{t-k+1}^{t-1}), X_{t-k} - EL(X_{t-k}|\mathcal{H}_{t-k+1}^{t-1})\right)\\ &= \alpha_k(k)\mathrm{Var}\left(X_{t-k} - EL(X_{t-k}|\mathcal{H}_{t-k+1}^{t-1})\right).\end{aligned}$$

- *On va maintenant obtenir une autre expression pour cette covariance. En ajoutant et retranchant la variable X_t dans le terme de gauche de cette covariance et en utilisant un argument d'othogonalité, montrer que l'on a :*

$$\begin{aligned}&Cov\left(EL(X_t|\mathcal{H}_{t-k}^{t-1}) - EL(X_t|\mathcal{H}_{t-k+1}^{t-1}), X_{t-k} - EL(X_{t-k}|\mathcal{H}_{t-k+1}^{t-1})\right)\\ &= Cov\left(X_t - EL(X_t|\mathcal{H}_{t-k+1}^{t-1}), X_{t-k} - EL(X_{t-k}|\mathcal{H}_{t-k+1}^{t-1})\right).\end{aligned}$$

- *Conclure en admettant l'invariance en temps du terme*

$$\mathrm{Var}\left(X_{t-k} - EL(X_{t-k}|\mathcal{H}_{t-k+1}^{t-1})\right).$$

Définition 3.17. *Le coefficient de corrélation*

$$\alpha_k(k) = \frac{Cov\left(X_t - EL(X_t|\mathcal{H}_{t-k+1}^{t-1}), X_{t-k} - EL(X_{t-k}|\mathcal{H}_{t-k+1}^{t-1})\right)}{\sqrt{\mathrm{Var}\left(X_t - EL(X_t|\mathcal{H}_{t-k+1}^{t-1})\right)}\sqrt{\mathrm{Var}\left(X_{t-k} - EL(X_{t-k}|\mathcal{H}_{t-k+1}^{t-1})\right)}}$$

est appelée ***autocorrélation partielle d'ordre*** *k et est noté $r_X(k)$.*

Attention, seul le paramètre $\alpha_k(k)$ est une autocorrélation partielle ! L'autocorrrélation partielle est une corrélation entre X_t et X_{t-k} quand on leur a retiré leurs meilleures explications données par les variables intermédiaires.

Proposition 3.18. *Il est équivalent de connaître le vecteur $(\rho_X(1), \ldots \rho_X(k))$ ou le vecteur $(r_X(1), \ldots, r_X(k))$.*

4. Tests

Abordons maintenant brièvement la question des tests statistiques dans le domaine des séries temporelles. Les résultats vus sur l'estimation de la fonction moyenne, des fonctions d'autocovariance ou d'autocorrélation permettent, nous l'avons déjà dit, de déduire des intervalles de confiance pour leurs valeurs.

Une question supplémentaire apparaîtra lorsque nous aborderons, dans le prochain chapitre, la modélisation ARMA. Nous allons en effet décomposer la partie stationnaire de la série temporelle (obtenue éventuellement après estimation ou élimination des parties tendances et saisonnalités) en une partie exploitable pour la prévision (un modèle de processus stationnaire pour lequel on peut utiliser la structure de covariance pour faire de la prévision) et une partie bruit blanc (n'ayant donc plus aucune information utilisable en vue d'une prévision des valeurs futures). C'est pourquoi il est important de pouvoir se prononcer sur la blancheur des résidus.

Mathématiquement, la question revient à tester l'hypothèse

$$H_0 : (X_t)_{t\in\mathbb{Z}} \text{ est un bruit blanc}$$

contre

$$H_1 : (X_t)_{t\in\mathbb{Z}} \text{ n'est pas un bruit blanc.}$$

Pour réaliser ce test, on suppose que l'on observe les τ premières valeurs de la série temporelle, i.e. $X_1, \ldots, X_\tau$. Les résultats du théorème suivant permettent de tester les hypothèses considérées.

Théorème 3.19. *Les statistiques de* ***Portmanteau***

$$Q_k = \tau \sum_{h=1}^{k} \hat{\rho}^2(h)$$

et de ***Ljung-Box***

$$Q_k^* = \tau(\tau+2) \sum_{h=1}^{k} \frac{\hat{\rho}^2(h)}{\tau - h}$$

convergent, quand τ tend vers $+\infty$, vers une loi de χ^2_k.

Sous l'hypothèse H_0 d'un bruit blanc, la fonction d'autocovariance ou d'autocorrélation est nulle sauf en sa valeur initiale $h = 0$. Ainsi les statistiques de Portmanteau et de Ljung-Box ont elles tendance à être faibles sous H_0 et élevées sous H_1. On rejettera donc H_0 pour de grandes valeurs observées de ces statistiques. Concrètement, à un niveau α, on rejettera H_0 dès que Q_k ou Q_k^* est supérieure au quantile $\chi^2_k(1-\alpha)$ de la loi de χ^2_k.

5. Travaux Dirigés

Exercice TD 3.1

Soit $(\varepsilon_t)_{t\in\mathbb{Z}}$ un bruit blanc de variance σ^2 et $(X_t)_{t\in\mathbb{Z}}$ un processus vérifiant :

$$X_t = \varepsilon_t - \theta\varepsilon_{t-1}.$$

(1) On suppose que $|\theta| < 1$.
   (a) Calculer la prévision linéaire optimale de X_{t+1} à partir du passé $(X_i)_{i\leq t}$.
   (b) Quelle est la variance de l'erreur de prévision linéaire de X_{t+1} à partir du passé $(X_i)_{i\leq t}$?

(2) On suppose que $\theta = 1$.
   Donner l'expression de la prévision linéaire optimale de X_{t+1} à partir de $(X_i)_{1\leq i\leq t}$, ainsi que la variance de son erreur de prévision.

Exercice TD 3.2

Soit $(\varepsilon_t)_{t\in\mathbb{N}}$ un bruit blanc de variance σ^2 avec $\varepsilon_0 = 0$. Soit $(X_t)_{t\in\mathbb{N}}$ un processus vérifiant :

$$\forall t \in \mathbb{N}^* : X_t = X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}$$

où $|\theta| < 1$, et $X_0 = 0$. On peut montrer par récurrence que $\varepsilon_t \perp \mathcal{H}_0^{t-1}(X)$ pour $t \in \mathbb{N}$.

(1) Construire $\widehat{X}_1(1)$ le prédicteur de X_2 sachant X_1, puis $\widehat{X}_2(1)$ le prédicteur de X_3 sachant X_1 et X_2. Trouver ensuite une relation de récurrence simple exprimant $\widehat{X}_t(1)$ en fonction de $\widehat{X}_{t-1}(1)$ et de X_t.

(2) Pour $h > 1$, montrer que $\widehat{X}_t(h)$ est constant (ne dépend pas de h).

(3) Montrer que la solution $\widehat{X}_t(h)$ est égale à un lissage exponentiel simple du processus $(X_t)_{t\in\mathbb{Z}}$ (on indiquera le coefficient de lissage).

(4) Calculer l'erreur de prévision de X_{t+h} et trouver sa variance. Que se passe-t-il lorsque h tend vers $+\infty$?

CHAPITRE 4

Modèles ARMA

Dans ce chapitre, nous allons introduire les modèles ARMA qui sont très couramment utilisés dans l'étude des séries temporelles. Ces modèles paramétriques linéaires de séries temporelles ont été proposés par Box et Jenkins. Leurs écriture et analyse utilisent abondamment les opérateurs retard et avance. C'est pourquoi nous allons en premier lieu présenter proprement ces opérateurs ainsi que quelques résultats sur des polynômes ou séries "en B" que nous exploiterons dans le cadre des processus ARMA.

1. Polynômes et séries en B

1.1. Définitions. Nous nous restreignons dans cette partie aux processus stationnaires du second ordre. Nous rappelons que l'opérateur B (resp. F) appliqué à un processus du stationnaire $(X_t)_{t\in\mathbb{Z}}$ donne le processus $(Y_t)_{t\in\mathbb{Z}}$ tel que $Y_t = BX_t = X_{t-1}$ (resp. $Y_t = FX_t = X_{t+1}$). Les opérateurs B^k (resp. F^k) transforment eux le processus $(X_t)_{t\in\mathbb{Z}}$ en un processus $(Y_t)_{t\in\mathbb{Z}}$ tel que $Y_t = B^kX_t = X_{t-k}$ (resp. $Y_t = F^kX_t = X_{t+k}$).

On peut alors naturellement étendre la définition de ces opérateurs en considérant des **polynômes en** B (ou F). Ainsi, par exemple, le polynôme $\sum_{i=1}^p a_iB^i$ appliqué aux processus $(X_t)_{t\in\mathbb{Z}}$ donne le processus $(Y_t)_{t\in\mathbb{Z}}$ tel que

$$Y_t = \left(\sum_{i=1}^p a_iB^i\right) X_t = \sum_{i=1}^p a_iX_{t-i}.$$

Nous avons utilisé ce genre d'opérateur, et aussi des opérateurs utilisant conjointement opérateurs retard et avance, pour définir les moyennes mobiles.

Nous avons également considéré la notion de série en B quand nous avons introduit la notion de filtrage linéaire. Rappelons que nous avons vu que si $(X_t)_{t\in\mathbb{Z}}$ est un processus stationnaire, alors le processus $(Y_t)_{t\in\mathbb{Z}}$ défini par

$$Y_t = \sum_{i\in\mathbb{Z}} a_iX_{t-i}$$

est également stationnaire, sous la condition de la convergence normale de la série des coefficients réels, i.e. sous l'hypothèse $\sum_{i\in\mathbb{Z}} |a_i| < +\infty$.

DÉFINITION 4.1. *Soit $(a_i)_{i\in\mathbb{Z}}$ une famille absolument sommable de réels, i.e. telle que $\sum_{i\in\mathbb{Z}} |a_i| < +\infty$. On appelle série en B de coefficients $(a_i)_{i\in\mathbb{Z}}$ l'opérateur*

$$P(B) = \sum_{i\in\mathbb{Z}} a_iB^i$$

sur les processus stationnaires qui transforme le processus stationnaire $(X_t)_{t\in\mathbb{Z}}$ en un autre processus stationnaire $(Y_t)_{t\in\mathbb{Z}}$ défini, pour tout $t \in \mathbb{Z}$, par :

$$Y_t = P(B)X_t = (\sum_{i\in\mathbb{Z}} a_i B^i)X_t = \sum_{i\in\mathbb{Z}} a_i X_{t-i}.$$

La manipulation des séries en B se fait comme celle des séries réelles, en particulier pour les combinaisons linéaires et composées de séries.

PROPOSITION 4.2. *On a les résultats suivants.*

- *La* ***combinaison linéaire*** *de deux séries en B est encore une série en B.*
- *La* ***composée*** *de deux séries en B est encore une série en B.*

Preuve. Soient $P(B) = \sum_{i\in\mathbb{Z}} a_i B^i$ et $Q(B) = \sum_{i\in\mathbb{Z}} b_i B^i$ deux séries en B, donc telles que $\sum_{i\in\mathbb{Z}} |a_i| < +\infty$ et $\sum_{i\in\mathbb{Z}} |b_i| < +\infty$. Considérons l'opérateur $\lambda P(B) + Q(B)$ et appliquons le à un processus stationnaire $(X_t)_{t\in\mathbb{Z}}$. On obtient, pour tout $t \in \mathbb{Z}$:

$$\begin{aligned} (\lambda P(B) + Q(B))\, X_t &= \lambda P(B)X_t + Q(B)X_t = \sum_{i\in\mathbb{Z}} \lambda a_i X_{t-i} + \sum_{i\in\mathbb{Z}} b_i X_{t-i} \\ &= \lim_{n\to+\infty} \sum_{i=-n}^{i=+n} \lambda a_i X_{t-i} + \lim_{n\to+\infty} \sum_{i=-n}^{i=+n} b_i X_{t-i} \\ &= \lim_{n\to+\infty} \sum_{i=-n}^{i=+n} (\lambda a_i X_{t-i} + b_i X_{t-i}) = \sum_{i\in\mathbb{Z}} (\lambda a_i + b_i) X_{t-i}, \end{aligned}$$

où la dernière égalité est justifiée par l'absolue convergence de la série $\sum(\lambda a_i + b_i)$ garantie par l'inégalité

$$\sum_{i\in\mathbb{Z}} |\lambda a_i + b_i| \leq \lambda \sum_{i\in\mathbb{Z}} |a_i| + \sum_{i\in\mathbb{Z}} |b_i| < +\infty.$$

On a bien alors

$$\lambda \sum_{i\in\mathbb{Z}} a_i B^i + \sum_{i\in\mathbb{Z}} b_i B^i = \sum_{i\in\mathbb{Z}} (\lambda a_i + b_i) B^i.$$

Considérons maintenant la composition des deux séries $P(B)$ et $Q(B)$. On a

$$P(B) \circ Q(B) X_t = P(B) Y_t,$$

où

$$Y_t = \sum_{j\in\mathbb{Z}} b_j X_{t-j}.$$

On a donc

$$\begin{aligned}
P(B)\circ Q(B)X_t &= \sum_{i\in\mathbb{Z}} a_i Y_{t-i} = \lim_{n\to+\infty}\sum_{i=-n}^{n} a_i Y_{t-i}\\
&= \lim_{n\to+\infty}\sum_{i=-n}^{n} a_i\left(\lim_{m\to+\infty}\sum_{j=-m}^{m} b_j X_{t-i-j}\right)\\
&= \lim_{n\to+\infty}\lim_{m\to+\infty}\sum_{i=-n}^{n}\sum_{j=-m}^{m} a_i b_j X_{t-i-j}\\
&= \lim_{n\to+\infty}\lim_{m\to+\infty}\sum_{k=-n-m}^{n+m}\left(\sum_{i=\max(-n,k-m)}^{\max(n,k+m)} a_i b_{k-i}\right) X_{t-k},
\end{aligned}$$

où la dernière égalité est obtenue en faisant un changement d'indice de j à $k = i + j$.

On sait que les séries $\sum a_i$ et $\sum b_j$ étant absolument sommables, la série produit de Cauchy (appelé aussi parfois convoluée) $\sum c_k$ de ces deux suites, où

$$c_k = \sum_{i\in\mathbb{Z}} a_i b_{k-i},$$

pour $k \in \mathbb{Z}$, l'est aussi. On a donc :

$$P(B)\circ Q(B)X_t = \sum_{k\in\mathbb{Z}} c_k B^k,$$

ce qui montre que la composition des deux séries en B est bien une série en B. □

1.2. Inversion de $I - \lambda B$.

Dans l'étude des modèles ARMA, on va considérer des polynômes en B et on aura besoin de savoir quand et comment on peut les inverser. Considérons en premier lieu un polynôme de degré 1 en B, le polynôme $I - \lambda B$.

Cet opérateur associe à tout processus stationnaire $(X_t)_{t\in\mathbb{Z}}$ un nouveau processus stationnaire $(Y_t)_{t\in\mathbb{Z}}$ tel que :

$$Y_t = (I - \lambda B)X_t = X_t - \lambda X_{t-1}.$$

Le problème de l'inversion du polynôme $I - \lambda B$ revient, une fois donné un processus stationnaire $(Y_t)_{t\in\mathbb{Z}}$, à déterminer s'il existe un processus stationnaire $(X_t)_{t\in\mathbb{Z}}$ vérifiant :

$$Y_t = (I - \lambda B)X_t = X_t - \lambda X_{t-1}$$

et de donner son expression en fonction du processus $(Y_t)_{t\in\mathbb{Z}}$.

On va constater que ce polynôme $I - \lambda B$ est inversible seulement si $|\lambda| \neq 1$ et on obtiendra dans ce cas l'expression de son inverse.

1.2.1. *1er Cas : la racine du polynôme est à l'extérieur strictement du disque unité, i.e.* $|\lambda| < 1$.

On suppose donc ici le module de λ strictement inférieur à 1. Considérons alors la suites $(a_i)_{i\in\mathbb{Z}}$ définie par $a_i = 0$ pour $i \in \mathbb{Z} \setminus \mathbb{N}$ et $a_i = \lambda^i$ pour $i \in \mathbb{N}$. On a alors

$$\sum_{i\in\mathbb{Z}} |a_i| = \sum_{i\in\mathbb{N}} |\lambda|^i < +\infty.$$

La série en B de coefficients (a_i) est donc bien définie et on peut écrire

$$(I - \lambda B) \sum_{i\in\mathbb{Z}} a_i B^i = (I - \lambda B) \sum_{i\in\mathbb{N}} \lambda^i B^i = \sum_{i\in\mathbb{N}} \lambda^i B^i - \sum_{i\in\mathbb{N}} \lambda^{i+1} B^{i+1} = \lambda^0 B^0 = I.$$

Ceci prouve, qu'une fois le processus stationnaire $(Y_t)_{t\in\mathbb{Z}}$ donné, le processus $(X_t)_{t\in\mathbb{Z}}$ défini par $X_t = \sum_{i=0}^{+\infty} \lambda^i Y_{t-i}$ est solution de l'équation

$$Y_t = (I - \lambda B) X_t.$$

Mais il n'est pas l'unique solution. En effet considérons un processus $(X_t^*)_{t\in\mathbb{Z}}$ tel que $(I - \lambda B) X_t^* = 0$. Le processus $(\tilde{X}_t)_{t\in\mathbb{Z}}$ défini par

$$\tilde{X}_t = X_t + X_t^*$$

est tel que

$$Y_t = (I - \lambda B) \tilde{X}_t.$$

Or, il est facile de voir qu'un processus $(X_t^*)_{t\in\mathbb{Z}}$ vérifiant $X_t^* - \lambda X_{t-1}^* = 0$ est alors de la forme $X_t^* = \lambda^t A$, pour tout t dans $\mathbb{Z}$, où A est une v.a.r. de L^2. Ainsi les solutions générales, le processus $(Y_t)_{t\in\mathbb{Z}}$ étant donné, de l'équation

$$Y_t = (I - \lambda B) \tilde{X}_t,$$

pour tout $t \in \mathbb{Z}$ sont de la forme :

$$\tilde{X}_t = X_t + X_t^* = \sum_{i=0}^{+\infty} \lambda^i Y_{t-i} + \lambda^t A.$$

En revanche, la dernière partie (le processus $(X_t^*)_{t\in\mathbb{Z}}$) n'étant pas stationnaire (s'en convaincre est facile en regardant la fonction d'autocovariance), la solution stationnaire de l'équation

$$Y_t = (I - \lambda B) X_t,$$

le processus $(Y_t)_{t\in\mathbb{Z}}$ étant donné, est unique et est le processus $(X_t)_{t\in\mathbb{Z}}$ défini par

$$X_t = \sum_{i=0}^{+\infty} \lambda^i Y_{t-i}.$$

L'opérateur $(I - \lambda B)$ est donc inversible dans la classe des processus stationnaire d'inverse $\sum_{i=0}^{+\infty} \lambda^i B^i$ quand $|\lambda| < 1$, ou encore quand la racine du polynôme est à l'extérieur du disque unité.

1.2.2. *2ème Cas : la racine du polynôme est à l'intérieur strictement du disque unité, i.e.* $|\lambda| > 1$.

On considère ici que le module de λ est strictement supérieur à 1. On peut écrire :

$$I - \lambda B = -\lambda B(I - \frac{1}{\lambda}F).$$

Par le même raisonnement que dans le cas précédent, la racine du polynôme $I - \frac{1}{\lambda}F$ étant à l'extérieur du disque unité, ce polynôme en F est inversible dans la classe des processus stationnaire et d'inverse $\sum_{i=0}^{+\infty} \lambda^{-i}F^i$. Le processus $-\lambda B$ est lui inversible d'inverse $(-1/\lambda)F$.

Ainsi le processus $I - \lambda B$, quand $|\lambda| > 1$, est inversible dans la classe des processus stationnaires d'inverse

$$\left(\frac{-1}{\lambda}F\right)\sum_{i=0}^{+\infty}\frac{1}{\lambda^i}F^i = -\sum_{i=0}^{+\infty}\frac{1}{\lambda^{i+1}}F^{i+1} = -\sum_{j=1}^{+\infty}\frac{1}{\lambda^j}F^j$$

Le processus $(Y_t)_{t\in\mathbb{Z}}$ étant donné, la solution stationnaire unique de l'équation

$$Y_t = (I - \lambda B)X_t,$$

est donc $(X_t)_{t\in\mathbb{Z}}$ défini par

$$X_t = -\sum_{j=1}^{+\infty}\frac{1}{\lambda^j}Y_{t+j}.$$

Remarquons que, cette fois-ci, son expression est fonction du futur du processus $(Y_t)_{t\in\mathbb{Z}}$ et non en fonction de son passé comme dans le cas précédent.

1.2.3. *3ème Cas : la racine du polynôme est sur le cercle unité, i.e.* $|\lambda| = 1$.

Dans le cas où le module de λ est égal à 1, il est facile de constater que l'opérateur $I - \lambda B$ n'est pas inversible dans la classe des processus stationnaires. Prenons en effet, par exemple, $\lambda = 1$ et $(X_t)_{t\in\mathbb{Z}}$ le processus constant, i.e. $X_t = m$, pour tout t dans $\mathbb{Z}$. On a alors

$$(I - \lambda B)X_t = m - m = 0,$$

quelque soit la valeur de m. Ceci prouve que l'opérateur n'est pas injectif et donc pas bijectif.

On peut aussi montrer qu'il n'est pas surjectif. En effet si le processus constant $(Y_t)_{t\in\mathbb{Z}}$ égal à $m \neq 0$ avait un antécédent $(X_t)_{t\in\mathbb{Z}}$ stationnaire, ce dernier serait tel que :

$$X_t - X_{t-1} = m,$$

pour tout t dans $\mathbb{Z}$. On arrive à une contradiction en considérant l'espérance de cette équation :

$$\mathbb{E}(X_t - X_{t-1}) = 0$$

et ne peut être égale à $m \neq 0$.

1.3. Inverse d'un polynôme en B.
Considérons maintenant comme opérateur, un polynôme en B de la forme :

$$\Phi(B) = I - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p.$$

On cherche toujours à déterminer si ce polynôme en B est inversible. C'est à dire, le processus $(Y_t)_{t\in\mathbb{Z}}$ étant donné, s'il existe un unique processus stationnaire $(X_t)_{t\in\mathbb{Z}}$ tel que

$$\Phi(B)X_t = Y_t,$$

pour tout t dans $\mathbb{Z}$.

On sait que dans le corps des complexes, ce polynôme a p racines, non nécessairement distinctes. Les polynômes que nous considérons étant non constants (sinon, ce n'est pas utile !), on sait qu'ils sont scindés. Ainsi, les résultats obtenus pour le polynôme $I - \lambda B$ nous permettent d'envisager sans peine les résultats suivants.

- Si au moins une des racines est de module 1, alors il n'existe pas de processus stationnaire $(X_t)_{t\in\mathbb{Z}}$ solution de l'équation $\Phi(B)X_t = Y_t$.
- Si toutes les racines sont de module différent de 1, alors il existe une série en B, notée $\Psi(B) = \sum \psi_i B^i$, telle que :
  - $\Phi(B)\Psi(B) = I$
  - $(X_t)_{t\in\mathbb{Z}}$, solution de $\Phi(B)X_t = Y_t$, est stationnaire.
- Si toutes les racines sont à l'extérieur du disque unité, alors l'inverse est une série en puissances positives de B uniquement, i.e.

$$\Psi(B) = \sum_{i\in\mathbb{N}} \psi_i B^i$$

- Si toutes les racines sont à l'intérieur du disque unité, alors l'inverse est une série en puissances strictement positives de F uniquement, i.e.

$$\Psi(B) = \sum_{i\in\mathbb{N}^*} \psi_i F^i$$

Le polynôme $\Phi(B)$ étant donné, son inverse quand il existe, peut être obtenu par une des méthodes suivantes.

- *Identification.* On écrit, par exemple (quand toutes les racines sont à l'extérieur du disque unité) :

$$\left(1 - \varphi_1 z - \varphi_2 z^2 - \cdots - \varphi_p z^p\right) \sum_{i\in\mathbb{N}} \psi_i z^i = 1$$

  dont on tire des équations donnant les expressions des coefficients ψ_i en fonction des φ_i.
- *Décomposition en éléments simples* de $\Phi(z)$. On écrit

$$\frac{1}{\Phi(z)} = \frac{1}{\prod_{i=1}^p (1-\lambda_i z)} = \sum_{i=1}^p \frac{a_i}{1-\lambda_i z} = \sum_{i=1}^p a_i \sum_{j\in\mathbb{Z}} \lambda_i^j z^j = \sum_{j\in\mathbb{Z}} \left(\sum_{i=1}^p a_i \lambda_i^j\right) z^j = \Psi(z).$$

- *Division du polynôme* 1 par les puissances croissantes de $\Phi(z)$.

2. Processus AR

Dans ce chapitre, on se restreint, sans perte de généralités, à des **processus centrés**. S'ils ne le sont pas, il suffit d'enlever leur moyenne pour se retrouver dans ce cas.

2.1. Définition.

Dans de nombreuses situations pratiques la valeur en un instant t d'une série temporelle peut s'écrire comme la somme d'une combinaison linéaire des valeurs précédentes de la série et d'un terme de bruit. Un tel modèle est connu sous le nom d'un processus AR (AutoRegressif).

DÉFINITION 4.3. *On dit qu'un processus stationnaire $(X_t)_{t\in\mathbb{Z}}$ admet une représentation **auto-régressive d'ordre** p **(noté AR(p))** s'il vérifie l'équation récurrente :*

$$X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} - \cdots - \varphi_p X_{t-p} = \varepsilon_t,$$

où $(\varepsilon_t)_{t\in\mathbb{Z}}$ est un bruit blanc $bb(0,\sigma^2)$ et $\varphi_1,\ldots,\varphi_p$ sont des coefficients réels, avec $\varphi_p \neq 0$. Autrement dit, en utilisant le polynôme en B

$$\Phi(B) = I - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p,$$

un processus AR(p) vérifie l'équation :

$$\Phi(B)X_t = \varepsilon_t,$$

pour tout t dans $\mathbb{Z}$.

*Cette représentation est dite **canonique** si le bruit blanc $(\varepsilon_t)_{t\in\mathbb{Z}}$ est tel que $\varepsilon_t \perp \mathcal{H}_{-\infty}^{t-1}(X)$, pour tout t dans $\mathbb{Z}$.*

Remarque. Par définition d'un processus AR, puisqu'il doit être stationnaire, **les racines du polynômes $\Phi(z)$ doivent toutes être de module différent de 1**. En effet, sinon il n'existe pas de processus stationnaire vérifiant l'équation récurrente d'un AR.

EXERCICE 4.1. *Considérons un processus AR(1) vérifiant*

$$X_t - \varphi X_{t-1} = \varepsilon_t$$

avec $\varphi = \pm 1$.

- *En écrivant X_t en fonction de X_0 et du bruit blanc, montrer que l'on a :*

$$\mathrm{Var}(X_t) + \mathrm{Var}(\varphi^t X_0) - 2\varphi^t Cov(X_t, X_0) = \sum_{j=0}^{t-1} \varphi^{2j}\mathrm{Var}(\varepsilon_{t-j}).$$

- *En supposant que le processus $(X_t)_{t\in\mathbb{Z}}$ est stationnaire, déduire de ce qui précède l'égalité :*

$$2 - 2\varphi^t \rho_X(t) = \frac{t\sigma^2}{\gamma_X(0)}.$$

- *En faisant tendre t vers $+\infty$, montrer qu'il ne peut exister de solution stationnaire qui vérifie l'équation*
$$X_t - \varphi X_{t-1} = \varepsilon_t$$
avec $\varphi = \pm 1$.

2.2. Un exemple simple mais très instructif : le processus AR(1).

On considère $(X_t)_{t\in\mathbb{Z}}$, un processus AR(1), vérifiant donc l'équation de récurrence
$$X_t - \varphi X_{t-1} = \varepsilon_t,$$
pour tout t dans $\mathbb{Z}$ et où $(\varepsilon_t)_{t\in\mathbb{Z}}$ est un bruit blanc $bb(0,\sigma^2)$.

D'après ce que l'on a vu dans la première partie sur l'inversibilité du polynôme en $I - \varphi B$, trois cas se présentent suivant les valeurs du paramètre φ.

2.2.1. *Cas* $|\varphi| = 1$.

On a vu dans le premier paragraphe de ce chapitre que, dans ce cas, il n'existe pas de processus stationnaire $(X_t)_{t\in\mathbb{Z}}$ vérifiant l'équation
$$(I - \varphi B)X_t = \varepsilon_t.$$
Il n'existe donc pas de processus AR(1) de paramètre 1 ou -1.

2.2.2. *Cas* $|\varphi| < 1$.

Exercice 4.2.

Étudions si la représentation du processus $(X_t)_{t\in\mathbb{Z}}$ est canonique.

- *Écrire X_t en fonction du bruit blanc uniquement.*
- *Que dire alors de $\mathcal{H}^t_{-\infty}(X)$ et $\mathcal{H}^t_{-\infty}(\varepsilon)$? La représentation du processus AR(1) est-elle canonique ?*

Étudions maintenant le problème de la prévision linéaire optimale. Ayant observé $X_1, \ldots, X_\tau$, on cherche donc à prédire de manière linéaire et optimale la valeur suivante de la série, i.e. $X_{\tau+1}$.

- *Donner l'expression de la prévision linéaire optimale de $X_{\tau+1}$.*
- *Quel est le processus des innovations ?*

2.2.3. *Cas* $|\varphi| > 1$.

Exercice 4.3.

On se pose les mêmes questions que dans le cas précédent.

- *Écrire X_t en fonction du bruit blanc uniquement.*
- *La représentation du processus AR(1) est-elle canonique ? On pourra calculer $Cov(\varepsilon_t, X_{t-1})$ pour répondre à la question.*
- *Que dire de la prévision linéaire optimale de $X_{\tau+1}$?*
- *Quel dire du processus des innovations ?*

Les problèmes rencontrés dans ce cas $|\varphi| > 1$ viennent de la représentation non-canonique du processus AR(1) dans ce cas. Mais on va voir facilement que le processus AR(1) ainsi défini (avec $|\varphi| > 1$), possède un autre représentation, celle-ci étant canonique.

Exercice 4.4.

- *De l'équation définissant le processus AR(1)*
$$(I - \varphi B)X_t = \varepsilon_t.$$
déduire que l'on a :
$$f_X(\omega) = \frac{\sigma^2}{2\pi} \frac{1}{|1 - \varphi e^{-i\omega}|^2},$$
où $f_X(\omega)$ est la densité spectrale du processus $(X_t)_{t\in\mathbb{Z}}$. On pourra utiliser la Proposition 2.33.
- *Considérons alors le processus $(\eta_t)_{t\in\mathbb{Z}}$ défini, pour tout t dans $\mathbb{Z}$, par :*
$$\eta_t = X_t - \frac{1}{\varphi} X_{t-1}.$$
- *Toujours avec l'aide de la Proposition 2.33, montrer que l'on a*
$$f_\eta(\omega) = \frac{\sigma^2}{2\pi\varphi^2} \frac{|1 - \frac{1}{\varphi} e^{-i\omega}|^2}{|\frac{1}{\varphi} - e^{-i\omega}|^2}.$$
- *En utilisant l'égalité $|e^{i\omega}|^2 = 1$, montrer que l'on a finalement*
$$f_\eta(\omega) = \frac{\sigma^2}{2\pi\varphi^2}.$$
- *En déduire que le processus $(\eta_t)_{t\in\mathbb{Z}}$ est aussi un bruit blanc.*
- *Donner une représentation canonique du processus $(X_t)_{t\in\mathbb{Z}}$.*
- *Conclure sur la question de la prévision linéaire optimale.*

Cette petite étude du cas particulier du processus AR(1) nous a donc permis de voir que la représentation AR(1) n'est pas unique et qu'il suffit de prendre celle avec le coefficient φ ou $1/\varphi$ inférieur à 1 pour "tomber" sur la représentation canonique. Ce résultat se généralise aux processus AR d'ordre quelconque.

2.3. Propriétés.

Avant d'étudier plus en détail les propriétés des processus AR, notons, et c'est la moindre des choses, que de tels processus existent ! La proposition suivante généralise les résultats vus sur la surjectivité et l'inversibilité du polynôme en $I - \lambda B$ et décrit le processus AR en fonction du bruit blanc.

Proposition 4.4. *Soit $(\varepsilon_t)_{t\in\mathbb{Z}}$ un bruit blanc et $\Phi(B)$ un polynôme en B.*

- *Il existe une infinité de processus du second ordre $(X_t)_{t\in\mathbb{Z}}$ vérifiant l'équation :*
$$\Phi(B)X_t = \varepsilon_t, \text{ pour tout } t \in \mathbb{Z}.$$

- *Si les racines du polynôme $\Phi(B)$ sont toutes de module différent de 1, il existe une seule solution stationnaire $(X_t)_{t\in\mathbb{Z}}$ de cette même équation. Celle-ci s'écrit alors sous la forme d'un filtrage linéaire (ou encore moyenne mobile infinie) du bruit blanc $(\varepsilon_t)_{t\in\mathbb{Z}}$, i.e. :*

$$X_t = \Psi(B)\varepsilon_t = \sum_{i\in\mathbb{Z}} \psi_i \varepsilon_{t-i}, \text{ pour tout } t \in \mathbb{Z}.$$

- *Si toutes les racines du polynômes sont à l'extérieur du disque unité, alors l'écriture moyenne mobile du processus AR(p) ne considère que les valeurs passées du bruit blanc, i.e. :*

$$X_t = \Psi(B)\varepsilon_t = \sum_{i\in\mathbb{N}} \psi_i \varepsilon_{t-i}, \text{ pour tout } t \in \mathbb{Z}.$$

Preuve. Pour s'en convaincre, il suffit de factoriser le polynôme $\Phi(B)$ de degré p en fonction de ses p racines et d'appliquer p fois le résultat d'inversibilité vu pour le polynôme $I - \lambda B$. C'est ce que nous avons décrit dans le paragraphe 1.3 de ce même chapitre. □

Comme nous l'avons constaté dans le cas particulier du processus AR(1), il n'existe pas qu'une seule représentation d'un processus AR(p). En revanche une seule de ces représentations est canonique. Tout ceci est précisé dans la proposition suivante.

Proposition 4.5.
Soit $(\varepsilon_t)_{t\in\mathbb{Z}}$ un bruit blanc et $\Phi(B)$ un polynôme en B de degré p exactement dont toutes les racines sont de module différent de 1. Notons $z_i = 1/\lambda_i$, pour $i = 1, \dots, p$, les racines (non nécessairement distinctes) de $\Phi(z)$.
Soit $(X_t)_{t\in\mathbb{Z}}$ l'unique processus stationnaire AR(p) vérifiant l'équation

$$\Phi(B)X_t = \varepsilon_t, \text{ pour tout } t \in \mathbb{Z}.$$

- *Le processus $(X_t)_{t\in\mathbb{Z}}$ satisfait plusieurs représentations AR(p) différentes. Celles-ci sont obtenues en remplaçant certaines racines par leur inverse. Plus précisément, en considérant $\tilde{\Phi}(B)$ un nouveau polynôme en B, toujours de degré p, ayant pour racines celles de $\Phi(B)$ ou leurs inverses, le processus $(X_t)_{t\in\mathbb{Z}}$ admet également la représentation :*

$$\tilde{\Phi}(B)X_t = \eta_t, \text{ pour tout } t \in \mathbb{Z}.,$$

  où $(\eta_t)_{t\in\mathbb{Z}}$ est un nouveau bruit blanc.
- *Toutes les représentations ont même ordre.*
- *Une seule représentation est canonique, elle est obtenue en prenant le polynôme avec toutes les racines à l'extérieur du disque unité. Concrètement, la représentation canonique du processus AR(p) est donnée par :*

$$\tilde{\Phi}(B)X_t = \eta_t, \text{ pour tout } t \in \mathbb{Z},$$

où

$$\tilde{\Phi}(B) = \prod_{i:|z_i|>1} (I - \lambda_i B) \prod_{i:|z_i|<1} \left(I - \frac{1}{\lambda_i} B \right).$$

Exercice 4.5. *Preuve de cette proposition.*

Il suffit de montrer le dernier point. Le second est en effet évident et le premier point s'obtiendrait de manière tout à fait identique à ce que l'on va faire en faisant un changement de racines qui ne retient pas uniquement celles à l'extérieur du disque unité.

La démonstration est assez proche de ce que l'on a fait pour le processus AR(1) et utilise donc les densités spectrales.

- *Montrer que l'on a :*

$$f_\eta(\omega) = \frac{\left|\tilde{\Phi}(e^{-i\omega})\right|^2}{\left|\Phi(e^{-i\omega})\right|^2} f_\varepsilon(\omega) = \frac{\sigma^2}{2\pi} \frac{\left|\prod_{j:|z_j|<1}\left(1 - \frac{1}{\lambda_j} e^{-i\omega}\right)\right|^2}{\left|\prod_{j:|z_j|<1}\left(1 - \lambda_j e^{-i\omega}\right)\right|^2}.$$

- *On suppose dans un premier temps que toutes les racines de $\Phi(z)$ sont réelles. Montrer, en s'inspirant du raisonnement fait pour le AR(1), que le processus $(\eta_t)_{t\in\mathbb{Z}}$ est un bruit blanc.*
- *On va regarder ce qu'il se passe si au moins une racine, disons $z = 1/\lambda$, n'est pas réelle. Le polynôme $\Phi(z)$ étant à coefficient réel, son conjugué $\bar{z}$ est également racine. Considérer dans le ratio du premier point uniquement les termes correspondants à ces deux racines conjuguées et montrer que l'on a :*

$$\frac{\left|\left(1 - \frac{1}{\lambda} e^{-i\omega}\right)\left(1 - \frac{1}{\bar{\lambda}} e^{-i\omega}\right)\right|^2}{\left|\left(1 - \lambda e^{-i\omega}\right)\left(1 - \bar{\lambda} e^{-i\omega}\right)\right|^2} = \frac{1}{|\lambda|^4}.$$

- *En déduire que $(\eta_t)_{t\in\mathbb{Z}}$ est toujours un bruit blanc, y compris dans le cas où certaines racines ne sont pas réelles et qu'ainsi, avec ce nouveau polynôme $\tilde{\Phi}(B)$ obtenu en remplaçant racines à l'intérieur du disque unité par leur inverse, on a obtenu une nouvelle représentation AR(p) du processus $(X_t)_{t\in\mathbb{Z}}$.*
- *Reste à vérifier que l'on a ainsi obtenu la représentation canonique. Montrer l'égalité des histoires des processus $(X_t)_{t\in\mathbb{Z}}$ et $(\eta_t)_{t\in\mathbb{Z}}$, i.e.*

$$\mathcal{H}_{-\infty}^t(X) = \mathcal{H}_{-\infty}^t(\eta).$$

On dit dans ce cas que le processus est ***régulier****. Conclure.*

Proposition 4.6. *Si la représentation du processus AR(p)*

$$\Phi(B)X_t = \varepsilon_t, \text{ pour tout } t \in \mathbb{Z},$$

est canonique, le bruit blanc $(\varepsilon_t)_{t\in\mathbb{Z}}$ utilisé dans sa représentation est aussi le processus des innovations.

EXERCICE 4.6. *Démontrer cette Proposition.*

PROPOSITION 4.7. *Tout processus AR(p) de représentation canonique*

$$\Phi(B)X_t = \varepsilon_t, \text{ pour tout } t \in \mathbb{Z},$$

admet la représentation MA(∞), i.e.

$$X_t = \varepsilon_t + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-i}.$$

EXERCICE 4.7. *Démontrer cette Proposition.*

2.4. Liaisons temporelles.

On considère dans cette partie un processus $(X_t)_{t\in\mathbb{Z}}$ de représentation AR(p) canonique :

$$\Phi(B)X_t = \varepsilon_t, \text{ pour tout } t \in \mathbb{Z}.$$

2.4.1. *Autocovariances et autocorrélations simples.*

EXERCICE 4.8.

- *Montrer que l'on a*

$$\mathrm{Var}(X_t) = \sum_{i=1}^{p} \varphi_i \gamma_X(i) + \mathrm{Var}(\varepsilon_t)$$

puis que

$$\gamma_X(0) = \frac{\sigma^2}{1 - \sum_{i=1}^{p} \varphi_i \rho_X(i)}.$$

- *Montrer que l'on a de même :*

$$Cov(X_t, X_{t+h}) = \sum_{i=1}^{p} \varphi_i \gamma_X(h-i),$$

puis que

$$\rho_X(h) = \sum_{i=1}^{p} \varphi_i \rho_X(h-i), \tag{4}$$

pour $h \in \mathbb{N}^*$.

L'exercice précédent nous permet de constater que les autocorrélations d'un AR(p) sous sa représentation canonique vérifient une équation linéaire récurrente d'ordre p. Elles sont donc entièrement déterminées par la donnée des p premières et celles-ci vérifient l'équation matricielle dite de **Yule-Walker** :

$$\begin{aligned}
\begin{pmatrix} \rho_X(1) \\ \vdots \\ \rho_X(p) \end{pmatrix} &= R(p) \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{pmatrix} = \begin{pmatrix} 1 & \rho_X(1) & \cdots & \rho_X(p-1) \\ \rho_X(1) & 1 & \ddots & \rho_X(p-2) \\ \vdots & \ddots & \ddots & \vdots \\ \rho_X(p-1) & \cdots & \rho_X(1) & 1 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{pmatrix} \\
\text{et } \gamma_X(0) &= \frac{\sigma^2}{1 - \sum_{i=1}^p \varphi_i \rho_X(i)}.
\end{aligned}$$

Proposition 4.8 (**Autocorrélations simples d'un AR(p)**).
Les autocorrélations simples d'un processus AR(p) décroissent, de manière exponentielle ou sinusoïdale amortie, vers 0.

Exercice 4.9. *Démonstration de cette proposition.*

- *Montrer que le polynôme caractéristique de l'équation récurrente linéaire (4) vérifiée par les autocorrélations d'un AR(p) peut s'écrire sous la forme :*

$$r^p \Phi\left(\frac{1}{r}\right).$$

  Justifier au passage que cette expression a toujours un sens en précisant pourquoi les racines du polynôme caractéristiques sont toujours non nulles.
- *Montrer que la solution de l'équation (4) dans $\mathbb{C}$ est alors une combinaison linéaire de solutions de la forme*

$$(a_{i,0} + a_{i,1}h + \cdots + a_{i,k-1}h^{k-1})\lambda_i^h,$$

  où k est l'ordre de multiplicité de la racine $z_i = 1/\lambda_i$ de $\Phi(z)$.
- *Que dire de l'évolution des autocorrélations en fonction de h si toutes les racines sont réelles ?*
- *On suppose maintenant qu'au moins une racine, disons $z_i = 1/\lambda_i$ de $\Phi(z)$, n'est pas réelle. Montrer que les combinaisons linéaires des solutions réelles associées aux deux racines complexes conjuguées sont simusoïdales amorties.*

Exercice 4.10. *On considère un AR(1) de représentation canonique*

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

Vérifier sur ce cas précis la décroissance exponentielle des autocorrelations.

2.4.2. *Autocorrélations partielles.*

PROPOSITION 4.9 (**Autocorrélations partielles d'un AR(p)**).
Les autocorrélations partielles d'un processus AR(p) sont nulles à partir du rang $p+1$. Plus précisément, on a : $r_X(p) = \varphi_p \neq 0$ et $r_X(k) = 0$, pour $k > p$.

EXERCICE 4.11.

- *On rappelle que le coefficient $r_X(k)$ est donné par le coefficient $\alpha_k(k)$ dans la projection de X_t sur $\mathcal{H}_{t-k}^{t-1}$, i.e.*
$$EL(X_t|\mathcal{H}_{t-k}^{t-1}) = \alpha_1(k)X_{t-1} + \cdots + \alpha_k(k)X_{t-k}.$$
En déduire que $r_X(p) = \varphi_p \neq 0$.
- *En déduire également que $r_X(k) = 0$ dès que $k > p$.*

Remarques.

- La propriété $r_X(p) = \varphi_p$ n'est vraie que dans le cas d'une représentation canonique.
- On peut montrer que la propriété d'autocorrélations partielles nulles à partir d'un certain rang $p+1$ est caractéristique d'un AR(p).

3. Processus MA

3.1. Definition.

On introduit la définition des processus MA (Moving Average).

DÉFINITION 4.10. *On dit qu'un processus stationnaire $(X_t)_{t\in\mathbb{Z}}$ admet une représentation **Moyenne Mobile d'ordre** q **(noté MA(q))** s'il vérifie l'équation :*
$$X_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}$$
où $(\varepsilon_t)_{t\in\mathbb{Z}}$ est un bruit blanc $bb(0,\sigma^2)$ et $\theta_1, \ldots, \theta_q$ sont des coefficients réels, avec $\theta_q \neq 0$. Autrement dit, en utilisant le polynôme en B
$$\Theta(B) = I + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q,$$
un processus MA(q) vérifie l'équation :
$$X_t = \Theta(B)\varepsilon_t,$$
pour tout t dans $\mathbb{Z}$.

*Cette représentation est dite **canonique** si les racines du polynôme $\Theta(z)$ sont toutes à l'extérieur du disque unité.*

Remarques.

- On constate, qu'à la différence d'un processus AR, un processus MA est entièrement spécifié. À bruit blanc et paramètres $\theta_1, \ldots, \theta_q$ fixés, il ne correspond qu'un seul processus MA(q) qui est donc le filtrage linéaire du processus $(\varepsilon_t)_{t\in\mathbb{Z}}$ par le filtre
$$\Theta(B) = I + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q.$$
- Tout processus MA admet plusieurs représentations, il suffit de remplacer une ou plusieurs racines du polynôme $\Theta(z)$ par leur(s) inverse(s). Ce résultat se montre par une méthode similaire à celle utilisée pour les processus AR.
- Il n'existe qu'une seule représentation canonique, celle donc où toutes les racines du polynôme sont à l'extérieur du disque unité.
- **Un processus MA, quelque soit sa représentation, est stationnaire**, puisque obtenu par filtrage linéaire d'un bruit blanc, donc stationnaire.
- Un processus MA est toujours centré.

3.2. Ecriture AR(∞).

Proposition 4.11. *Soit $(X_t)_{t\in\mathbb{Z}}$ un processus MA(q) de représentation canonique*
$$X_t = \Theta(B)\varepsilon_t,$$
où $(\varepsilon_t)_{t\in\mathbb{Z}}$ est un bruit blanc $bb(0,\sigma^2)$.

Le processus des innovations correspond alors au bruit blanc $(\varepsilon_t)_{t\in\mathbb{Z}}$ de sa représentation canonique. De plus, le processus $(X_t)_{t\in\mathbb{Z}}$ possède une représentation AR(∞) :
$$\varepsilon_t = \sum_{i=0}^{+\infty} \pi_i X_{t-i},$$
avec $\pi_0 = 1$.

Exercice 4.12. *Démonstration de cette Proposition.*

- *Établir d'abord l'écriture AR(∞).*
- *En déduire que le processus $(X_t)_{t\in\mathbb{Z}}$ est régulier.*
- *Calculer alors $EL(X_t|\mathcal{H}_{-\infty}^{t-1}(X))$ et conclure en déterminant le processus des innovations.*

3.3. Liaisons temporelles.

3.3.1. *Autocovariances et autocorrélations simples.*

Exercice 4.13.

- *Montrer que l'on a :*
$$\mathrm{Var}(X_t) = \sigma^2(1 + \sum_{j=1}^{q} \theta_j^2).$$
- *Que dire de $\gamma_X(h)$ pour $h > q$?*
- *Calculer $\gamma_X(1)$ et $\gamma_X(2)$.*

- *En déduire une expression générale de $\gamma_X(h)$ pour $h \leq q$.*

On a démontré la proposition suivante.

PROPOSITION 4.12 (**Autocorrélations simples d'un MA(q)**).

Les autocorrélations simples d'un processus MA(q) sont nulles à partir du rang $q+1$. Plus précisément, on a : $\rho_X(q) \neq 0$ et $\rho_X(h) = 0$, pour $h > q$ et

$$\rho_X(h) = \frac{\theta_h + \sum_{i=1}^{q-h} \theta_i \theta_{i+h}}{1 + \sum_{j=1}^{q} \theta_j^2}.$$

Remarque. On peut montrer que la propriété d'autocorrélations simples nulles à partir d'un certain rang $q+1$ est caractéristique d'un MA(q).

3.3.2. *Autocorrélations partielles.*

On pourrait démontrer le résultat suivant.

PROPOSITION 4.13 (**Autocorrélations partielles d'un MA(q)**).

Les autocorrélations partielles d'un processus MA(q) sont solutions d'une équation linéaire récurrente d'ordre q. Elles décroissent, de manière exponentielle ou sinusoïdale amortie, vers 0.

4. Processus ARMA

Nous allons maintenant introduire un modèle de processus stationnaire comportant une partie AR et une partie MA. C'est pourquoi il porte le nom de processus ARMA (AutoRegressive Moving Average). Les processus ARMA sont très importants en pratique car on peut montrer que tout processus stationnaire peut être approché par un processus ARMA.

4.1. Définition.

DÉFINITION 4.14. *On dit qu'un processus stationnaire $(X_t)_{t\in\mathbb{Z}}$ admet une représentation **ARMA**(p,q) s'il vérifie l'équation :*

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

où $(\varepsilon_t)_{t\in\mathbb{Z}}$ est un bruit blanc $bb(0,\sigma^2)$ et $\varphi_1,\ldots,\varphi_p,\theta_1,\ldots,\theta_q$ sont des coefficients réels, avec $\varphi_p \neq 0$ ainsi que $\theta_q \neq 0$. Autrement dit, en utilisant les polynômes en B

$$\begin{aligned} \Phi(B) &= I - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p \\ \textit{et } \Theta(B) &= I + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q, \end{aligned}$$

un processus ARMA(p,q) vérifie l'équation :

$$\Phi(B) X_t = \Theta(B) \varepsilon_t,$$

pour tout t dans $\mathbb{Z}$.

Naturellement un processus AR(p) est un ARMA($p,0$) et un processus MA(q) est un ARMA($0,q$).

Définition 4.15. *La représentation d'un processus ARMA(p,q)*

$$\Phi(B)X_t = \Theta(B)\varepsilon_t,$$

est dite

- ***minimale*** *si les polynômes $\Phi(z)$ et $\Theta(z)$ n'ont pas de racine commune ;*
- ***causale*** *si le polynôme $\Phi(z)$ a toutes ses racines à l'extérieur du disque unité ;*
- ***inversible*** *si le polynôme $\Theta(z)$ a toutes ses racines à l'extérieur du disque unité ;*
- ***canonique*** *si elle est causale et inversible.*

Disons quelques mots sur le critère de représentation minimale. Si la représentation n'est pas minimale, c'est à dire si les polynômes $\Phi(z)$ et $\Theta(z)$ ont une ou des racine(s) commune(s), alors deux cas peuvent se présenter.

- Soit aucune de ces racines communes n'est sur le cercle unité. Dans ce cas, le processus $(X_t)_{t\in\mathbb{Z}}$ admet aussi la représentation
$$\tilde{\Phi}(B)X_t = \tilde{\Theta}(B)\varepsilon_t,$$
pour tout t dans $\mathbb{Z}$, où les nouveaux polynômes sont obtenus à partir des précédants en enlevant les racines communes.
- Si au moins une des racines communes est sur le cercle unité, alors il peut y avoir plus d'un unique processus stationnaire vérifiant l'équation.

C'est pourquoi, dans la suite, on ne considèrera que des représentations minimales.

4.2. Écritures MA(∞) et AR(∞).

Proposition 4.16 (**Écriture MA(∞)**). *Soit $(X_t)_{t\in\mathbb{Z}}$ un processus ARMA(p,q) de représentation minimale et causale. Il admet alors la représentation MA(∞)*

$$X_t = \Phi^{-1}(B)\Theta(B)\varepsilon_t = \varepsilon_t + \sum_{i=1}^{+\infty}\psi_i\varepsilon_{t-i},$$

où les coefficients $(\psi_i)_{i\in N}$ forment une famille absolument sommable et vérifient l'équation de récurrence linéaire :

$$\psi_i - \sum_{j=1}^{p}\varphi_j\psi_{i-j} = \theta_i, \text{ pour } i\in\mathbb{N},$$

avec $\psi_i = 0$, pour $i<0$, $\psi_0 = 1$, $\theta_0 = 1$ et $\theta_i = 0$ pour $i > q$.

Exercice 4.14. *Démonstration de cette proposition.*

- *Justifier que l'on peut écrire X_t, pour tout $t\in\mathbb{Z}$, en fonction du passé du bruit blanc sous la forme*
$$X_t = \varepsilon_t + \sum_{i=1}^{+\infty}\psi_i\varepsilon_{t-i}.$$

- *En injectant cette écriture MA(∞) de $(X_t)_{t\in\mathbb{Z}}$ dans l'équation*
$$\Phi(B)X_t = \Theta(B)\varepsilon_t,$$
montrer que les coefficients doivent vérifier les équations données dans la Proposition pour $i = 1, 2, 3$ et ainsi de suite.

Proposition 4.17 (**Écriture AR(∞)**).
Soit $(X_t)_{t\in\mathbb{Z}}$ un processus ARMA(p,q) de représentation minimale et inversible. Il admet alors la représentation AR(∞)
$$\varepsilon_t = \Theta^{-1}(B)\Phi(B)X_t = X_t + \sum_{i=1}^{+\infty} \pi_i X_{t-i},$$
où les coefficients $(\pi_i)_{i\in N}$ forment une famille absolument sommable et vérifient l'équation de récurrence linéaire :
$$\pi_i + \sum_{j=1}^{q} \theta_j \pi_{i-j} = -\varphi_i, \text{ pour } i \in \mathbb{N},$$
avec $\pi_i = 0$, pour $i < 0$, $\varphi_0 = -1$ et $\varphi_i = 0$ pour $i > p$.

4.3. Liaisons temporelles.

Exercice 4.15.
Calculons les autocorrélations simples d'un processus ARMA(p, q). On peut obtenir deux expressions différentes, l'une en utilisant l'écriture ARMA(p, q), l'autre en utilisant l'écriture MA(∞).

- *En utilisant l'écriture ARMA(p, q), montrer que l'on a*
$$\gamma_X(h) - \varphi_1\gamma_X(h-1) - \cdots - \varphi_p\gamma_X(h-p)$$
$$= Cov(\varepsilon_{t+h} + \theta_1\varepsilon_{t+h-1} + \theta_2\varepsilon_{t+h-2} + \cdots + \theta_q\varepsilon_{t+h-q}, \varepsilon_t + \sum_{i=1}^{+\infty}\psi_i\varepsilon_{t-i}).$$
- *En déduire, pour $h > q$, la relation de récurrence :*
$$\gamma_X(h) - \varphi_1\gamma_X(h-1) - \cdots - \varphi_p\gamma_X(h-p) = 0.$$
- *Donner l'expression de la relation de récurrence pour les valeurs $0 \le h \le q$.*
- *Quelle expression de $\gamma_X(h)$ obtient-on en utilisant cette fois l'écriture MA(∞) du processus $(X_t)_{t\in\mathbb{Z}}$.*

L'expression obtenue ci-dessus, sous forme de relation de récurrence, permet de montrer que les autocorrélations simples décroissent de manière exponentielle ou sinusoïdale amortie vers 0 avec h. On peut montrer le même genre de résultat pour les autocorrélations partielles.

On constate aussi, qu'à la différence des cas particuliers des processus AR(p) ou MA(q), il n'existe pas de caractérisation aisée pour les modèles ARMA(p, q). Les autocorrélations simples ou partielles ne s'annulent pas à partir d'un certain rang.

Proposition 4.18 (**Méthode du coin**).
Soient les matrices suivantes et leurs déterminants respectifs, pour $(i,j) \in \mathbb{N}^2$:

$$\Omega_{i,j} = \begin{bmatrix} \rho_X(i) & \rho_X(i-1) & \dots & \dots & \rho_X(i-j+1) \\ \rho_X(i-1) & \rho_X(i) & \rho_X(i-1) & \dots & \rho_X(i-j) \\ \rho_X(i-2) & \rho_X(i-1) & \rho_X(i) & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \rho_X(i-1) \\ \rho_X(i-j+1) & \rho_X(i-j) & \dots & \rho_X(i-1) & \rho_X(i) \end{bmatrix},$$

$$\Delta_{i,j} = \det(\Omega_{i,j}).$$

Pour un processus $ARMA(p,q)$, on a :

- $\forall (i,j) \in \mathbb{N}^2, i > q, j > p : \Delta_{i,j} = 0,$
- $\forall (i,j) \in \mathbb{N}^2, i \leq q : \Delta_{i,p} \neq 0,$
- $\forall (i,j) \in \mathbb{N}^2, j \leq p : \Delta_{q,j} \neq 0.$

On peut visualiser ce résultat sous forme matricielle en représentant la matrice (pour k assez grand) $M = (\Delta_{i,j})_{(i,j)\in\{1,\dots,k\}^2}$ et faire ainsi apparaître un coin :

$$M = \left[\begin{array}{ccc|ccc} \Delta_{1,1} & \dots & \Delta_{1,p} & \Delta_{1,p+1} & \dots & \Delta_{1,k} \\ \vdots & & \vdots & \vdots & & \vdots \\ \Delta_{q,1} & \dots & \Delta_{q,p} & \Delta_{q,p+1} & \dots & \Delta_{q,k} \\ \hline \Delta_{q+1,1} & \dots & \Delta_{q+1,p} & & & \\ \vdots & & \vdots & & 0 & \\ \Delta_{k,1} & \dots & \Delta_{k,p} & & & \end{array}\right].$$

5. Processus ARIMA

Les processus étant dans la pratique rarement stationnaires, on a introduit une généralisation des processus ARMA vus précédemment de manière à les étendre à des processus non stationnaires. L'idée générale, essentiellement conçue pour les processus non stationnaires à tendance polynomiale, est de différencier suffisamment le processus initial afin d'obtenir un processus sans tendance et sur la partie différenciée appliquer un modèle ARMA. Cette classe de modèle est connue sous le nom de processus ARIMA (AutoRegressive Integrated Moving Average).

Définition 4.19.
*On dit qu'un processus $(X_t)_{t\in\mathbb{Z}}$ admet une représentation **ARIMA**(p,d,q) s'il vérifie l'équation :*

$$\Phi(B)\nabla^d X_t = \Theta(B)\varepsilon_t,$$

pour tout t dans $\mathbb{Z}$, où

$$\begin{aligned} \nabla^d &= (I-B)^d \\ \Phi(B) &= I - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p \\ \Theta(B) &= I + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q, \end{aligned}$$

et où $(\varepsilon_t)_{t\in\mathbb{Z}}$ *est un bruit blanc* $bb(0,\sigma^2)$. *Les coefficients* $\varphi_1,\ldots,\varphi_p,\theta_1,\ldots,\theta_q$ *sont réels et tels que* $\varphi_p \neq 0$ *ainsi que* $\theta_q \neq 0$.

On peut montrer que, dans le cas d'un processus ARIMA(p,d,q), le processus $\nabla^d X_t$ est asympotiquement, au sens quand $t \to +\infty$, un processus ARMA(p,q). On peut aussi établir des représentations AR∞) et MA(∞) pour les processus ARIMA.

6. Travaux Dirigés

Exercice TD 4.1

Soit $(X_t)_{t\in\mathbb{Z}}$ le processus $MA(1)$:

$$X_t = \varepsilon_t - 2\varepsilon_{t-1},$$

où $(\varepsilon_t)_{t\in\mathbb{Z}}$ est un bruit blanc fort de variance 1 tel que :

$$\forall t \in \mathbb{Z} : \mathbb{E}\left(\varepsilon_t^3\right) = c < +\infty.$$

Soit $(\eta_t)_{t\in\mathbb{Z}}$ le processus tel que :

$$X_t = \eta_t - \frac{1}{2}\eta_{t-1}.$$

(1) Vérifier que $(\eta_t)_{t\in\mathbb{Z}}$ est un bruit blanc dont on calculera la variance.
(2) Montrer que $(\eta_t)_{t\in\mathbb{Z}}$ est l'innovation du processus $(X_t)_{t\in\mathbb{Z}}$.
(3) Montrer que :

$$\eta_t = \varepsilon_t - \sum_{i=1}^{+\infty} \frac{3}{2^i}\varepsilon_{t-i}.$$

(4) Calculer $\mathbb{E}\left(\eta_1^2\eta_2\right)$.
(5) En déduire que η_1 et η_2 ne sont pas des v.a.r. indépendantes ?

Exercice TD 4.2

Soit $(\varepsilon_t)_{t\in\mathbb{Z}}$ une suite de v.a.r. indépendantes de loi $\mathcal{N}(0,1)$. Soit $(X_t)_{t\in\mathbb{Z}}$ un processus vérifiant :

$$X_t = \varepsilon_t - \theta\varepsilon_{t-1}$$

où $\theta \neq 0$ et $|\theta| < 1$.
Soit $(Y_t)_{t\in\mathbb{Z}}$ un processus vérifiant :

$$Y_t = \begin{cases} 1 & \text{si } X_t > 0 \\ 0 & \text{si } X_t \leq 0 \end{cases}.$$

(1) (a) Quelle est la loi des v.a.r. Y_t pour $t \in \mathbb{Z}$?
    (b) Quelle est la loi conjointe du vecteur aléatoire (X_t, X_{t-1}) ?
(2) Montrer que le processus $(Y_t)_{t\in\mathbb{Z}}$ est stationnaire et donner sa fonction d'autocovariance.
(3) Déduire de ce qui précède que le processus $(Y_t)_{t\in\mathbb{Z}}$ admet une représentation du type :

$$Y_t = \mu + u_t - \alpha u_{t-1}$$

où μ est une constante à préciser et $(u_t)_{t\in\mathbb{Z}}$ est un bruit blanc de variance σ^2.
(4) Proposer une méthode pour calculer α et σ^2.

Indication

On admettra que :

$$P\left(X_t > 0, X_{t-1} > 0\right) = \frac{1}{2\pi} \arctan\left(\frac{\theta}{\sqrt{1+\theta^2+\theta^4}}\right)$$

et que :

$$P\left(X_t > 0, X_{t-1} > 0\right) \leq \frac{1}{8}.$$