

الجمهورية الجزائرية الديمقراطية الشعبية

MINISTRE DE L'ENSEIGNEMENT
SUPERIEUR ET DE LA RECHERCHE SIENTIFIQUE
UNIVERSITE 8 MAI 1945
GUELMA
Faculté des Mathématiques et de l'Informatique
et des Sciences de la Matière
Département des Mathématiques



وزارة التعليم العالي
والبحث العلمي
جامعة 8 ماي 1945
قلمة
كلية الرياضيات والإعلام الآلي
وعلوم المادة
قسم الرياضيات

Polycopié du cours :

Statistique des processus

Niveau deuxième année
Master Probabilités et Applications

Auteur :

Dr: Rabah DEBBAR

Guelma, 2016



Statistique des processus

Dr: DEBBAR.Rabah

Master Probabilités et Applications
deuxième année

Université 08 mai 1945-Guelma

12 janvier 2017

*Je dédie ce polycopié
A mes chers parents, symbole de sacrifice, de tendresse et d'amour..*

Notes au lecteur

Les sections, les définitions et les équations sont numérotées par deux chiffres, le premier indique le chapitre et le deuxième indique la position dans le chapitre en question.

Exemple :

Section **2.1** désigne la **première** section du Chapitre **2**.

les théorèmes et les lemmes sont numérotés séparément en utilisant **2** chiffres, le premier chiffre indique le chapitre, le deuxième il désigne la position dans le Chapitre en question.

Exemple :

Lemme **2.1** indique le **premier** lemme dans la **deuxième** Chapitre.

Résumé

Objectifs de l'enseignement Il s'agit d'une initiation à la théorie des processus stochastiques et à l'inférence statistique relative à ces processus.

Connaissances préalables recommandées Connaissances solides en probabilités et éventuellement notions élémentaires de statistiques.

Table des matières

Introduction	v
1 Processus à temps discret et continu	1
1.1 Variables de carré intégrable	1
1.2 Processus stochastiques	5
1.3 Opérateurs avance et retard	7
1.3.1 Inversion de $I - \lambda B$	9
1.3.2 Inversion d'un polynôme en B	10
1.4 Processus stationnaires du second ordre	11
1.4.1 Définitions	11
1.4.2 Fonctions d'autocovariance et d'autocorrélation	15
1.4.3 Estimations des fonctions caractéristiques	21
2 Processus ARMA	24
2.1 Processus linéaires	25
2.2 Processus moyennes mobiles	30
2.2.1 Définition d'un MA (q)	30
2.2.2 Forme autorégressive infinie d'un MA (q)	32
2.2.3 Bruit blanc d'innovation d'un MA (q)	33

2.2.4	Fonctions caractéristiques et propriétés d'un MA (q)	34
2.2.5	Prédiction d'un MA (q)	37
2.2.6	Erreur et intervalle de confiance pour la prévision d'un MA (q)	38
2.3	Processus autorégressifs	38
2.3.1	Définition d'un AR (p)	38
2.3.2	Ecriture moyenne mobile infinie d'un AR (p)	39
2.3.3	Bruit blanc d'innovation d'un AR (p)	40
2.3.4	Fonctions caractéristiques d'un AR (p)	41
2.3.5	Prédiction d'un AR (p)	45
2.3.6	Equations de Yule-Walker d'un AR (p)	47
2.4	Les processus mixtes ARMA	50
2.4.1	Définition d'un ARMA(p, q)	51
2.4.2	Expression d'un ARMA(p, q)	52
2.4.3	Bruit blanc d'innovation d'un ARMA(p, q)	54
2.4.4	Autocovariances et propriétés d'un ARMA(p, q)	55
3	Processus de Poisson	57
3.1	La loi de Poisson	57
3.2	La loi exponentielle	59
3.3	Processus de comptage	59
3.4	Processus de Poisson	60
3.5	Quelques propriétés du processus de Poisson	62
3.6	Processus de Poisson marqué	62
3.7	Processus de Poisson composé	63
3.8	Processus Markovien de saut	64

4	Processus de Wiener	65
5	Intégrale d'Itô et Processus de diffusion	69
5.1	Processus de diffusion	69
5.1.1	Espace mesurable.	69
5.1.2	Semi-groupe de Feller.	71
5.1.3	Processus de diffusion.	72
6	Estimation paramétrique des processus stationnaires	73
6.1	L'approche de Box et Jenkins	73
6.1.1	Pré-sélection de modèles	75
6.1.2	Estimation des paramètres	76
6.1.3	Sélection de modèles	78
6.1.4	Test sur les résidus	79
7	Prévision	81
7.1	Prévision des modèles ARMA (p, q)	81
7.1.1	Formule déduite de la forme autorégressive	81
7.1.2	Formule déduite de la forme moyenne mobile	83
7.1.3	Remarques et propriétés	84
	Bibliographie	85

Introduction

Le premier objectif de ce cours est de modéliser la partie aléatoire d'une série temporelle. Dans le cours du premier semestre, nous avons vu qu'une série temporelle X peut s'écrire sous la forme

$$X_t = Z_t + S_t + \epsilon_t;$$

avec Z_t et S_t des séries déterministes représentant respectivement la tendance et la saisonnalité et ϵ_t une série aléatoire représentant le résidu ou bruit. Dans le cadre du cours,

- nous avons vu comment isoler les parties déterministes.
- nous avons proposé des techniques pour les estimer.
- nous avons supposé que la série ϵ_t était un bruit blanc.
- enfin nous vérifions qu'une fois tendance et saisonnalité supprimées, le résidu était bien un bruit blanc.

Nous allons nous intéresser dans ce cours à la partie aléatoire (ϵ_t) de la série temporelle et travailler dans un cadre plus général que le précédent : nous ne supposons plus que (ϵ_t) est un bruit blanc mais qu'il est seulement stationnaire ; ce qui signifie grossièrement que sa moyenne, sa variance et son autocovariance sont constantes au cours du temps. A noter que le seul fait de supprimer la tendance et la saisonnalité ne rend pas nécessairement la série résiduelle stationnaire puisque cela n'affecte pas la variance et l'autocovariance qui

doivent être constantes pour un processus stationnaire. Nous proposerons des techniques de modélisation de ce type de processus. Il faudra choisir le modèle le plus simple possible avec le nombre le plus petit de paramètres possibles. Cela nous amènera à considérer tout particulièrement une famille de processus linéaires très couramment employée, les processus ARMA. Nous verrons ensuite comment faire de la prédiction à partir de ces séries.

En pratique, les séries temporelles résiduelles ne sont pas nécessairement stationnaires. Un prétraitement est alors nécessaire pour supprimer la tendance et la saisonnalité d'une part comme usuellement mais aussi pour "stationnariser" la série résiduelle. Par exemple, on peut supposer seulement que (ϵ_t) est à accroissements stationnaires i.e. que la loi de $(\epsilon_{t+h} - \epsilon_t)_{t,h}$ ne dépend pas de t . Les processus ARIMA et SARIMA mais aussi les processus de Poisson et le mouvement brownien sont des exemples de processus à accroissements stationnaires. On abordera dans ce cours le cas des processus ARIMA et SARIMA. Une fois la série "stationnarisée" analysée, et les valeurs futures prédites, il sera ensuite nécessaire de revenir à la série initiale.

Nous terminons ce cours avec des références bibliographiques.

Chapitre 1

Processus à temps discret et continu

L'objectif de ce chapitre est de rappeler l'essentiel des notions et résultats utilisés tout au long de ce travail. Pour faciliter l'approche et la lecture de ce chapitre nous avons omis volontiers certains détails et présentons-nous des versions particulières de certains résultats. Pour les lecteurs désireux plus de détails, des références à la littérature seront systématiquement données.

1.1 Variables de carré intégrable

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et X une application mesurable de Ω vers \mathbb{R} .

Définition 1.1. *On dit que X est une variable aléatoire réelle (v.a.r.) de **carré intégrable** ou d'ordre 2 si et seulement si $E(X^2) < \infty$.*

Notations. On note $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ (ou plus simplement \mathcal{L}^2 lorsqu'il n'y a pas d'ambiguïté) l'ensemble des v.a.r. de carré intégrable et $L^2(\Omega, \mathcal{A}, \mathbb{P})$ (ou L^2) l'ensemble des classes

d'équivalence pour l'égalité \mathbb{P} -ps de v.a.r. de $\mathcal{L}^2 : L^2$ est ainsi constitué des classes d'équivalence de v.a.r. égales \mathbb{P} -ps. On peut montrer que l'application

$$L^2 \times L^2 \longrightarrow \mathbb{R}$$

$$(X, Y) \longmapsto \mathbb{E}(XY)$$

est un produit scalaire que L^2 . La norme associée à ce produit scalaire, appelée norme L^2 , est définie par

$$\|X\|_2^2 = \mathbb{E}(X^2).$$

L'ensemble L^2 muni de cette norme est un espace vectoriel réel normé.

Définition 1.2. 1. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. de L^2 et $X \in L^2$. On dit que X_n **converge vers X dans L^2** si et seulement si

$$\lim_{n \rightarrow \infty} \|X_n - X\|_2^2 = \lim_{n \rightarrow \infty} \mathbb{E}[(X_n - X)^2] = 0$$

On note alors

$$X_n \xrightarrow{L^2} X$$

2. Deux variables aléatoires X et Y de L^2 sont **orthogonales** si et seulement si $\mathbb{E}(XY) = 0$. On note alors

$$X \perp Y$$

Proposition 1.1. L'ensemble L^2 muni du produit scalaire défini ci-dessus est un espace de Hilbert (espace vectoriel muni d'un produit scalaire complet).

Rappel : 1. Une suite $(X_n)_{n \in \mathbb{N}}$ est dite **Cauchy** dans L^2 si et seulement si

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N}, \forall n, m \in \mathbb{N}, \|X_n - X_m\|_2 < \varepsilon$$

2. Dire qu'un espace Ω est complet signifie que toute suite de Cauchy de L^2 converge dans L^2 : soit $(X_n)_{n \in \mathbb{N}}$ une suite de Cauchy dans L^2 alors il existe $X \in L^2$ telle que $X_n \xrightarrow{L^2} X$.

L'espérance conditionnelle ou régression

Considérons maintenant deux v.a.r. X et Y de L^2 . On note \mathcal{B}_X la tribu engendrée par X : \mathcal{B}_X est une tribu de parties de Ω , sous-tribu de \mathcal{A} , définie par

$$\mathcal{B}_X = \{X^{-1}(B) / B \in \mathcal{B}_{\mathbb{R}}\}.$$

On définit ainsi un nouvel espace probabilisé $(\Omega, \mathcal{B}, \mathbb{P})$, \mathbb{P} étant restreinte à la tribu \mathcal{B}_X . On note L_X^2 l'espace des v.a.r. d'ordre 2 définies sur ce nouvel espace.

Proposition 1.2. *L'espace L_X^2 est un sous-espace vectoriel fermé de L^2 .*

Remarque 1.1. *On montre que*

$$L_X^2 = \{\phi(X), \phi \text{ mesurable de } \mathbb{R} \text{ dans } \mathbb{R} \text{ et } \mathbb{E}(\phi^2(X)) < \infty\}$$

Ainsi L_X^2 est l'espace vectoriel constitué des fonctions mesurables de la variable X .

Définition 1.3. *L'espérance conditionnelle de $Y \in L^2$ sachant X est la variable aléatoire, notée $\mathbb{E}(Y|X)$, égale à la projection de Y sur L_X^2 .*

Remarque 1.2. 1. L'existence de la projection est assurée par le fait que L_X^2 est fermé.

2. D'après la définition (1.3), on a

$$\mathbb{E}(Y|X) = \arg \min_{Z \in L_X^2} \mathbb{E}[(Y - Z)^2]$$

3. D'après la remarque (1.1), il existe une application r de \mathbb{R} dans \mathbb{R} telle que

$\mathbb{E}(Y|X) = r(X)$. La fonction $r(x) = \mathbb{E}(Y|X = x)$ est appelée **régression** de Y en X .

4. On peut étendre la définition de l'espérance conditionnelle sachant X à toute variable aléatoire Y (i.e. pas nécessairement L^2). Dans ce cas bien évidemment, on ne la définira pas comme une projection.

La régression affine

Considérons maintenant une famille $(X_i)_{i \in I}$ (I étant fini ou dénombrable) de v.a.r. de carré intégrable.

Soit $L_{((X_i)_{i \in I})}$ le plus petit sous-espace vectoriel fermé de L^2 contenant toutes les combinaisons affines des variables X_i (et donc leurs limites au sens L^2). L'existence dans L^2 d'une notion d'orthogonalité et d'une notion de convergence permet d'introduire la notion de projection orthogonale.

Définition 1.4. On appelle **régression affine** de Y sur $(X_i, i \in I)$ la projection orthogonale de Y sur $L_{((X_i)_{i \in I})}$ notée Y^* .

Etant donnée la variable Y de carré intégrable, la variable Y^* est la meilleure approximation de Y au sens de L^2 par un élément de $L_{((X_i)_{i \in I})}$ puisqu'elle vérifie

$$\mathbb{E}[(Y - Y^*)^2] = \min_{Z \in L_{((X_i)_{i \in I})}} \mathbb{E}[(Y - Z)^2]$$

De plus, on a

$$Y = Y^* + R = a_0 + \sum_{j \in J} a_j X_j + R,$$

avec $R \perp X_i, i \in I$.

Remarque 1.3. Lorsqu'on considère deux v.a.r. X et Y la régression affine de Y sur X est donnée par

$$\frac{\text{Cov}(X, Y)}{\text{Var}(X)} X + \left(\mathbb{E}(Y) - \frac{\text{Cov}(X, Y)}{\text{Var}(X)} \mathbb{E}(X) \right)$$

Lien entre espérance conditionnelle et régression affine

Propriété 1.1. Soit $X = (X_1, \dots, X_n)$. Si (X, Y) est un vecteur gaussien, la régression affine de Y sur $X = (X_1, \dots, X_n)$ coïncide avec la projection orthogonale de Y sur L_X^2 :

$$Y^* = \mathbb{E}(Y|X)$$

1.2 Processus stochastiques

Définition 1.5. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Un **processus stochastique** X est une famille de variables aléatoires réelles $(X_t)_{t \in \Theta}$, où $\Theta \subset \mathbb{R}$ est appelé **l'espace des temps**. Si $\Theta \subset \mathbb{Z}$, le processus est dit à temps **dicret**. Si est un intervalle de \mathbb{R} , le processus est dit à temps **continu**.

Remarque 1.4. 1. La définition précédente se généralise à un ensemble d'indices Θ quelconque et à des variables aléatoires à valeurs dans $(\hat{\Omega}, \hat{\mathcal{A}})$, espace probabilisable quelconque.

Si Ω est fini ou dénombrable, le processus est dit à **valeurs discrètes**. Si $\Omega = \mathbb{R}^d$, $d \geq 2$, le processus est dit **multidimensionnel**. Dans la suite, nous ne considèrerons que des processus à temps discret, réels unidimensionnels.

2. La loi du processus $(X_t)_{t \in \mathbb{Z}}$ est caractérisée par les lois de toutes les sous-familles finies X_{t_1}, \dots, X_{t_n} , $n \in \mathbb{N}^*$, $t_1, \dots, t_n \in \mathbb{Z}$.

Une série chronologique est l'observation à T instants d'un processus stochastique.

Exemple 1.1. Bruits blancs. On appelle **bruit blanc faible** tout processus $\epsilon = (\epsilon_t)_{t \in \mathbb{Z}}$ tel que

$$\mathbb{E}(\epsilon_t) = 0 \quad \text{et} \quad \mathbb{E}(\epsilon_t \epsilon_{t'}) = \sigma^2 \delta_{tt'}, \quad t, t' \in \mathbb{Z}$$

où $\delta_{tt'}$ est le symbole de Kronecker

$$\delta_{tt'} = \begin{cases} 1 & \text{si } t = t' \\ 0 & \text{si } t \neq t' \end{cases}$$

On dit qu'il est **fort** lorsque de plus les variables ϵ_t sont indépendantes. On appelle **bruit blanc gaussien** tout bruit blanc fort ϵ tel que $\forall t, \epsilon_t \sim \mathcal{N}(0, \sigma^2)$. Dans les deux cas, la variance commune σ^2 des ϵ_t est appelée variance du bruit blanc.

Remarquons que si ϵ est un bruit blanc alors les variables ϵ_t sont non corrélées. Si de plus, ϵ est un bruit blanc gaussien, elles sont indépendantes.

Exemple 1.2. Processus gaussien. On appelle **processus gaussien** tout processus $X = (X_t)_{t \in \mathbb{Z}}$ tel que $\forall k \in \mathbb{N}$, $t_1, \dots, t_k \in \mathbb{Z}$, $a_1, \dots, a_k \in \mathbb{R}$,

$$a_1 X_{t_1} + \dots + a_k X_{t_k}$$

suit une loi normale.

1.3 Opérateurs avance et retard

Définition 1.6. On appelle opérateur retard l'opérateur B qui à tout processus $(X_t)_{t \in \mathbb{Z}}$ associe le processus $(Y_t)_{t \in \mathbb{Z}}$ défini par

$$\forall t \in \mathbb{Z}, \quad Y_t = BX_t = X_{t-1}$$

Remarque 1.5. - L'opérateur B est linéaire et inversible. Son inverse $B^{-1} = F$ est défini par

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

L'opérateur F est appelé opérateur avance. - Si on compose B avec lui-même on obtient $B^2 = B \circ B$ tel que

$$\forall t \in \mathbb{Z}, \quad B^2 X_t = X_{t-2}$$

On peut itérer cette application et définir par récurrence

$$B^k X_t = X_{t-k}, \quad k \in \mathbb{N}$$

. Par convention, B^0 est l'opérateur identité I .

De manière générale, on peut combiner de façon linéaire ces différentes puissances pour construire un nouvel opérateur qui s'écrit sous la forme d'un polynôme en B : étant donné $n \in \mathbb{N}^*$ et $n + 1$ réels a_0, \dots, a_n on définit un polynôme en B noté $P(B) = a_0 I + a_1 B + \dots + a_n B^n$ et défini par

$$\forall t \in \mathbb{Z}, \quad P(B)X_t = a_0 X_t + a_1 X_{t-1} + \dots + a_n X_{t-n}.$$

Par extension, en faisant tendre n vers l'infini, on définit un nouvel opérateur s'écrivant comme une série en B

$$\Psi(B) = \sum_{i=0}^{+\infty} \psi_i B^i;$$

cet opérateur est défini sous réserve que la série de terme général ψ_i est absolument convergente (i.e. $\sum |\psi_i| < \infty$). Il transforme le processus (X_t) en

$$Y_t = \Psi(B) X_t = \sum_{i=0}^{+\infty} \psi_i X_{t-i}.$$

Ces opérateurs en B se comportent comme les séries entières d'une variable réelle.

De la même manière, on définit les puissances de l'opérateur F , les polynômes en F ainsi que les séries en F .

- L'opérateur B constitue l'exemple de base d'un filtre. Un autre exemple de filtre est le filtre identité I : c'est le filtre qui ne fait rien : $I(X_t) = X_t \quad \forall t$. La notion de filtre A est très vaste : d'une manière générale, il s'agit d'un opérateur défini sur un certain espace vectoriel de séries temporelles et satisfaisant des conditions telles que

- la linéarité : $A(X + Y) = A(X) + A(Y)$;
- l'invariance dans le temps : $A \circ B(X) = B \circ A(X)$.

Un autre exemple de filtre a été déjà vu : les moyennes mobiles.

Proposition 1.3. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus du second ordre.

1. Si $\mathbb{E}(X_t) = at + b$, alors $\mathbb{E}((I - B)X_t) = a$
2. Si $\mathbb{E}(X_t) = \sum_{i=0}^k a_i t^i$, alors $\mathbb{E}((I - B)^k X_t) = c$, où c est une cste indépendante de t .

Exercice Démontrer les propriétés de la proposition précédente.

Remarque 1.6. L'opérateur $I - B$, noté Δ , est appelé **opérateur de différenciation**. La proposition précédente montre que si on a une série chronologique ayant une tendance polynomiale, on peut annuler celle-ci par différenciations successives du processus, le nombre de différenciation étant égal au degré du polynôme de la tendance.

Sous certaines conditions, l'opérateur série en $B, \Psi(B)$, peut être inversible, c'est à dire que l'on peut trouver un autre opérateur $\pi(B)$ tel que

$$\pi(B) \circ \Psi(B) = \Psi(B) \circ \pi(B) = Id.$$

Cet opérateur inverse de $\Psi(B)$ sera noté indifféremment

$$\Psi^{-1}(B) \quad \text{ou} \quad \frac{1}{\Psi(B)}$$

Par exemple, nous allons étudier dans les deux sections suivantes l'inversion de $I - \lambda B$ et celle d'un polynôme en B .

1.3.1 Inversion de $I - \lambda B$

Proposition 1.4. L'application $I - \lambda B$ est inversible si et seulement si $|\lambda| \neq 1$. Son inverse est alors donnée par

$$\begin{cases} \sum_{i=0}^{+\infty} \lambda^i B^i & \text{pour } |\lambda| < 1 \\ -\sum_{i=1}^{+\infty} \frac{1}{\lambda^i} B^i & \text{pour } |\lambda| > 1 \end{cases} = \begin{cases} \sum_{i=0}^{+\infty} \lambda^i B^i & \text{pour } |\lambda| < 1 \\ -\sum_{i=-1}^{-\infty} \lambda^i B^i & \text{pour } |\lambda| > 1 \end{cases}$$

Exercice 1) Cas où $|\lambda| < 1$:

- Montrer que la série réelle $(a_i)_{i \in \mathbb{Z}}$ définie par

$$a_i = \begin{cases} \lambda^i & \text{si } i \geq 0 \\ 0 & \text{si } i < 0 \end{cases}$$

est absolument sommable.

- En effectuant le produit de cette série par $I - \lambda B$, déduire que $I - \lambda B$ est inversible et déterminer son inverse.

2) **Cas où $|\lambda| > 1$** : démontrer le résultat de la même façon.

3) **Cas où $|\lambda| = 1$** : montrer dans ce cas que l'opérateur n'est pas inversible en étudiant le processus constant.

1.3.2 Inversion d'un polynôme en B

Considérons maintenant le cas plus général de l'opérateur polynomial : plus généralement, soit le polynôme

$$\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p.$$

Proposition 1.5. *L'opérateur $\Phi(B)$ est inversible si et seulement si les racines du polynôme Φ sont de module différent de 1.*

Exercice On se propose de démontrer le résultat de la proposition précédente par étapes.

1. Supposons que les racines $z_j = \frac{1}{\lambda_j}$ de Φ sont distinctes, réelles et de module strictement supérieur à 1.

- Vérifier que le polynôme en z est Φ inversible. On appellera son inverse.

- Vérifier que

$$\Phi(z) = \prod_{i=1}^p (1 - \lambda_i z)$$

- Décomposer la fraction $\frac{1}{\Phi(z)}$ en éléments simples et effectuer le développement en série de chacun des termes de la décomposition afin de déterminer l'inverse de Φ .

2. Le cas des racines complexes ou multiples (plus lourd au niveau de l'écriture) se traite de manière habituelle et analogue.

3. Supposons que les r premières racines sont de module supérieur à 1 et que les $p-r$ dernières sont de module inférieur à 1.

- En décomposant Φ selon ses racines, écrire $\Phi(B)$ comme le produit de deux polynômes (l'un en B et l'autre en F) et de $\prod_{i=r+1}^p (-\lambda_i B)$.
- En utilisant les résultats précédents en déduire que $\Phi(B)$ est inversible et déterminer son inverse.

Définition 1.7. Les λ_j sont appelés les **co-racines** du polynôme Φ .

1.4 Processus stationnaires du second ordre

Dans de très nombreux cas, on ne peut pas renouveler la suite de mesures dans des conditions parfaitement identiques (météo, économie...). Alors pour que le modèle déduit à partir d'une suite d'observations ait un sens, il faut que toute portion de trajectoire observée fournisse des informations sur la loi de X et que des portions différentes mais de même longueur fournissent les mêmes indications.

C'est ce qui nous amène à définir la notion de stationnarité.

1.4.1 Définitions

Considérons une suite de variables aléatoires $(X_t)_{t \in \mathbb{Z}}$. On dit que cette suite est **stationnaire en moyenne** lorsque la moyenne de chacune des variables de la suite est identique :

$$\mathbb{E}(X_t) = \mathbb{E}(X_0) =: \mu_X, \quad \forall t \in \mathbb{Z}.$$

De même, cette suite sera **stationnaire en variance** lorsque la variance de chacune des variables de la suite est identique :

$$\text{Var}(X_t) = \text{Var}(X_0) =: \sigma_X^2, \quad \forall t \in \mathbb{Z}.$$

La définition suivante caractérise les suites qui sont stationnaires en moyenne et dont la structure de covariance reste elle aussi constante.

Définition 1.8. *Un processus $X = (X_t)_{t \in \mathbb{Z}}$ est **stationnaire du second ordre** (ou **faiblement stationnaire**) si et seulement si*

(i) $\mathbb{E}(X_t) = \mu, \quad \forall t \in \mathbb{Z};$

(ii) X_t est de carré intégrable pour tout $t \in \mathbb{Z} : \mathbb{E}(X_t^2) < \infty;$

(iii) $\text{Cov}(X_s, X_{s+t}) = \text{Cov}(X_{s-1}, X_{s-1+t}) = \dots = \text{Cov}(X_0, X_t) \quad \forall t, s \in \mathbb{Z}.$

Dans cette définition, la propriété (i) exprime la stationnarité en moyenne, (ii) assure que la variance de chaque variable est finie et (iii) précise ce que l'on entend par "invariance de la structure de covariance". Par cette propriété, on peut introduire la fonction

$$\gamma(h) := \text{Cov}(X_t, X_{t+h})$$

qui est indépendante de t et n'est définie que si le processus est stationnaire du second ordre (et vérifie donc (iii)). Nous reviendrons sur cette fonction et ses propriétés dans la section suivante.

Propriété 1.2. *Si X est un processus stationnaire, $\text{Var}(X_t) = \gamma(0) = \sigma^2$ est indépendante de t . On dit que le processus est **homoscédastique**.*

Remarque 1.7. On peut également définir un concept de stationnarité à partir des lois jointes des variables aléatoires du processus X : Un processus $X = (X_t)_{t \in \mathbb{Z}}$ est **strictement stationnaire** si et seulement si $\forall k \in \mathbb{N}, t_1, \dots, t_k \in \mathbb{Z} \text{ et } \forall h \in \mathbb{Z}$, la loi conjointe de $(X_{t_1+h}, \dots, X_{t_k+h})$ ne dépend pas de h i.e.

$$\mathfrak{L}(X_{t+h_1}, X_{t+h_2}, \dots, X_{t+h_k}) = \mathfrak{L}(X_{h_1}, X_{h_2}, \dots, X_{h_k}), \quad \forall t, k(h_1, h_2, \dots, h_k).$$

Notons que si X est un processus strictement stationnaire, la loi de X_t ne dépend pas de t : les variables X_t , pour $t \in \mathbb{Z}$ sont identiquement distribuées.

Cette dernière définition de stationnarité est plus exigeante que le concept de stationnarité du second ordre puisque tout processus strictement stationnaire est stationnaire du second ordre. La réciproque est fausse. La stationnarité à l'ordre 2 est bien plus facile à étudier et vérifier que la stationnarité stricte.

L'importance pratique de la stationnarité à l'ordre 2 tient surtout aux problèmes de prédiction ou de régression. En effet, on se limite souvent à des critères de moindres carrés pour avoir des estimateurs calculables. Cela signifie alors utiliser des prédicteurs linéaires optimaux dont le calcul ne fait pas intervenir dans sa totalité la structure probabiliste du processus observé X mais seulement la géométrie (angles et longueurs) de la suite (X_k) considérée comme une suite de vecteurs dans l'espace de Hilbert $L^2(\Omega, \mathbb{P})$. Or cette géométrie ne dépend que des moments d'ordre 2 de X ; la notion naturelle de stationnarité est donc l'invariance de ces moments d'ordre 2 par translation dans le temps.

Dans la suite, nous utiliserons le terme "stationnaire" pour un processus stationnaire de second ordre et nous nous cantonnerons simplement à ce type de stationnarité.

Exemple 1.3. Un bruit blanc est un processus stationnaire.

Remarque 1.8. L'opérateur $I - \lambda B$ est un opérateur de l'ensemble des processus stationnaires dans lui-même

Exemple 1.4. *Un exemple de processus non stationnaire : la marche aléatoire.* Le processus $(X_t)_{t \in \mathbb{N}}$ est une **marche aléatoire** (random walk en anglais) lorsque

$$X_t = X_{t-1} + \epsilon_t, \quad t \in \mathbb{N}^*$$

où $(\epsilon_t)_{t \in \mathbb{N}^*}$ est une suite de variables indépendantes et identiquement distribuées de loi $\mathcal{N}(0, \sigma^2)$.

1. Réécrire X_t en fonction de X_0 et des ϵ_t .
2. Calculer l'espérance de X_t .
3. **On suppose que X_0 n'est pas aléatoire mais que c'est une donnée initiale (pour le cadre général se reporter au TD).** Calculer la covariance du processus. En déduire qu'une marche aléatoire n'est pas stationnaire.
4. En déduire la variance de X_t . Que peut-on dire ?

L'évolution de X_t est aléatoire et on ne peut pas faire de prévision car il manque une structure adéquate.

Pour illustrer ce type de processus non stationnaire, on peut penser aux prix des actifs d'aujourd'hui donnant une information sur les prix du lendemain mais dont la différence est quand même aléatoire, avec un certain contrôle de la variabilité ($\sigma^2 < \infty$). On ne peut pas faire de prévision car ϵ_t ne possède pas de structure de corrélation.

1.4.2 Fonctions d'autocovariance et d'autocorrélation

Définition 1.9. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire. On appelle **fonction d'autocovariance** la fonction γ définie de \mathbb{Z} dans \mathbb{R} par

$$\forall h, t \in \mathbb{Z}, \quad \gamma(h) := \text{Cov}(X_t, X_{t+h})$$

Le graphe de cette fonction est appelé **variogramme**.

Propriété 1.3. La fonction d'autocovariance d'un processus stationnaire vérifie

1. $\forall h \in \mathbb{Z}, \quad \gamma(-h) = \gamma(h)$: elle est paire;
2. $\forall n \in \mathbb{N}, \forall (a_i) \in \mathbb{R}^{\mathbb{N}}, \forall (t_i) \in \mathbb{Z}^{\mathbb{N}}, \quad \sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma(t_i - t_j) > 0$: elle est de type positif;
3. $\gamma(0) = \text{Var}(X_t)$;
4. $|\gamma(h)| \leq \gamma(0), \quad \forall h$.

Démonstration. A faire en exercice. □

La fonction d'autocovariance peut être "normalisée" et la nouvelle fonction obtenue est la fonction d'autocorrélation :

Définition 1.10. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire. On appelle **fonction d'autocorrélation** la fonction ρ définie de \mathbb{Z} dans \mathbb{R} par

$$\forall h \in \mathbb{Z}, \quad \rho(h) := \frac{\gamma(h)}{\gamma(0)}.$$

Le graphe de cette fonction est appelé **corrélogramme**.

Propriété 1.4. La fonction d'autocorrélation d'un processus stationnaire vérifie

1. $\forall h \in \mathbb{Z}, \rho(-h) = \rho(h)$: elle est paire ;
2. $\rho(0) = 1$;
3. $|\rho(h)| \leq 1, \forall h$.

La fonction $\rho(h)$ est l'expression du lien linéaire entre X_t et X_{t-h} . Si t est l'instant présent et $h > 0$, $\rho(h)$ est l'expression du lien linéaire entre le présent et le passé d'ordre h (mais aussi par parité de la fonction γ entre le présent et le futur d'ordre h) ; plus $|\rho(h)|$ est proche de 1 et plus ce lien est fort.

Exemple 1.5. Considérons le processus X_t stationnaire centré défini par

$$X_t = \phi X_{t-1} + \eta_t,$$

où η_t est un bruit blanc avec $|\phi| < 1$. Nous verrons dans la Section 2.3 que ce processus est appelé processus autorégressif d'ordre 1 ou encore AR(1).

1. Calculer $\rho(1)$ et $\rho(2)$.
2. En déduire que $\rho(h) = \phi^k$ pour $h \geq 0$ et donc que $\rho(h) = \phi^{|h|}$ pour tout h .

On sent bien où se situe le problème de la fonction d'autocorrélation : il semblerait, par la valeur de cette fonction, qu'il y ait une relation entre X_t et X_{t-2} .

Or cette relation n'est qu'indirecte puisqu'elle n'existe que par X_{t-1} et par les relations entre ce-dernier et X_t et X_{t-2} .

Comme nous venons de le voir sur l'exemple précédent, les valeurs $\rho(h)$, $h=1, 2, \dots$ ne sont souvent pas suffisantes pour expliquer comment présent et passé sont reliés car les variables du passé sont elles mêmes reliées entre elles. Autrement dit, X_{t-h} dépend en général des variables $X_{t-h+1}, \dots, X_{t-1}$. On a donc recours à une nouvelle fonction, la **fonction d'autocorrélation partielle** $\tau(h)$ qui exprime le lien entre X_t et X_{t-h} lorsqu'on a retiré leur meilleure explication affine en termes de $X_{t-h+1}, \dots, X_{t-1}$.

Cette fonction existe donc pour connaître jusqu'à quel niveau de décalage ou ordre, il existe une relation directe entre X_t et les valeurs précédentes.

Définition 1.11. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire. Définissons tout d'abord la régression affine de X_t sur $(X_{t-1}, \dots, X_{t-h})$ notée $X_{t,h}^*$. On a

$$X_t = X_{t,h}^* + R_{t,h} = \lambda_{0,h} + \sum_{s=1}^h \lambda_{s,h} X_{t-s} + R_{t,h}$$

où $R_{t,h}$ est une variable aléatoire non corrélée avec $(X_{t-1}, \dots, X_{t-h})$. La **fonction d'autocorrélation partielle** τ est définie par

$$\forall h \in \mathbb{Z}, \quad \tau(h) = \lambda_{h,h}$$

Le graphe de cette fonction est appelé **corrélogramme partiel**.

Proposition 1.6. L'autocorrélation partielle d'ordre h $\tau(h)$ représente le coefficient de corrélation linéaire entre

- le résidu $X_t - X_{t,h-1}^*$ de la régression de X_t par $X_{t-h+1}, \dots, X_{t-1}$

et

- le résidu $X_{t-h} - X_{t-h,h-1}^*$ de la régression de X_{t-h} par $X_{t-h+1}, \dots, X_{t-1}$.

En d'autres termes,

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_{h-1} X_{t-h+1} + U$$

$$X_{t-h} = \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_{h-1} X_{t-h+1} + V$$

et

$$\tau(h) = \text{Cor}(U, V)$$

Il faut bien comprendre que l'autocorrélation partielle est la corrélation entre X_t et X_{t-h} une fois que l'on a expliqué ceux-ci par les valeurs entre eux deux $X_{t-h+1}, \dots, X_{t-1}$.

Démonstration. En effet, on a

$$\frac{\text{Cov}(X_t - X_{t,h-1}^*, X_{t-h} - X_{t-h,h-1}^*)}{\sqrt{\text{Var}(X_t - X_{t,h-1}^*) \text{Var}(X_{t-h} - X_{t-h,h-1}^*)}} = \frac{\text{Cov}(R_{t,h-1}, R_{t-h,h-1})}{\sqrt{\text{Var}(R_{t,h-1}) \text{Var}(R_{t-h,h-1})}} \quad (1.1)$$

De plus, la stationnarité de X entraîne que

$$\text{Var}(R_{t,h-1}) = \text{Var}(R_{t-h,h-1})$$

et par conséquent le second membre de (1.1) ne dépend pas de t . Notons cette expression $r(h)$; sa valeur représente le coefficient de régression de $X_t - X_{t,h-1}^*$ sur $X_{t-h} - X_{t-h,h-1}^*$.

Ecrivons maintenant la régression affine de X_t sur X_{t-h}, \dots, X_{t-1} ; on a

$$X_t = \lambda_{0,h} + \lambda_{1,h}X_{t-1} + \dots + \lambda_{h,h}X_{t-h} + R_{t,h}.$$

En projetant chaque membre de l'égalité sur $L_{X_{t-h+1}, \dots, X_{t-1}}$, on obtient

$$\begin{aligned} X_{t,h-1}^* &= \lambda_{0,h} + \lambda_{1,h}X_{t-1,h-1}^* + \dots + \lambda_{h,h}X_{t-h,h-1}^* \\ &= \lambda_{0,h} + \lambda_{1,h}X_{t-1} + \dots + \lambda_{h-1,h}X_{t-h-1} + \lambda_{h,h}X_{t-h,h-1}^* \end{aligned}$$

Et par conséquent,

$$X_t - X_{t,h-1}^* = \lambda_{h,h}(X_{t-h} - X_{t-h,h-1}^*) + R_{t,h},$$

où $R_{t,h}$ est orthogonal à $X_{t-h} - X_{t-h,h-1}^*$. De cette expression, on déduit que le coefficient de régression $r(h)$ de $X_t - X_{t,h-1}^*$ sur $X_{t-h} - X_{t-h,h-1}^*$ est égal à $\lambda_{h,h} = \rho(h)$. D'où le résultat. \square

Exemple 1.6. Revenons à l'exemple du processus AR(1).

1. Calculer $\tau(1)$.

2. En faisant la régression de X_{t-1} sur X_{t-2} et X_t , calculer $\tau(2)$. **Rappel:** Dans une régression

$$Y_i = \alpha X_i + \epsilon_i,$$

le coefficient α peut être calculé comme

$$\alpha = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{\text{Cor}(X, Y)\sigma_X\sigma_Y}{\text{Var}(X)} = \text{Cor}(X, Y)\frac{\sigma_Y}{\sigma_X}$$

C'est à dire que si X et Y ont le même écart-type, que l'on fasse la régression de X sur Y ou la régression de Y sur X on retrouve le même coefficient $\alpha := \text{Cor}(X, Y)$. Pour résumer, si on dispose de deux variables aléatoires X et Y de moyennes nulles et de même écart-type, si on fait la régression de X sur Y

$$Y = \alpha X + \epsilon$$

alors la régression de Y sur X est

$$X = \alpha Y + \delta$$

et non comme on pourrait s'y attendre

$$X = \frac{1}{\alpha} Y + \delta$$

3. En déduire que $\tau(h) = 0, \quad \forall h > 1$.

Relation entre les fonctions d'autocorrélation et d'autocorrélation partielle

Nous avons tout d'abord

$$\tau(1) = \lambda_{1,1} = \frac{\text{Cov}(X_t, X_{t-1})}{\sigma^2} = \rho(1).$$

Déterminons maintenant une expression explicite de $\tau(2)$. Pour cela, écrivons

$$X_t = \lambda_{0,2} + \lambda_{1,2}X_{t-1} + \lambda_{2,2}X_{t-2} + R_{t,2}$$

avec $R_{t,2}$ indépendant à la fois de X_{t-1} et X_{t-2} . En multipliant les deux membres de cette égalité par X_{t-1} d'une part et X_{t-2} d'autre part, puis en prenant l'espérance, on obtient le système d'équations

$$\begin{cases} \rho(1) = \lambda_{1,2} + \lambda_{2,2}\rho(1) \\ \rho(2) = \lambda_{1,2}\rho(1) + \lambda_{2,2} \end{cases}$$

On en déduit

$$\tau(2) = \lambda_{2,2} = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2} = \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}}$$

En généralisant la démarche précédente, on obtient pour déterminer $\tau(h)$ un système de h équations linéaires

$$\begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(h) \end{pmatrix} = \begin{pmatrix} 1 & \rho(1) & \dots & \rho(h-1) \\ \rho(1) & 1 & \dots & \rho(h-2) \\ \vdots & \vdots & \dots & \vdots \\ \rho(h-1) & \rho(h-2) & \dots & 1 \end{pmatrix} \begin{pmatrix} \lambda_{1,h} \\ \lambda_{2,h} \\ \vdots \\ \lambda_{h,h} \end{pmatrix}$$

D'où l'on déduit que

$$\tau(h) = \lambda_{h,h} = \frac{\begin{vmatrix} 1 & \rho(1) & \dots & \rho(h-2) & \rho(1) \\ \rho(1) & 1 & \dots & \rho(h-3) & \rho(2) \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \rho(h-1) & \rho(h-2) & \dots & \rho(1) & \rho(h) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \dots & \rho(h-2) & \rho(h-1) \\ \rho(1) & 1 & \dots & \rho(h-3) & \rho(h-2) \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \rho(h-1) & \rho(h-2) & \dots & \rho(1) & 1 \end{vmatrix}}$$

Ceci montre que $\tau(h)$ est fonction de $\rho(1), \dots, \rho(h)$. Inversement, $\rho(h)$ est fonction de $\tau(1), \dots, \tau(h)$ et bien sûr

$$\tau(1) = \rho(1).$$

La connaissance de $\rho(h)$ est donc équivalente à celle de $\tau(h)$. Selon les cas, il sera plus commode d'utiliser l'une ou l'autre de ces fonctions.

Nous terminons le paragraphe par une dernière définition

Définition 1.12. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus de carré intégrable et X_t^* la régression affine de X_t sur $(X_s, s < t)$. On appellera innovation du processus à la date t la variable aléatoire réelle $\epsilon_t^* := X_t - X_t^*$.

Remarque 1.9. La v.a.r. X_t^* représente la régression affine de X_t sur le passé du processus à la date t . Il s'agit donc de la projection de X_t sur l'espace $L_{(X_s, s < t)}$. C'est la meilleure approximation affine de X_t fonction du passé du processus. L'innovation est alors la partie de X_t non corrélée au passé.

Exemple 1.7. Lorsque $(X_t)_{t \in \mathbb{Z}}$ est stationnaire, la suite des innovations $(\epsilon_t^*)_{t \in \mathbb{Z}} := (X_t - X_t^*)_{t \in \mathbb{Z}}$ est un bruit blanc.

1.4.3 Estimations des fonctions caractéristiques

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire de moyenne μ ($\mu = \mathbb{E}(X_t) \forall t$), de fonctions d'autocorrélation ρ et d'autocorrélation partielle τ . On dispose en pratique de l'observation du processus jusqu'à l'instant T soit de X_1, \dots, X_T .

Estimation de la moyenne On estime généralement la moyenne du processus par la **moyenne empirique**

$$\bar{X} := \frac{1}{T} \sum_{i=1}^T X_i.$$

Propriété 1.5. 1. \bar{X} est un estimateur sans biais de μ : $\mathbb{E}(\bar{X}) = \mu$.

2.
$$\text{Var}(\bar{X}) = \frac{1}{T} \sum_{|h| < T} \left(1 - \frac{|h|}{T}\right) \gamma(h).$$

Démonstration. A faire en exercice. □

Remarque 1.10. Les variables ne sont cependant pas nécessairement indépendantes ce qui nous empêche d'appliquer la loi forte des grands nombres. Par contre, la proposition suivante donne une condition suffisante de convergence en moyenne quadratique.

Proposition 1.7. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire de moyenne μ et de fonction d'autocovariance γ . Si $\sum_{h=0}^{+\infty} |\gamma(h)| < \infty$, alors

$$\text{Var}(\bar{X}) = \mathbb{E} \left((\bar{X} - \mu)^2 \right) \xrightarrow{T \rightarrow \infty} 0.$$

On dit alors que ce processus est **ergodique** (pour l'espérance).

Démonstration. A faire en exercice. □

Estimation de la fonction d'autocovariance Pour construire un estimateur de la fonction d'autocovariance, rappelons que si $(X_1, Y_1), \dots, (X_T, Y_T)$ sont des observations bivariées i.i.d. de variance finie, un estimateur de la covariance entre X et Y est donné par

$$\frac{1}{T} \sum_{t=1}^T (X_t - \bar{X})(Y_t - \bar{Y})$$

On estime donc naturellement la fonction d'autocovariance par

$$\hat{\gamma}(h) := \frac{1}{T} \sum_{t=1}^{T-|h|} (X_t - \bar{X})(X_{t+|h|} - \bar{X})$$

.Il est appelé l'**autocovariance empirique**.

Propriété 1.6. *On peut montrer que*

1. $\hat{\gamma}(h)$ est un estimateur biaisé de $\gamma(h)$.
2. Sous certaines conditions, on peut montrer que $\hat{\gamma}(h)$ est asymptotiquement non biaisé.
3. Si $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire gaussien et si la série de terme général $\gamma(h)$ est absolument convergente, alors $\hat{\gamma}(h)$ converge en moyenne quadratique vers $\gamma(h)$.

Estimation des fonctions d'autocorrélation On estime ensuite naturellement la fonction d'autocorrélation à partir de l'autocovariance empirique de la façon suivante

$$\hat{\rho}(h) := \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} = \frac{\sum_{t=1}^{T-|h|} (X_t - \bar{X})(X_{t+|h|} - \bar{X})}{\sum_{t=1}^T (X_t - \bar{X})^2}$$

Il est appelé l'**autocorrélation empirique**.

Propriété 1.7. 1. $\hat{\rho}(h)$ est de type positif.

2. $\forall h \in \mathbb{Z}, \quad -1 \leq \hat{\rho}(h) \leq 1.$

Remarque 1.11. *Sous certaines conditions, on peut démontrer un théorème central limite qui précise le comportement asymptotique de $\hat{\gamma}(h)$.*

Enfin, on obtient un estimateur $\hat{\tau}(h)$ de $\tau(h)$ en remplaçant dans l'expression de $\tau(h)$ vue à la section précédente, $\rho(h)$ par $\hat{\rho}(h)$.

Remarque 1.12. *L'algorithme de Durbin-Watson, que nous ne présentons pas ici permet d'estimer les autocorrélations partielles d'un processus stationnaire.*

Chapitre 2

Processus ARMA

Dans ce chapitre, nous entamons l'étude détaillée de la partie aléatoire d'une série chronologique. Les processus linéaires constituent le modèle le plus simple pour décrire cette composante. Nous avons déjà vu dans les chapitres précédents des exemples de processus linéaires : les processus moyennes mobiles (au premier semestre) et le processus autorégressif d'ordre 1 (en exemple au chapitre précédent). Ces processus sont des combinaisons linéaires de processus de bruit blanc (fort ou faible) dont l'intérêt tient au comportement de leur fonction d'autocovariance, qui tend vers 0 lorsque l'on compare deux variables éloignées dans le temps :

$$\gamma(h) = \text{Cov}(X_t, X_{t+h}) \xrightarrow{|h| \rightarrow +\infty} 0.$$

Intuitivement, cela signifie que la "mémoire" du processus est contrôlée : la liaison du second ordre est plus importante pour deux variables proches dans le temps que pour des variables éloignées. Ce modèle est donc à mémoire appelée "courte" même si elle est "infinie" pour un processus autorégressif.

Nous étudions dans ce chapitre les processus linéaires ainsi que deux cas particuliers de tels processus : les processus autorégressifs et les moyennes mobiles. Nous verrons dans

le prochain chapitre un autre exemple très important de processus linéaire : le processus ARMA.

2.1 Processus linéaires

Un processus linéaire est un processus stochastique X_t formé par une combinaison linéaire (non nécessairement finie) de bruits blancs forts. On définit également la classe des processus linéaires généraux, qui sont constitués de combinaisons linéaires de bruits blancs faibles. Introduisons formellement ces deux types de processus.

Définition 2.1. $(X_t)_{t \in \mathbb{Z}}$ est un **processus linéaire** (resp. **linéaire général**) de moyenne μ s'il peut être écrit sous la forme

$$X_t = \mu + \sum_{i=-\infty}^{+\infty} b_i \epsilon_{t-i},$$

où $(\epsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc fort (resp. faible), de variance σ^2 et où la suite des coefficients b_i est supposée telle que

$$\sum_{i=-\infty}^{+\infty} b_i^2 < +\infty.$$

Le résultat suivant montre que les processus linéaires que nous avons définis sont des processus stationnaires.

Proposition 2.1. Si $(X_t)_{t \in \mathbb{Z}}$ est un processus linéaire général alors $(X_t)_{t \in \mathbb{Z}}$ est stationnaire et on a

$$\left\{ \begin{array}{l} \gamma(0) = \text{Var} X_t = \sigma^2 \sum_{i=-\infty}^{+\infty} b_i^2 \\ \gamma(h) = \text{Cov}(X_t, X_{t+h}) = \sigma^2 \sum_{i=-\infty}^{+\infty} b_i b_{i+h} \quad \forall h \in \mathbb{Z} \end{array} \right.$$

Démonstration. A faire en exercice. □

Nous présentons maintenant deux résultats portant sur l'image de processus stationnaires par certaines transformations.

Proposition 2.2. Soit $X = (X_t)_{t \in \mathbb{Z}}$ un processus stationnaire et soit $(a_i)_{i \in \mathbb{Z}}$ une suite de réels absolument sommable (i.e. telle que $\sum_{i=-\infty}^{+\infty} |a_i| < +\infty$). Alors le processus $Y = (Y_t)_{t \in \mathbb{Z}}$ défini par

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

est stationnaire.

Démonstration. A faire en exercice. □

Exemple 2.1. Soit $(\epsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc et soit $(a_i)_{i \in \mathbb{Z}}$ une suite de réels absolument sommable (i.e. telle que $\sum_{i=-\infty}^{+\infty} |a_i| < +\infty$). Alors le processus $Y = (Y_t)_{t \in \mathbb{Z}}$ défini par

$$Y_t = \sum_{i=-\infty}^{+\infty} a_i \epsilon_{t-i}$$

est un processus stationnaire. Il est naturellement appelé moyenne mobile infinie. On a alors

$$\mathbb{E}(Y_t) = 0 \quad \text{et} \quad \text{Var}(Y_t) = \sigma^2 \sum_{i=-\infty}^{+\infty} a_i^2$$

et plus généralement

$$\gamma(h) = \sigma^2 \sum_{i=-\infty}^{+\infty} a_i a_{i-h}.$$

Terminons cette section par quelques définitions. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire centré. On note $H(X)$ et on appelle **histoire de X** l'espace engendré par les $(X_t)_{t \in \mathbb{Z}}$ (i.e. l'espace des combinaisons linéaires des $(X_t)_{t \in \mathbb{Z}}$). Le **passé linéaire de X jusqu'à l'instant t**, noté $H_{\leq t}(X)$ est l'espace engendré par les $(X_u)_{u \leq t}$. Les $H_{\leq t}(X)$ forment une chaîne croissante de sous-espaces de $H(X)$. Leur réunion est dense dans $H(X)$ et leur intersection $H_{-\infty}(X) = \cap_t H_{\leq t}(X)$ est appelée **passé lointain de X**.

Définissons maintenant pour tout t , P_t la projection orthogonale sur $H_{\leq t}(X)$. Notons alors

$$\epsilon_t := X_t - P_{t-1}(X_t).$$

Définition 2.2. La série temporelle $\epsilon = (\epsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc appelé **bruit blanc d'innovation** et les ϵ_t sont les **innovations fondamentales** du processus X .

La relation

$$X_t = P_{t-1}(X_t) + \epsilon_t$$

montre que ϵ_t est ce qui arrive de "nouveau" à X à l'instant t . La variance de ϵ est nécessairement inférieure ou égale à celle de X .

Définition 2.3. 1. On dit que X est **purement déterministe** si son bruit blanc d'innovation est nul. On peut alors (en théorie) prévoir exactement les valeurs futures de X en fonction de la connaissance de toutes ses valeurs du passé. Ce qui est équivalent à $H_{\leq t}(X) = H_{\leq t+1}(X)$ pour au moins un t ; ce qui est encore équivalent à $H_{-\infty}(X) = H(X)$.

2. L'extrême opposé est le suivant : on dit que X est purement innovant si X est non nul et si $H_{-\infty}(X) = \{0\}$.

Théorème 2.1. (Wold)

Pour tout processus stationnaire centré X , il existe (de manière unique) un processus purement innovant (ou nul) Y et un processus purement déterministe η non corrélés tels que

$$X = Y + \eta.$$

De plus, la partie purement innovante s'exprime (de manière unique) comme moyenne mobile en fonction du bruit blanc d'innovation ϵ de X :

$$Y_t = \sum_{j \geq 0} c_j \epsilon_{t-j} \quad \forall t,$$

avec $c_0 = 1$ et $\sum_{j \geq 0} c_j^2 < \infty$. Le bruit blanc d'innovation de X est également bruit blanc d'innovation pour Y .

Définition 2.4. 1. Un processus $(X_t)_{t \in \mathbb{Z}}$ admet une **représentation inversible** s'il peut s'écrire comme combinaison linéaire des valeurs d'un autre processus, c'est-à-dire qu'il existe une suite $(\psi_i, i \in \mathbb{Z})$ et un processus $(Y_t)_{t \in \mathbb{Z}}$ tels que

$$\forall t \in \mathbb{Z}, \quad X_t = \sum_{i \in \mathbb{Z}} \psi_i Y_{t-i}.$$

2. Un processus $(X_t)_{t \in \mathbb{Z}}$ admet une **représentation causale** s'il peut s'écrire comme combinaison linéaire des valeurs passées d'un autre processus, c'est-à-dire qu'il existe une suite

$(\psi_i, i \in \mathbb{Z})$ et un processus $(Y_t)_{t \in \mathbb{Z}}$ tels que

$$\forall t \in \mathbb{Z}, \quad X_t = \sum_{i \in \mathbb{N}} \psi_i Y_{t-i}.$$

Le théorème ci-dessous traite du cas fondamental des filtres causaux et de leur inversion et découle immédiatement des résultats de la Section (1.3) :

Théorème 2.2. Soit $(X_t)_{t \in \mathbb{Z}}$ le processus stationnaire solution de l'équation suivante :

$$\forall t \in \mathbb{Z}, \quad \Phi(B)X_t = Y_t,$$

où $(Y_t, t \in \mathbb{Z})$ est un processus stationnaire et $\Phi(B)$ le polynôme en B de la forme

$$\Phi(B) = I - \phi_1 B - \dots - \phi_p B^p$$

avec $p \in \mathbb{N}^*$. Alors

1. Si Φ n'a pas de racine de module égal à 1, alors il existe une **représentation inversible** du processus $(X_t)_{t \in \mathbb{Z}}$, c'est-à-dire qu'il existe une suite $(\psi_i, i \in \mathbb{Z})$ telle que

$$\forall t \in \mathbb{Z}, \quad X_t = \sum_{i \in \mathbb{Z}} \psi_i Y_{t-i}.$$

2. Si de plus toutes les racines de Φ sont de module supérieur à 1, alors il existe une **représentation causale** du processus $(X_t)_{t \in \mathbb{Z}}$, c'est-à-dire qu'il existe une suite $(\psi_i, i \in \mathbb{Z})$ telle que

$$\forall t \in \mathbb{Z}, \quad X_t = \sum_{i \in \mathbb{N}} \psi_i Y_{t-i}.$$

Autrement dit, dans le premier cas, on garantit que X_t puisse s'écrire en fonction du processus $Y_t, t \in \mathbb{Z}$ et dans le second cas, que X_t puisse s'écrire en fonction seulement du passé de Y_t . Si ϕ a au moins une racine de module supérieur à 1, alors il n'existe pas de représentation inversible. Le lemme suivant permet de bien comprendre l'importance du rôle des racines de ϕ dans le cas particulier où $\phi(B) = I - \lambda B$.

Lemme 2.1. Soit $(X_t, t \in \mathbb{Z})$ le processus stationnaire d'ini comme dans le théorème ci-dessus avec ϕ de la forme

$$\phi(B) = I - \lambda B.$$

Alors

1. Si $|\lambda| < 1$, il existe une **représentation causale** du processus $(X_t, t \in \mathbb{Z})$.
2. Si $|\lambda| > 1$, il existe une **représentation inversible** du processus $(X_t, t \in \mathbb{Z})$.

Dans la suite de ce chapitre, nous étudions en particulier deux familles de processus linéaires que nous supposerons de plus centrés : les processus MA, les processus AR. Il suffira dans le cas où X est non centré de moyenne μ ($\mathbb{E}(X_t) = \mu \quad \forall t$) de poser $Y_t = X_t - \mu$ et d'appliquer les résultats à Y .

Dans la suite du cours, nous verrons aussi les processus ARMA.

2.2 Processus moyennes mobiles

2.2.1 Définition d'un MA (q)

Ces processus forment une classe flexible de modèles pour de nombreux phénomènes observés. Ils sont construits à partir de l'idée que l'observation au temps t s'explique linéairement par les observations d'un bruit blanc ; ils sont donc définis par la relation

$$\forall t \in \mathbb{Z}, \quad X_t = \eta_t - \theta_1 \eta_{t-1} - \dots - \theta_q \eta_{t-q}$$

où $\theta_1, \theta_2, \dots, \theta_q$ sont des réels fixés et $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc de variance σ^2 .

L'observation au temps t est donc la somme d'un choc aléatoire η_t à l'instant t et d'une fonction linéaire du passé de ce choc $-\sum_{i=1}^q \theta_i \eta_{t-i}$.

Nous avons déjà rencontré ce type de processus au premier semestre.

Définition 2.5. On appelle **processus moyenne mobile** [Moving Average] d'ordre q tout processus $(X_t)_{t \in \mathbb{Z}}$, stationnaire tel que

$$\forall t \in \mathbb{Z}, \quad X_t = \eta_t - \theta_1 \eta_{t-1} - \dots - \theta_q \eta_{t-q}$$

où $\theta_1, \theta_2, \dots, \theta_q$ sont des réels fixés et $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc de variance σ^2 . Un tel processus est dit MA(q) (Moving Average of order q).

Posons $\Theta(B) = I - \theta_1 B - \dots - \theta_q B^q$, on peut écrire :

$$X_t = \Theta(B)\eta_t, \quad t \in \mathbb{Z}.$$

On suppose bien évidemment que q est inférieur au nombre d'observations.

Remarque 2.1. Notons immédiatement que

- un processus à moyenne mobile est défini de manière explicite ;
- un processus à moyenne mobile est automatiquement centré et stationnaire ;
- un processus à moyenne mobile est par définition purement innovant et causal.

Mais attention son bruit blanc innovant n'est pas nécessairement égal à η .

On peut étendre cette définition aux $MA(\infty)$ en faisant croître q . On pourra alors vérifier que X est stationnaire ssi $\sum_{j \in \mathbb{Z}} \theta_j^2 < \infty$.

2.2.2 Forme autorégressive infinie d'un MA (q)

Dans le cas régulier où le polynôme Θ ne s'annule pas sur le cercle unité, on a le théorème

Théorème 2.3. (i) Si Θ n'a pas de racine de module égal à 1, alors il existe un choix unique de coefficients π_j pour lesquels on a

$$\eta_t = \sum_{j \in \mathbb{Z}} \pi_j X_{t-j},$$

avec $\sum_{i=-\infty}^{+\infty} |\pi_i| < \infty$. Dans ce cas, on dit que le processus est **inversible**.

(ii) Si de plus les racines de Θ sont toutes de module strictement supérieur à 1, alors il existe un choix unique de coefficients π_j pour lesquels on a

$$\eta_t = \sum_{j \in \mathbb{N}} \pi_j X_{t-j}.$$

Les coefficients π_j convergent rapidement vers 0 lorsque $j \rightarrow \infty$, en effet ce sont les coefficients du filtre associé au polynôme $1/\Theta$. Dans ce cas, on dit que le processus est **causal** et η_t est l'innovation du processus à la date t .

Dans le cas où Θ n'a pas de racine multiple, on a la formule explicite suivante

$$\forall j \geq 0, \quad \pi_j = \sum_{1 \leq k \leq q} m_k \mu_k^j$$

avec

$$m_k = \frac{1}{\prod_{l \neq k} \left(1 - \frac{\mu_l}{\mu_k}\right)}$$

et

$$\Theta(z) = 1 - \theta_1 z - \dots - \theta_q z^q = \prod_j (1 - \mu_j z).$$

Par contre, si Θ s'annule sur le cercle unité, on a le théorème suivant :

Théorème 2.4. *Si Θ s'annule en un point du cercle unité, alors il n'existe aucun choix de coefficients π_j pour lesquels $\sum_{j \geq 0} \pi_j X_{t-j}$ converge et coïncide avec η_t . Cependant, on peut représenter η_t comme limite de telles combinaisons linéaires des X_u , $u \leq t$.*

2.2.3 Bruit blanc d'innovation d'un MA (q)

Comme nous l'avons dit précédemment, le bruit blanc d'innovation $(\epsilon_t)_{t \in \mathbb{Z}}$ du processus $MA(q)$ $(X_t)_{t \in \mathbb{Z}}$ n'est pas nécessairement le processus $(\eta_t)_{t \in \mathbb{Z}}$ c'est-à-dire que η_t ne représente pas l'information ajoutée au passé pour obtenir la valeur présente de X_t . Mais nous avons le résultat suivant

Théorème 2.5. *La relation d'un $MA(q)$ avec son bruit blanc d'innovation est aussi du type $X_t = \tilde{\Theta}(B)\epsilon_t$, $t \in \mathbb{Z}$. Le polynôme $\tilde{\Theta}$ de degré q , dit **polynôme canonique**, s'obtient à partir de Θ en remplaçant toutes les racines éventuelles à l'intérieur du cercle unité par leur inversion $(\mu \mapsto \frac{1}{\bar{\mu}})$, les plaçant ainsi à l'extérieur du cercle unité. Notons μ_1, \dots, μ_q les co-racines de $\Theta(z)$:*

$$\Theta(z) = \prod_j (1 - \mu_j z) = 1 - \theta_1 z - \dots - \theta_q z^q.$$

Le polynôme $\tilde{\Theta}$ qui relie X à son bruit blanc d'innovation est donné par

$$\tilde{\Theta}(z) = \prod_{j, |\mu_j| \leq 1} (1 - \mu_j z) \prod_{j, |\mu_j| > 1} \left(1 - \frac{1}{\mu_j} z\right).$$

Nous verrons en TD quelle est la relation qui lie $(\eta_t)_t$ à $(\epsilon_t)_t$ dans le cadre d'un exemple simple.

On d'eduit du théorème précédent que

Proposition 2.3. La variance du bruit blanc d'innovation ϵ est donnée par

$$\text{Var}(\epsilon_t) = \gamma_\epsilon(0) = \gamma_\eta(0) \prod_{j, |\mu_j| > 1} |\mu_j|^2.$$

A partir de maintenant, nous ne considérerons plus que des processus $MA(q)$ donnés sous leur représentation canonique i.e. vérifiant l'équation

$$X_t = \Theta(B)\eta_t$$

avec $\Theta(z) = 1 - \theta_1 z - \dots - \theta_q z^q = \prod_{i=1}^q (1 - \mu_i z)$. où $|\mu_i| \leq 1$ puisque nous venons de voir qu'il est toujours possible de se ramener à ce cas quitte à changer de bruit blanc.

2.2.4 Fonctions caractéristiques et propriétés d'un MA (q)

Exemple 2.2. Soit X un $MA(1)$. (X_t) est donc généré par un bruit blanc (η_t) sous la forme

$$X_t = \eta_t - \theta \eta_{t-1} = (I - \theta B) \eta_t$$

La fonction de transfert du filtre se réduit à un seul terme.

1. Par simples calculs, on peut montrer que

$$\text{Var}(X_t) = (1 + \theta^2)\sigma^2, \quad \forall t.$$

Ce qui signifie qu'en modélisant, on diminue la variance du phénomène ce qui est par nature le propre de toute modélisation!

Et plus généralement,

$$\gamma(h) = \begin{cases} -\sigma^2\theta & \text{si } |h| = 1 \\ 0 & \text{si } |h| \geq 2 \end{cases}$$

2. De même, on peut montrer que

$$\rho(h) = \begin{cases} -\frac{\theta}{1+\theta^2} & \text{si } |h| = 1 \\ 0 & \text{si } |h| \geq 2 \end{cases}$$

3. Enfin à partir de l'expression générale de la fonction d'autocorrélation partielle, on peut montrer que

$$\tau(h) = -\theta^h \frac{1 - \theta^2}{1 - \theta^{2h+2}}, \quad h \in \mathbb{N}^*.$$

On voit que cette fonction décroît exponentiellement vers 0 quand h augmente.

Proposition 2.4. 1. La variance de X_t est donnée par

$$\text{Var}(X_t) = \gamma^X(0) = (\theta_1^2 + \dots + \theta_q^2)\gamma^\eta(0) > \gamma^\eta(0) = \sigma^2, \quad \forall t$$

Ayant de plus $\mathbb{E}(X_t) = 0$, on en déduit que tout processus moyenne mobile est un processus stationnaire.

2. Plus généralement, la fonction d'autocovariance est donnée par

$$\gamma(h) = \begin{cases} (-\theta_h + \theta_{h+1}\theta_1 + \dots + \theta_q\theta_{q-h})\sigma^2 & \text{si } 0 < |h| \leq q \\ 0 & \text{si } |h| > q \end{cases}$$

3. La fonction d'autocorrélation est donnée par

$$\rho(h) = \begin{cases} \frac{-\theta_h + \theta_{h+1}\theta_1 + \dots + \theta_q\theta_{q-h}}{1 + \theta_1^2 + \dots + \theta_q^2} & \text{si } 0 < |h| \leq q \\ 0 & \text{si } |h| > q \end{cases}$$

4. L'expression de la fonction d'autocorrélation partielle est compliquée. Notons simplement qu'elle ne s'annule pas à partir d'un certain rang mais qu'il existe un nombre $r \in]0, 1[$ tel que

$$|\tau(h)| \leq r^h, \quad h \geq 2.$$

Démonstration. A faire en exercice. □

En particulier, $\rho(q) = -\theta_q / (1 + \theta_1^2 + \dots + \theta_q^2)$ ne s'annule que si $\theta_q = 0$, ce qui revient à dire que le processus n'est pas d'ordre q .

Insistons sur le fait que si $h > q$ la fonction d'autocorrélation d'un processus $MA(q)$ s'annule. Cette observation sera utile pour la modélisation : si à partir de données X_1, \dots, X_T , la fonction d'autocorrélation empirique n'est pas significativement différente de zéro au-delà d'un certain nombre q_0 , on sera alors guidé pour choisir d'ajuster un modèle $MA(q_0)$ aux données.

Ce choix est en plus motivé par le résultat suivant sur la signification d'un processus $MA(q)$:

Proposition 2.5. Soit (X_t) un processus linéaire stationnaire corrélé d'ordre q , c'est à dire dont $\gamma(h) = 0$ pour tout $|h| > q$. Alors (X_t) possède une représentation comme processus $MA(q)$.

Proposition 2.6. *La somme d'un $MA(q_1)$ et d'un $MA(q_2)$ (non corrélés) est un $MA(\leq \max(q_1, q_2))$.*

Dans les figures Fig.1-3, nous avons tracé des graphes de trajectoire de MA ainsi que les corrélogramme et corrélogramme partiel empiriques associés.

2.2.5 Prédiction d'un $MA(q)$

Soit X un $MA(q)$ de polynôme Φ et de bruit blanc d'innovation η (on suppose que η est causal). La relation qui le lie à son bruit blanc d'innovation η :

$$X_t = \eta_t - \theta_1 \eta_{t-1} - \dots - \theta_q \eta_{t-q}.$$

On a donc

Théorème 2.6.

$$\hat{X}_t(1) = -\theta_1 \eta_t - \dots - \theta_q \eta_{t-(q-1)}$$

$$\hat{X}_t(2) = -\theta_2 \eta_t - \dots - \theta_q \eta_{t-(q-2)}$$

...

$$\hat{X}_t(q) = -\theta_q \eta_t$$

$$\hat{X}_t(j) = 0, \quad j > q$$

Il suffit ensuite d'exprimer explicitement les η_u en fonction des X_t comme nous l'avons fait dans la Section 2.2.2.

Par conséquent, les $MA(q)$ ne sont pas bien adaptés à la prévision puisque d'une part $j > q \Rightarrow \widehat{X}_t(j) = 0$ et d'autre part, les prévisions nécessitent la combinaison de toutes les valeurs du passé de X .

2.2.6 Erreur et intervalle de confiance pour la prévision d'un MA (q)

Notons immédiatement que $X_{t+1} - \widehat{X}_t(1) = \eta_{t+1}$. Ainsi

$$\mathbb{E}(X_{t+1} - \widehat{X}_t(1)) = 0 \quad \text{et} \quad \mathbb{E}((X_{t+1} - \widehat{X}_t(1))^2) = \sigma^2$$

Enfin, si on suppose que le bruit est gaussien, les variables aléatoires X_t sont elles aussi gaussiennes, tout comme l'erreur de prédiction. Ainsi, il sera possible de construire des intervalles de confiance sur la prédiction.

2.3 Processus autorégressifs

2.3.1 Définition d'un AR (p)

Ces processus forment une classe flexible de modèles pour de nombreux phénomènes observés. Ils sont construits à partir de l'idée que l'observation au temps t s'explique linéairement par les observations précédentes ; ils sont donc définis implicitement par la relation

$$\forall t \in \mathbb{Z}, \quad X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \eta_t,$$

où $\phi_1, \phi_2, \dots, \phi_p$ sont des réels fixés et $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc de variance σ^2 .

L'observation au temps t est donc la somme d'un choc aléatoire η_t à l'instant t (indépendant de l'historique) et d'une fonction linéaire de son passé (qui peut-être vue comme la prédiction de X_t à partir des p dernières valeurs observées. Nous reviendrons en détail sur cette propriété dans la section consacrée à la prédiction).

Contrairement aux processus MA , leur définition pose quelques problèmes :

- de leur définition ne découle pas naturellement la stationnarité. C'est pourquoi nous ajoutons l'hypothèse de stationnarité dans la définition suivante.
- ils sont définis de manière implicite. Il va donc s'agir de déterminer une forme explicite.

Définition 2.6. On appelle **processus autorégressif d'ordre p** tout processus $(X_t)_{t \in \mathbb{Z}}$ stationnaire tel que

$$\forall t \in \mathbb{Z}, \quad X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \eta_t, \quad (2.1)$$

où $\phi_1, \phi_2, \dots, \phi_p$ sont des réels fixés et $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc de variance σ^2 . Un tel processus est dit $AR(p)$ (AutoRegressive of order p).

Posons $\Phi(B) = I - \phi_1 B - \dots - \phi_p B^p$, on peut écrire :

$$\Phi(B)X_t = \eta_t, \quad \forall t \in \mathbb{Z}, \quad (2.2)$$

Remarque 2.2. 1. En prenant l'espérance de (2.1) et en utilisant la stationnarité du processus, on obtient que l'espérance μ du processus vérifie $\mu(1 - \sum_{i=1}^p \phi_i) = 0$. Donc lorsque 1 n'est pas racine de , on a nécessairement $\mu = 0$.

2. Dans ce cours, nous ne considérerons que des processus $AR(p)$ centrés. Un processus $AR(p)$ non centré serait défini par $X_t = \mu + \eta_t + \sum_{i=1}^p \phi_i X_{t-i}$.

2.3.2 Ecriture moyenne mobile infinie d'un $AR(p)$

Dans le cas régulier où le polynôme Φ ne s'annule pas sur le cercle unité, on a le théorème

Théorème 2.7. (i) Si Φ n'a pas de racine de module égal à 1, $\Phi(B)$ est inversible et on en déduit que l'équation (2.2) a une solution unique, avec une écriture moyenne mobile infinie

$$\forall t \in \mathbb{Z}, \quad X_t = (\Phi(B))^{-1} \eta_t = \sum_{j \in \mathbb{Z}} \psi_j \eta_{t-j},$$

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

Les coefficients ψ_j convergent rapidement vers 0 lorsque $j \rightarrow \infty$, en effet ce sont les coefficients du filtre associé au polynôme $1/\Phi$.

Ainsi le processus X est bien stationnaire, déterminé de manière unique par la relation précédente et la valeur présente de X dépend à la fois du passé, du présent et du futur du bruit blanc.

(ii) Si de plus les racines de sont toutes de module strictement supérieur à 1, alors l'opérateur inverse $\phi^{-1}(B)$ admet un développement ne faisant intervenir que les puissances positives de B :

$$\forall t \in \mathbb{Z}, \quad X_t = (\Phi(B))^{-1} \eta_t = \sum_{i=0}^{+\infty} \psi_i \eta_{t-i},$$

avec $\sum_{i=0}^{+\infty} |\psi_i| < +\infty$. C'est-à-dire que le processus X_t s'exprime en fonction de $\eta_s, s \leq t$ et d'après la définition, on voit que η_t n'est pas corrélé avec X_{t-1}, X_{t-2}, \dots . La variable η_t est donc l'innovation du processus à la date t et $\sum_{i=0}^{+\infty} \psi_i \eta_{t-i}$ est la régression affine de X_t sur $(X_s, s \leq t-1)$; il s'agit de la représentation canonique d'un processus $AR(p)$.

2.3.3 Bruit blanc d'innovation d'un $AR(p)$

Comme pour les MA , le bruit blanc η n'est pas nécessairement le bruit blanc d'innovation.

Théorème 2.8. *X est un AR(p) alors X est purement innovant et il existe un unique polynôme $\tilde{\Phi}$ (dit canonique) tel que le bruit blanc d'innovation de X soit de la forme $\epsilon = \tilde{\Phi}(B)X$. Ce polynôme est de degré p et s'obtient à partir de Φ par la même règle que pour les MA(q). Il est donc de la forme $\tilde{\Phi}(z) = \prod_{1 \leq k \leq p} (1 - \mu_k z)$ avec $\forall k \quad |\mu_k| < 1$.*

Nous verrons en TD quelle est la relation qui lie $(\eta_t)_t$ à $(\epsilon_t)_t$ dans le cadre d'un exemple simple.

A partir de maintenant, nous ne considérerons plus que des processus AR(p) donnés sous leur représentation canonique i.e. vérifiant l'équation

$$\Phi(B)X_t = \eta_t$$

avec $\phi(z) = 1 - \phi_1 z - \dots - \phi_q z^q = \prod_{i=1}^p (1 - \lambda_i z)$ où $|\lambda_i| \leq 1 \forall i$ puisque nous venons de voir qu'il est toujours possible de se ramener à ce cas quitte à changer de bruit blanc.

2.3.4 Fonctions caractéristiques d'un AR (p)

Exemple 2.3. *Supposons que X est un AR(1). On peut donc écrire*

$$X_t - \phi X_{t-1} = \eta_t \quad \text{i.e.} \quad (I - \phi B)X_t = \eta_t \quad (2.3)$$

où la série η_t est un bruit blanc. On peut remarquer que l'on fait une régression de la série décalée de 1 sur la série elle-même et les résidus forment un bruit blanc.

1) D'après les résultats de la Section (1.3), on peut écrire que si $|\phi| < 1$, on a

$$X_t = \sum_{j=0}^{+\infty} \phi^j \eta_{t-j}. \quad (2.4)$$

La fonction de transfert a donc une infinité de termes. On vérifie que X défini par la relation (2.4) est directement stationnaire. En effet, d'une part, de manière évidente

$$\mathbb{E}(X_t) = \mathbb{E}\left(\sum_{j=0}^{+\infty} \phi^j \eta_{t-j}\right) = \sum_{j=0}^{+\infty} \phi^j \mathbb{E}(\eta_{t-j}) = 0.$$

D'autre part, l'autocovariance est donnée par

$$\begin{aligned} \gamma(h) &= \text{Cov}(X_t, X_{t+h}) = \lim_{n \rightarrow \infty} \mathbb{E} \left[\left(\sum_{j=0}^n \phi^j \eta_{t-j} \right) \left(\sum_{k=0}^n \phi^k \eta_{t+h-k} \right) \right] \\ &= \sigma^2 \phi^{|h|} \sum_{j=0}^{+\infty} \phi^{2j} \quad (\text{série convergente car } |\phi| < 1) \\ &= \frac{\sigma^2 \phi^{|h|}}{1 - \phi^2} \end{aligned}$$

De plus, X défini par la relation (2.4) satisfait l'équation (2.3) et est par conséquent l'unique solution stationnaire de (2.3).

On a aussi

$$\text{Var}(X_t) = \gamma(0) = \frac{\sigma^2}{1 - \phi^2} > \sigma^2.$$

Ici aussi on diminue la variance en modélisant à condition que ϕ soit en valeur absolue inférieur à 1.

Les fonctions d'autocorrélation et d'autocorrélation partielle sont données par

$$\rho(h) = \phi^{|h|} \quad \text{et} \quad \tau(h) = \begin{cases} \phi & \text{si } |h| \leq 1 \\ 0 & \text{si } |h| \geq 2 \end{cases}$$

On obtient deux sortes de corrélogrammes suivant que ϕ est positif ou négatif. On voit ainsi que les variables X_t et X_{t-h} sont dépendantes avec une corrélation qui décroît exponentiellement vers 0 lorsque h tend vers l'infini.

2) Dans le cas où $|\phi| > 1$, la série (2.4) ne converge pas. On peut cependant contourner cette difficulté en réécrivant le processus AR(1) comme

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

Par les mêmes arguments que pour le cas $|\phi| < 1$ ou en utilisant le résultat de la Section (1.3), on montre que

$$X_t = -\sum_{j=1}^{+\infty} \phi^{-j} \eta_{t+j}.$$

est solution de l'équation implicite et que la série converge. Cependant, cette solution relie X_t aux valeurs futures des innovations $(\eta_{t+k})_{k>1}$. Cette solution n'est donc pas naturelle car le processus est corrélé avec les observations non encore observées du processus η_t . La solution n'est pas causale.

Par conséquent, quand on modélise des séries temporelles stationnaires, on se restreint aux processus AR(1) avec $|\phi| < 1$ de telle sorte que X a une représentation causale i.e. X_t a une représentation fonction des $(\epsilon_s, s \leq t)$. Tout processus AR(1) tel que $|\phi| > 1$ peut être réexprimé selon un AR(1) avec $|\phi| < 1$ et une nouvelle suite de bruit blanc. Ainsi on ne consid'era plus de AR(1) avec $|\phi| > 1$.

3) Dans le cas où $|\phi| = 1$, il n'y a pas de solution stationnaire à (2.3) et par conséquent les AR(1) n'existent pas selon la d'efinition des processus AR.

On peut généraliser cette notion de causalité aux processus AR(p) et ARMA.

Dorénavant, nous ne consid'ererons plus que des processus AR(p) donnés sous leur représentation canonique

$$\Phi(B)X_t = \eta_t$$

Plus généralement,

Proposition 2.7. Soit X un $AR(p)$ de polynôme générateur Φ et de bruit blanc η .

1) La fonction d'autocovariance d'un $AR(p)$ est donnée par

$$\gamma(h) = \sum_{i=1}^p \phi_i \gamma(h-i)$$

2) La fonction d'autocorrélation se déduit aisément $\rho(h) = \sum_{i=1}^p \phi_i \rho(h-i)$.

3) La fonction d'autocorrélation partielle est nulle pour $|h| > p$ et vaut ϕ_p pour $|h| = p$.

Démonstration. A faire en exercice. □

Si le corrélogramme partiel d'une série chronologique stationnaire s'annule à partir d'un rang $p+1$, on pourra donc envisager de modéliser celle-ci à l'aide d'un processus $AR(p)$.

Dans les figures Fig.4-8, nous avons tracé des graphes de trajectoire d'AR ainsi que les corrélogramme et corrélogramme partiel empiriques associés.

Exemple 2.4. A titre d'exemple, nous allons étudier le processus $AR(2)$ suivant :

$$X_t - 0.9X_{t-1} + 0.8X_{t-2} = \epsilon_t.$$

Dans la figure Fig.9, nous avons tracé le graphe d'une trajectoire de ce phénomène ainsi que les corrélogramme et corrélogramme partiel.

Le polynôme Φ est du second degré et son discriminant est négatif et vaut -2.39 . Les racines sont donc complexes. Ceci explique la forme "sinusoïdale" du corrélogramme. D'autre part, les corrélations d'ordre 1, 3 et 6 apparaissent assez élevées en module. Cette propriété devrait réapparaître sur les trajectoires : une forte valeur à une certaine date devrait en général impliquer une faible valeur trois

2.3.5 Prédiction d'un AR (p)

Les $AR(p)$ sont bien adaptés à la prédiction. Les prévisions optimales $\widehat{X}_t(h)$ pour $h > 0$ sont des combinaisons linéaires de $X_t, X_{t-1}, \dots, X_{t-p+1}$ comme d'habitude. On les calcule de façon récurrente.

A l'horizon 1 : on a

$$X_{t+1} = \phi_1 X_t + \phi_2 X_{t-1} + \dots + \phi_p X_{t+1-p} + \eta_{t+1}$$

que l'on prédit par

$$\widehat{X}_t(1) = \phi_1 X_t + \phi_2 X_{t-1} + \dots + \phi_p X_{t+1-p}$$

que l'on prédit par A l'horizon 2 : on a

$$X_{t+2} = \phi_1 X_{t+1} + \phi_2 X_t + \dots + \phi_p X_{t+2-p} + \eta_{t+2}$$

que l'on prédit par

$$\widehat{X}_t(2) = \phi_1 \widehat{X}_t(1) + \phi_2 X_t + \dots + \phi_p X_{t+2-p}$$

et ainsi de suite.

Ainsi on a le théorème suivant :

Théorème 2.9. Soit X un $AR(p)$ de bruit blanc d'innovation η et de polynôme canonique $\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$. Les prévisions optimales $\widehat{X}_t(h)$ pour $h > 0$ sont des combinaisons linéaires de $X_t, X_{t-1}, \dots, X_{t-p+1}$:

$$\widehat{X}_t(h) = \sum_{0 \leq j \leq p-1} a_j(h) X_{t-j}$$

qui s'obtiennent par récurrence selon

$$\widehat{X}_t(h) = \sum_{1 \leq k \leq p} \phi_k \widehat{X}_t(h-k)$$

avec les conditions initiales $\widehat{X}_t(-j) = X_{t-j}$ pour $j \geq 0$.

On peut donner des formules explicites pour les coefficients $a_j(h)$ mais c'est un peu compliqué, surtout si Φ a des racines multiples. Du point de vue du calcul par un ordinateur, il est en général plus simple, et tout aussi efficace, d'évaluer $\widehat{X}_t(h)$ par récurrence comme ci-dessus. Lorsque Φ n'a pas de racines multiples, on peut aussi procéder selon la remarque suivante.

Remarque 2.3. Dans le cas des processus autorégressifs causaux, le bruit blanc coïncide avec l'erreur de prédiction un pas en avant. En effet, dans ce cas,

$$\begin{aligned} \widehat{X}_t(1) &= \sum_{1 \leq k \leq p} \phi_k \widehat{X}_t(1-k) \\ &= \phi_1 \widehat{X}_t(0) + \phi_1 \widehat{X}_t(-1) + \dots + \phi_p \widehat{X}_{1-p}(0) \\ &= \phi_1 X_t + \phi_2 X_{t-1} + \dots + \phi_p X_{t-p} \\ &= X_{t+1} - \eta_{t+1} \end{aligned}$$

D'où l'on déduit que $X_{t+1} - \widehat{X}_t(1) = \eta_{t+1}$.

Proposition 2.8. Soit X un $AR(p)$ de bruit blanc d'innovation η et de polynôme canonique $\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p = \prod_{1 \leq k \leq p} (1 - \lambda_k z)$, sans racine multiple. Soit $\Phi_k(z) = \prod_{j \neq k} (1 - \lambda_j z)$ et $Y^k = \Phi_k(B)X$. Alors les Y^k sont des $AR(1)$ qui partagent avec X le même bruit blanc d'innovation. On a

$$\widehat{X}_t(h) = \sum_{1 \leq k \leq p} m_k \lambda_k^h Y_t^k$$

avec

$$m_k = \frac{1}{\prod_{j \neq k} (1 - \frac{\lambda_j}{\lambda_k})}$$

Intervalle de confiance pour la prévision

De la même façon que pour les MA , si on suppose que le bruit d'innovation est gaussien, les variables aléatoires X_t sont elles aussi gaussiennes, tout comme l'erreur de prédiction. Ainsi, il sera possible de construire des intervalles de confiance sur la prédiction.

En comparaison avec les $MA(q)$, on voit donc que pour un $AR(p)$, les prévisions $\widehat{X}_t(h)$ peuvent être non nulles même avec $h \gg 0$ et de plus s'expriment assez directement comme une combinaison linéaire des p dernières valeurs observées de X . Ces avantages paraissent décisifs... mais attention, dans la pratique, on ne connaît pas à l'avance les valeurs de ϕ_1, \dots, ϕ_p et on ne connaît pas non plus la variance $\gamma^\eta(0)$ du bruit blanc d'innovation η : pour obtenir des valeurs approchées significatives, il faudra disposer de nettement plus que p observations. On calculera alors d'abord des approximations des autocovariances de X puis on en déduira des valeurs pour ϕ_1, \dots, ϕ_p et $\gamma^\eta(0)$.

2.3.6 Equations de Yule-Walker d'un AR (p)

Soit X un $AR(p)$ causal de bruit blanc d'innovation η et de polynôme canonique associé $\Phi(x) = 1 - \phi_1 x - \dots - \phi_p x^p$.

Pour des raisons théoriques, on veut pouvoir calculer les autocovariances de X lorsque les ϕ_k et $\sigma^2 = \gamma^\eta(0)$ sont connus.

Pour des raisons pratiques, on veut pouvoir calculer ϕ_k et $\gamma^\eta(0)$ lorsque les autocovariances de X sont connues. Dans le cas d'un processus autorégressif, on peut utiliser les équations de Yule-Walker. Etudions le cas particulier des processus $AR(2)$.

Exemple 2.5. Soit le processus $AR(2)$ de la forme :

$$\forall t \in \mathbb{Z}, \quad X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = \eta_t. \tag{2.5}$$

On le suppose de plus centré et causal comme usuellement.

En multipliant l'équation (2.5) par X_{t-1} et X_{t-2} puis en prenant l'espérance de ces équations et enfin en divisant par $\gamma(0)$, on obtient les équations de Yule-Walker :

$$\begin{pmatrix} \rho(1) \\ \rho(2) \end{pmatrix} = \begin{pmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$$

On obtient donc la solution suivante :

$$\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \frac{1}{1 - \rho(1)^2} \begin{pmatrix} \rho(1)(1 - \rho(2)) \\ \rho(2) - \rho(1)^2 \end{pmatrix}.$$

Quant à la variance, partant de l'équation (2.5), on obtient que :

$$X_t^2 - \phi_1 X_{t-1} X_t - \phi_2 X_{t-2} X_t = \eta_t X_t$$

et par passage à l'espérance

$$\gamma(0) - \phi_1 \gamma(1) - \phi_2 \gamma(2) = \mathbb{E}(\eta_t X_t).$$

Mais

$$\begin{aligned} \mathbb{E}(\eta_t X_t) &= \mathbb{E}(\eta_t(\eta_t + \phi_1 X_{t-1} + \phi_2 X_{t-2})) \\ &= \sigma^2 + \phi_1 \mathbb{E}(\eta_t X_{t-1}) + \phi_2 \mathbb{E}(\eta_t X_{t-2}) \\ &= \sigma^2 \end{aligned}$$

D'où $\sigma^2 = \gamma(0) - \phi_1\gamma(1) - \phi_2\gamma(2)$. Comme on a des estimateurs convergents de la fonction d'autocovariance et donc aussi de la fonction d'autocorrélation, on pourra en déduire des estimateurs convergents de ϕ_1, ϕ_2 et σ^2 .

Sur cet exemple, on voit donc bien comment dériver les équations de Yule-Walker pour un $AR(2)$. En procédant de façon analogue, on obtient le théorème suivant

Théorème 2.10. (Yule-Walker) On a les équations suivantes :

$$\begin{aligned} (YW_0) \quad & \gamma^X(0) - \phi_1\gamma^X(-1) - \dots - \phi_p\gamma^X(-p) = \gamma^\eta(0) \\ (YW_1) \quad & \gamma^X(1) - \phi_1\gamma^X(0) - \dots - \phi_p\gamma^X(1-p) = 0 \\ & \vdots \\ (YW_j) \quad & \gamma^X(j) - \phi_1\gamma^X(j-1) - \dots - \phi_p\gamma^X(j-p) = 0 \end{aligned}$$

dites **équations de Yule-Walker**.

Les équations de Yule-Walker peuvent ensuite être utilisées pour l'estimation des paramètres d'un processus autorégressif. Notons $\hat{\phi} = (\hat{\phi}_1, \dots, \hat{\phi}_p)$ les estimateurs (dits de Yule-Walker) de $\phi = (\phi_1, \dots, \phi_p)$ et $\hat{\sigma}^2$ l'estimateur de σ^2 par cette méthode. Le théorème suivant donne la normalité asymptotique de ces estimateurs :

Théorème 2.11. Si $(X_t)_{t \in \mathbb{Z}}$ est un processus $AR(p)$ avec toutes ses racines de module supérieur à 1, alors :

$$\sqrt{T}(\hat{\phi} - \phi) \underset{d, T \rightarrow \infty}{\longrightarrow} \mathcal{N}(0, \sigma^2 \Gamma_p^{-1}),$$

avec $\Gamma_p = (\gamma(i-j))_{(i,j) \in [1,p]^2}$. De plus,

$$\hat{\sigma}^2 \underset{d, T \rightarrow \infty}{\longrightarrow} \sigma^2.$$

Note Rappelons que $\gamma^X(-j) = \gamma^X(j)$.

Théorème 2.12. *Etant donnés ϕ_1, \dots, ϕ_p et $\gamma^\eta(0)$, le système des $p + 1$ équations linéaires en les $p + 1$ inconnues $\gamma^X(0), \dots, \gamma^X(p)$ possède une unique solution. Les $\gamma^X(j)$ pour $j > p$ sont alors calculés par récurrence avec les YW_j , $j > p$.*

Théorème 2.13. *Etant donnés $\gamma^X(0), \dots, \gamma^X(p)$, le système des p équations linéaires en les p inconnues ϕ_1, \dots, ϕ_p possède une unique solution. L'équation YW_0 permet d'évaluer $\gamma^\eta(0)$.*

Dans la pratique (après avoir deviné une valeur raisonnable pour p), on calcule des valeurs empiriques des autocovariances $\gamma^X(0), \dots, \gamma^X(p)$. On en déduit les valeurs de ϕ_1, \dots, ϕ_p et $\gamma^\eta(0)$. Puis on cherche à valider le modèle en comparant avec les empiriques les valeurs des $\gamma^X(j)$ pour $j > p$ déduites des autres YW_j . Enfin, le modèle sera adapté (du point de vue de la prédiction) si $\gamma^\eta(0)$ est petit par rapport à $\gamma^X(0)$. Cela signifie que les racines (ou co-racines) de Φ doivent être proches du cercle unité.

2.4 Les processus mixtes ARMA

Les processus AR et MA ont des caractéristiques qui se révèlent grâce à leurs fonctions d'autocorrélations et leurs fonctions d'autocorrélations partielles. Pour un processus AR , nous avons vu que la fonction d'autocorrélation partielle possède un point de rupture après un certain nombre d'écart ; ce dernier détermine l'ordre du polynôme AR . Pour un processus MA , nous avons vu que c'est la fonction d'autocorrélation qui

possède un point de rupture après un certain nombre d'écarts ; ce dernier d'etermine l'ordre du polynôme MA . Cependant pour certains processus, ni la fonction d'autocorrélation, ni la fonction d'autocorrélation partielle ne possèdent de point de rupture. Dans de tels cas, il faut construire un modèle mixte. Nous définissons dans ce chapitre les séries $ARMA$ qui sont des généralisations directes des deux exemples introductifs, la combinaison des processus autorégressifs et moyennes mobiles. Cette classe de processus $ARMA$ est encore un cas particulier de processus linéaires et jouera un rôle prépondérant dans la modélisation concrète des processus stationnaires. Elle présente l'avantage d'être plus souple à l'utilisation et de fournir généralement de bonnes approximations des séries réelles avec moins de paramètres que les modèles purs. Nous aborderons finalement le problème de la prédiction sur ce modèle. En particulier, on exposera la méthode de Box-Jenkins qui est une des méthodes de prévision la plus couramment utilisée (en particulier sous R, SAS) en raison de sa simplicité, de l'économie de temps qu'elle permet et de la fiabilité de ses résultats.

2.4.1 Définition d'un $ARMA(p, q)$

Définition 2.7. On appelle *processus autorégressif moyenne mobile* d'ordre (p, q) tout processus $(X_t)_{t \in \mathbb{Z}}$ stationnaire tel que

$$\forall t \in \mathbb{Z}, \quad X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \eta_t - \theta_1 \eta_{t-1} - \dots - \theta_q \eta_{t-q} \quad (2.6)$$

où $\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q$ sont des réels fixés et $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc de variance σ^2 . Un tel processus est dit $ARMA(p, q)$ (AutoRegressive Moving Average of order (p, q)).

Posons comme précédemment $\Phi(B) = I - \phi_1 B - \dots - \phi_p B^p$ et $\Theta(B) = I - \theta_1 B - \dots - \theta_q B^q$.

On peut alors écrire

$$\forall t \in \mathbb{Z}, \quad \Phi(B)X_t = \Theta(B)\eta_t.$$

Remarque 2.4. *Il est immédiat qu'un ARMA(p, 0) est un AR pur et qu'un ARMA(0, q) est un MA pur. Les seuls processus admettant simultanément une représentation MA pure et une représentation AR pure correspondent au cas ARMA(0, 0), c'est-à-dire aux bruits blancs.*

2.4.2 Expression d'un ARMA(p, q)

a) Représentation moyenne mobile infinie

Propriété 2.1. (i) *Si le polynôme Φ a toutes ses racines de module différent de 1, l'opérateur $\Phi(B)$ est inversible et la relation admet une solution stationnaire donnée par*

$$\forall t \in \mathbb{Z}, \quad X_t = \frac{\Theta(B)}{\Phi(B)}\eta_t = \sum_{i \in \mathbb{Z}} \psi_i \eta_{t-i} \quad \text{avec} \quad \sum_{i \in \mathbb{Z}} |\psi_i| < +\infty,$$

c'est à dire sous une forme moyenne mobile infinie.

(ii) *Si de plus les racines du polynôme Φ sont de module strictement supérieur à 1, seules les valeurs présente et passées du bruit interviennent dans cette écriture MA(∞). Dans ce cas, les ψ_i de la représentation causale vérifient $\psi_0 = 1$ et*

$$\psi_k = \begin{cases} -\theta_k + \sum_{i=1}^k \phi_i \psi_{k-i} & \text{pour } 1 \leq k < \max(p, q+1) \\ \sum_{i=1}^p \phi_i \psi_{k-i} & \text{pour } k \geq \max(p, q+1) \end{cases}$$

en posant $\theta_0 = -1, \theta_i = 0$ si $i > q$ et $\phi_i = 0$ si $i > p$.

Ces équations peuvent se résoudre successivement pour $\psi_0, \psi_1 \dots$ Ainsi

$$\begin{aligned} \psi_0 &= 1 \\ \psi_1 &= -\theta_1 + \psi_0\phi_1 = -\theta_1 + \phi_1 \\ \psi_2 &= -\theta_2 + \psi_0\phi_2 + \psi_1\phi_1 = -\theta_2 + \phi_2 - \theta_1\phi_1 + \theta_1^2 \\ &\dots \end{aligned}$$

Exemple 2.6. On considère X défini par

$$X_t - X_{t-1} + \frac{1}{4}X_{t-2} = \eta_t + \eta_{t+1}$$

avec (η_t) un bruit blanc de variance σ^2 .

On a $\phi_1 = 1$, $\phi_2 = -\frac{1}{4}$, $\theta_1 = -1$, $p = 2$ et $q = 1$. En utilisant les équations précédentes, on obtient

$$\psi_0 = 1 \quad \psi_1 = -\theta_1 + \psi_0\phi_1 = -\theta_1 + \phi_1 = 2$$

et les autres valeurs de ψ_k s'obtiennent successivement à partir de l'équation

$$\psi_k - \psi_{k-1} + \frac{1}{4}\psi_{k-2} = 0 \quad k \geq 2$$

b) Représentation autorégressive

Une démarche analogue peut-être suivie pour le polynôme Θ :

Propriété 2.2. (i) Supposons que le polynôme Θ a toutes ses racines de module différent de 1, l'opérateur $\Theta(B)$ est inversible et on obtient la forme autorégressive infinie

$$\forall t \in \mathbb{Z}, \quad \eta_t = \frac{\Phi(B)}{\Theta(B)} X_t = \sum_{i \in \mathbb{Z}} \pi_i X_{t-i} \quad \text{avec} \quad \sum_{i \in \mathbb{Z}} |\pi_i| < +\infty$$

(ii) Si de plus les racines du polynôme Θ sont de module strictement supérieur à 1, cette représentation $AR(\infty)$ ne fait intervenir que les valeurs présente et passées du processus. Dans ce cas, les π_i de la représentation causale vérifient $\pi_0 = 1$ et

$$\pi_k = \begin{cases} -\phi_k + \sum_{i=1}^k \theta_i \pi_{k-i} & \text{pour } 1 \leq k < \max(p+1, q) \\ \sum_{i=1}^q \theta_i \pi_{k-i} & \text{pour } k \geq \max(p+1, q) \end{cases}$$

Remarque 2.5. Lorsque Φ et Θ ont des racines de module strictement supérieur à 1, les deux représentations ci-dessus sont vérifiées et η_t est non corrélé avec X_{t-1}, X_{t-2}, \dots et représente l'innovation du processus X_t à la date t .

2.4.3 Bruit blanc d'innovation d'un ARMA(p, q)

Comme pour les AR et les MA, le bruit blanc η n'est pas nécessairement le bruit blanc d'innovation :

Théorème 2.14. Pour que η soit le bruit blanc d'innovation de X il faut et il suffit que

$$\Phi(z) = \prod_{1 \leq k \leq p} (1 - \lambda_k z) \quad \text{et} \quad \Theta(z) = \prod_{1 \leq j \leq q} (1 - \mu_j z)$$

avec $\forall k, |\lambda_k| < 1$ et $\forall j, |\mu_j| \leq 1$.

Théorème 2.15. Si X est un ARMA(p, q) donné selon $\Phi(B)X = \Theta(B)\eta$, alors la relation (dite minimale) qui le lie à son bruit blanc d'innovation est aussi du type $\tilde{\Phi}(B)X = \tilde{\Theta}(B)\epsilon$, où les polynômes $\tilde{\Phi}$ et $\tilde{\Theta}$ s'obtiennent à partir de Φ et Θ selon les mêmes règles que pour les AR(p) et les MA(q) puis en supprimant tout facteur commun éventuel.

A partir de maintenant, nous ne considérerons plus que des processus ARMA(p, q) donnés sous leur représentation canonique i.e. vérifiant l'équation

$$\Phi(B)X_t = \Theta(B)\eta_t$$

avec

$$\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p = \prod_{i=1}^p (1 - \lambda_i z) \quad \text{où} \quad |\lambda_i| \leq 1 \forall i$$

$\Theta(z) = 1 - \theta_1 z - \dots - \theta_q z^q = \prod_{i=1}^q (1 - \mu_i z)$ où $|\mu_i| \leq 1 \forall i$. **puisque nous venons de voir qu'il est toujours possible de se ramener à ce cas quitte à changer de bruit blanc.**

2.4.4 Autocovariances et propriétés d'un ARMA(p, q)

Propriété 2.3. La fonction d'autocovariance d'un processus ARMA(p, q) vérifie pour $h \geq \max(p, q + 1)$

$$\gamma(h) = \sum_{i=1}^p \phi_i \gamma(h - 1) \tag{2.7}$$

et pour $0 \leq h < \max(p, q + 1)$,

$$\gamma(h) = \sum_{i=1}^p \phi_i \gamma(h - 1) + \sigma^2 (\psi_0 - \psi_1 \theta_{h+1} - \dots - \psi_{q-h} \theta_q) \tag{2.8}$$

Exercice Démonstration du résultat précédent

1. A partir de la définition de l'ARMA, montrer que

$$\gamma(h) = \sum_{i=1}^p \phi_i \gamma(h - i) + Cov(X_{t-h}, \eta_t) - Cov\left(X_{t-h}, \sum_{i=1}^q \theta_i \eta_{t-i}\right).$$

2. En déduire le résultat souhaité pour $h \geq q + 1$.

3. En utilisant la représentation moyenne mobile infinie pour X_{t-h} et la relation de la question 1), montrer le résultat pour $h \leq q$.

Ainsi, la suite des covariances satisfait une équation de récurrence d'ordre p à partir du rang $\max(p, q + 1)$ et les valeurs de $\gamma(h)$ pour $h = 0, \dots, p$ sont données par un système linéaire. On peut donc en déduire les valeurs de $\gamma(h)$ et donc $\rho(h)$ pour tout h :

En pratique

On trouve tout d'abord $\gamma(0), \dots, \gamma(p)$ à partir de ces équations avec $k = 0, 1, \dots, p$ puis on détermine $\gamma(p + 1), \gamma(p + 2), \dots$ de façon récursive. On complète ensuite par symétrie pour $\gamma(k)$ avec $k \leq 0$.

Exemple 2.7. On reprend l'exemple précédent avec X défini par

$$(I - B + \frac{1}{4}B^2)X_t = (I + B)\eta_t$$

avec (η_t) un bruit blanc de variance σ^2 .

Avec cette méthode on obtient

$$\gamma(0) = \frac{32}{3}\sigma^2, \quad \gamma(1) = \gamma(-1) = \frac{28}{3}\sigma^2, \quad \gamma(2) = \gamma(-2) = \frac{20}{3}\sigma^2$$

et les valeurs suivantes sont calculées récursivement à partir de l'équation $\gamma(k) = \gamma(k-1) - \frac{1}{4}\gamma(k-2)$ pour $k \geq 3$.

Proposition 2.9. La somme d'un $ARMA(p_1, q_1)$ et d'un $ARMA(p_2, q_2)$ (non corrélés) est un $ARMA(\leq \max(p_1, p_2), \leq \max(q_1, q_2))$.

Chapitre 3

Processus de Poisson

3.1 La loi de Poisson

On introduit une nouvelle loi de probabilité sur \mathbb{N} , dépendant d'un paramètre $\lambda > 0$, appelée **loi de Poisson** et notée $\mathcal{P}(\lambda)$: une variable aléatoire discrète suit cette loi si on a pour tout $n \in \mathbb{N}$

$$\mathbb{P}(X = n) = \frac{\lambda^n}{n!} \exp(-\lambda).$$

Rappelons que pour tout $x \in \mathbb{R}$, on a le développement de $\exp(x)$ en série : $\exp(x) = \sum_{k=0}^{+\infty} \frac{x^k}{k!}$. Donc on a bien défini une loi de probabilité sur \mathbb{N} .

L'espérance et la variance de cette loi sont toutes deux égales à λ .

De plus, la fonction caractéristique vérifie $\phi(t) = \exp(\lambda(\exp(it) - 1))$ pour tout $t \in \mathbb{R}$.

Exercice Vérifier ces résultats sur l'espérance, la variance, la fonction caractéristique.

Voici un résultat fameux, en exercice (facultatif, plus probabilités "élémentaires" que processus stochastiques) :

Exercice Soit $\lambda > 0$ un paramètre fixé.

Pour tout $n \in \mathbb{N}$, donnons X_n une variable aléatoire suivant une loi binomiale $\mathfrak{B}(n; \frac{\lambda}{n})$:

pour tout $0 \leq k \leq n$, on a

$$\mathbb{P}(X_n = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}.$$

Montrer, pour tout $k \in \mathbb{N}$ fixé, la convergence suivante lorsque $n \rightarrow +\infty$:

$$\mathbb{P}(X_n = k) \rightarrow \mathbb{P}(X = k);$$

où X suit une loi de Poisson de paramètre λ .

L'interprétation du résultat de l'exercice précédent se fait en termes d'événements rares : on compte le nombre d'occurrences d'événements indépendants, se réalisant chacun avec une probabilité $\frac{\lambda}{n}$, parmi un nombre de tirages égal à n , de telle sorte que en moyenne le nombre de réalisations soit constant égal à λ . Lorsque n tend vers $+\infty$, ce nombre de réalisations suit donc une loi de Poisson ; le paramètre λ est interprété comme une intensité : il détermine avec quelle proportion en moyenne l'événement doit se produire.

Au niveau des applications, on peut ainsi considérer qu'on subdivise un intervalle de temps fixé, en n sous-intervalles de taille identique. Sur chaque sous-intervalle, de façon indépendante, peut ou ne pas se produire un événement (arrivée d'un client, panne d'un appareil...), avec une probabilité proportionnelle à la longueur des sous-intervalles.

Plus généralement, on peut aussi découper un rectangle, un disque (par exemple une carte), et considérer comme événement un incendie, une panne...

Cette interprétation est à la base de la considération du processus de Poisson, où on ajoutera une composante dynamique pour faire évoluer le comptage au cours du temps.

3.2 La loi exponentielle

Elle a été définie lors des rappels de probabilités : une variable aléatoire réelle suit la loi exponentielle de paramètre $\lambda > 0$, notée $\mathcal{E}(\lambda)$, si pour tout $t > 0$, on a

$$\mathbb{P}(X > t) = \exp(-\lambda t) :$$

En particulier, on a $P(X \leq 0) = 0$: X ne prend que des valeurs positives.

Exercice Démontrer que $\mathbb{E}[X] = \frac{1}{\lambda}$ et $Var(X) = \frac{1}{\lambda^2}$. De plus, montrer que la fonction caractéristique de la loi $\mathcal{E}(\lambda)$ est donnée par $\varphi(t) = \frac{1}{\lambda - it}$. La propriété fondamentale est celle d'absence de mémoire : pour tout $t; s > 0$,

$$\mathbb{P}(X > t + s \mid X > s) = \mathbb{P}(X > t).$$

Dans la suite, on va utiliser des variables aléatoires de loi exponentielle pour modéliser le temps écoulé entre deux occurrences d'un événement.

3.3 Processus de comptage

Avant de spécifier le cas du processus de Poisson, on énonce quelques généralités sur la façon de modéliser le comptage d'événements.

Le processus est indexé par $\mathbb{T} = \mathbb{R}^+$, et est à valeurs dans \mathbb{N} . On se donne la suite des instants où se produit un événement (dans la suite, on parlera d'instant d'arrivées, par référence au cas d'une file d'attente) : il s'agit d'une suite ordonnée de variables aléatoires réelles $T_1 \leq T_2 \leq \dots \leq T_n, \dots$, sur laquelle on effectue les hypothèses suivantes :

Hypothèse 3.1. *Presque sûrement on a en fait un ordre strict : $0 = T_0 < T_1 < \dots < T_n < \dots$; de plus on a $T_n \rightarrow +\infty$, presque sûrement, quand $n \rightarrow +\infty$.*

On notera $T_0 = 0$ l'instant initial. Cela signifie que deux (ou plus) événements ne peuvent pas se produire en même temps, et qu'en un temps fixé T il ne peut se produire qu'un nombre fini de ces événements. On peut compter les événements produits sur l'intervalle de temps $[0; t]$, pour tout $t > 0$, en introduisant la quantité suivante :

$$N_t = \sum_{n=1}^{+\infty} \mathbb{1}_{T_n \leq t} = \sup\{n \in \mathbb{N}; T_n \leq t\}.$$

Remarquer que $N_0 = 0$.

Le processus à temps continu $N = (N_t)_{t \in \mathbb{N}}$ est à valeurs dans \mathbb{N} . Ses trajectoires sont en escalier, continues à droite, avec des sauts de hauteur 1 aux instants T_n (faire un dessin). La connaissance du processus N ou des temps d'arrivées sont équivalentes. On vient de voir comment N dépend des T_n ; graphiquement, les T_n sont les instants de saut des trajectoires.

On peut aussi noter les égalités d'événements suivantes :

$$\begin{aligned} \{N_t = n\} &= \{T_n \leq t < T_{n+1}\}; \\ \{T_n \leq t\} &= \{N_t \geq n\} \end{aligned}$$

En rajoutant des hypothèses sur les temps d'arrivées, ou de façon équivalente sur le processus, on va pouvoir caractériser la loi du processus.

3.4 Processus de Poisson

On introduit tout d'abord une hypothèse sur les temps d'arrivées :

Hypothèse 3.2. *Le processus des temps d'arrivées $(T_n)_{n \in \mathbb{N}}$ est un P.A.I.S., tel que $T_1, T_2 - T_1, \dots$ sont iid de loi exponentielle de paramètre $\lambda > 0$.*

Interprétation : les sources respectives des événements successifs sont indépendantes, et suivent toute la même loi exponentielle. En parallèle, on peut introduire l'hypothèse suivante sur le processus de comptage :

Hypothèse 3.3. *Le processus de comptage $(N_n)_{n \in \mathbb{R}^+}$ est un P.A.I.S : pour tout n -uplet (n quelconque) $t_1 < t_2 < \dots < t_n$, les accroissements $N_{t_1} - N_0, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}}$ sont indépendants et suivent pour lois respectives, celle de $N_{t_1(-0)}, N_{t_2-t_1}, \dots, N_{t_n-t_{n-1}}$.*

On a alors le résultat suivant (admis) :

Théorème 3.1. *On suppose l'hypothèse (3.1) vérifiée. Si l'hypothèse (3.2) est satisfaite, alors le processus de comptage associé est un P.A.I.S. Réciproquement, si l'hypothèse (3.3) est satisfaite, alors il existe $\lambda > 0$ tel que l'hypothèse (3.2) est vérifiée. Dans ce cas, pour tout $t \in \mathbb{R}^+$, la variable aléatoire N_t suit une loi de Poisson de paramètre λ_t . On parle de **processus de Poisson (homogène) d'intensité λ** .*

Exercice Montrer que la somme de deux va suivant des lois de Poisson $\mathcal{P}(\lambda), \mathcal{P}(\mu)$, suit une loi de Poisson dont on précisera le paramètre.

Remarque 3.1. *On a deux moyens de simuler un processus de Poisson :*

- par la donnée de la suite des temps d'arrivées (tirages indépendants selon la loi $\mathcal{E}(\lambda)$ des temps entre deux arrivées successives) ;
- si on a fixé un temps final t : on commence par tirer aléatoirement un nombre $N_t = k$ selon la loi de Poisson $\mathcal{P}(\lambda t)$; puis on tire des nombres u_1, \dots, u_k , indépendants, uniformément sur l'intervalle $[0; t]$ (ils sont presque sûrement tous distincts) ; on les trie en ordre croissant, ce qui donne les temps d'arrivée successifs. Le deuxième moyen repose sur une démonstration (admise) que ce procédé fonctionne...

Remarque 3.2. tirer un nombre aléatoirement selon une loi donnée=générer (par un certain mécanisme) une réalisation d'un variable aléatoire, de la loi en question.

3.5 Quelques propriétés du processus de Poisson

Exercice Soit $N = (N_t)_{t \geq 0}$ un processus de Poisson, et soit $T > 0$ un instant fixé.

Montrer que le processus $N^{(T)}$ tel que $N_t^{(T)} = N_{T+t} - N_T$, est un autre processus de Poisson, indépendant du processus $(N_s)_{0 \leq s \leq T}$ - c'est-à-dire de tout vecteur aléatoire $(N_{s_1}, \dots, N_{s_n})$ avec $0 \leq s_1 \leq \dots \leq s_n \leq T$.

Estimation de l'intensité : $\frac{Nt}{t} \rightarrow \lambda$, quand $t \rightarrow +\infty$, presque sûrement.

3.6 Processus de Poisson marqué

On considère toujours une suite de variables aléatoires exponentielles indépendantes, modélisant le temps entre deux instants d'occurrence d'un événement (tel l'arrivée d'un client dans une file d'attente).

Maintenant, on va diviser en deux (il est possible de faire plus mais ça suffira à expliquer comment ça marche) les possibilités. Par exemple, les clients sont soit des hommes, soit des femmes. Et on ne va compter qu'un seul type de possibilité.

Autrement dit, à un instant T_n , on regarde la réalisation d'une variable aléatoire X_n , indépendante du reste, prenant la valeur 1 (le type qu'on compte) avec une probabilité p , et la valeur 0 (celui qu'on élimine) avec la probabilité $q = 1 - p$. Et on ajoute la valeur de X_n au compteur.

Mathématiquement, on se donne donc des variables $0 = T_0 \leq T_1 \leq \dots \leq T_n$ vérifiant l'hypothèse (3.2), tel que les variables $(T_{i+1} - T_i)_{i \in \mathbb{N}}$ sont iid de loi exponentielle de paramètre

λ . On se donne aussi une suite $(X_n)_{n \in \mathbb{N}^*}$ de va iid, de loi de Bernoulli de paramètre p . Les suites $(T_n)_{n \in \mathbb{N}^*}$ et $(X_n)_{n \in \mathbb{N}^*}$ sont également supposées indépendantes. On a ainsi 3 processus, définis par les formules :

- $N_t = \sum_{n=1}^{+\infty} \mathbb{1}_{T_n \leq t}$ quand on compte tous les types ;
- $N_t^1 = \sum_{n=1}^{+\infty} X_n \mathbb{1}_{T_n \leq t}$ quand on compte uniquement le type 1 ;
- $N_t^0 = \sum_{n=1}^{+\infty} (1 - X_n) \mathbb{1}_{T_n \leq t}$ quand on compte uniquement le type 0.

On a $N_t = N_t^0 + N_t^1$. De plus, les processus N^0 et N^1 sont deux processus de Poisson indépendants, d'intensités respectives $(1 - p)\lambda$ et $p\lambda$.

3.7 Processus de Poisson composé

Dans la section précédente, par rapport au Processus de Poisson (simple), on s'est donné la possibilité de sélectionner ou non chaque événement dans notre compteur. Pour cela, on a introduit une suite de va iid $(X_n)_{n \in \mathbb{N}^*}$ dont les valeurs déterminent la prise en compte ou non de l'événement. On observe qu'on obtient encore un Processus de Poisson simple d'intensité différente ; en particulier, les sauts sont tous de taille 1.

Maintenant, on veut que les sauts puissent prendre d'autres valeurs, aléatoires. Il suffit pour cela de considérer une suite de va iid $(Y_n)_{n \in \mathbb{N}^*}$, de loi quelconque (par exemple gaussienne), et de poser

$$N_t^Y = \sum_{n=1}^{+\infty} Y_n \mathbb{1}_{T_n \leq t}.$$

Il est important de noter que N_t ne prend plus uniquement des valeurs entières. Les Processus de Poisson simple et marqué sont deux cas particuliers.

3.8 Processus Markovien de saut

Dans les deux raffinements précédents du Processus de Poisson simple, on a modifié la loi des sauts. Mais on n'a pas modifié la loi des temps de saut : la suite des durées entre deux sauts successifs est constituée de τ_n indépendantes, de loi exponentielle, de même paramètre.

On va maintenant modifier une condition (et pas les autres) : celle que le paramètre est le même pour toutes les τ_n exponentielles.

On va aussi supposer que l'espace d'états du processus $(X_t)_{t \in \mathbb{R}^+}$ ainsi obtenu est fini ou dénombrable, il sera noté S . Plutôt que de parler des sauts de ces processus, il est commode de garder un point de vue "position", qui change (sans possibilité de rester sur place !) aux "instants de saut".

Un cas particulier est celui des processus de naissance et de mort : à valeurs dans \mathbb{N} , dont tous les sauts sont $+1$ ou -1 .

Le fait de conserver l'indépendance entre les sauts et les temps de saut, ainsi que l'hypothèse d'accroissements stationnaires et de loi exponentielle sur ces temps de saut, est lié à la propriété de Markov (en temps continu) : on veut que le futur du processus ne dépende du passé qu'à travers l'état présent.

On ne va pas détailler précisément comment exprimer cette condition sur nos processus, on va se contenter de décrire la dynamique (de deux façons équivalentes).

Chapitre 4

Processus de Wiener

Le processus de Wiener porte le nom du mathématicien américain qui en donna une définition axiomatique en 1923. Il est l'aboutissement de la modélisation mathématique du mouvement brownien mis en évidence un siècle auparavant par le botaniste écossais Brown, grâce aux observations qu'il fit du mouvement erratique de particules de pollen s'agitant à la surface d'un liquide chauffé. C'était le début d'une saga physico-mathématique particulièrement riche dont on donne un aperçu en fin de chapitre. Le processus de Wiener est un processus stationnaire à accroissements indépendants (voir la définition des processus de Poisson au 3.4) caractérisé par la normalité de ses élongations.

Un processus à accroissements indépendants noté PAI, et PAIS s'il est de plus SSL.

Théorème 4.1. *Tout PAIS $(X_t)_{t \in \mathbb{R}}$ centré, tel que $X_0 = 0$, vérifie les propriétés :*

(1) $\forall t : V(X_t) = K|t| \quad (K > 0) ;$

$$(2) \quad cov(X_s, X_t) = \begin{cases} K \cdot \min(|s|, |t|) & \text{si } s, t > 0; \\ 0 & \text{si } s, t < 0; \end{cases}$$

(3) $\forall s, t \text{ tels que } t > s > 0 : \quad \rho(s, t) = \sqrt{\frac{s}{t}}.$

Preuve (1) $\forall t_1 t_2 > 0$,

$$\begin{aligned} \text{Var}(X_{t_1+t_2}) &= \text{Var}(X_{t_1+t_2} - X_{t_1} + X_{t_1} - X_0) \\ &= (\text{indépendance des acc}) V(X_{t_1+t_2} - X_{t_1}) + V(X_{t_1} - X_0) \\ &= (\text{stationnarité}) V(X_{t_2}) + V(X_{t_1}); \end{aligned}$$

si on pose $V(X_t) = f(t)$, l'équation précédente s'écrit :

(*) $f(t_1 + t_2) = f(t_1) + f(t_2)$, où f est positive donc par (*), croissante. La seule fonction solution de (*) est linéaire : $f(t) = Kt$. Un raisonnement analogue pour des t_i négatifs permet de conclure.

(2) Pour calculer la covariance distinguons trois cas :

$$\begin{aligned} (a) \quad 0 \leq s \leq t : \text{cov}(X_s, X_t) &= \text{cov}(X_s, (X_t - X_s) + X_s) \\ &= \text{cov}(X_s, (X_t - X_s)) + \text{cov}(X_s, X_s) \\ &= \text{cov}(X_s - X_0, X_t - X_s) + K.s = K.s \end{aligned}$$

car les accroissements $(X_s - X_0, X_t - X_s)$ étant indépendants ;

(b) $s \leq t \leq 0$: idem ; on parvient à $\text{cov}(X_s, X_t) = -Kt$;

(c) $s \leq 0 \leq t$: $\text{cov}(X_s, X_t) = -\text{cov}(X_0 - X_s, X_t - X_0) = 0$.

(3) est une conséquence de (1) et (2).

Définition 4.1. Le processus de Wiener $(W_t)_{t \geq 0}$ sur \mathbb{R}^+ est défini par les axiomes :

(1) $W_0 = 0$;

(2) $(W_t)_{t \geq 0}$ est un PAIS ;

(3) $\forall t, W_t$ est de loi $\mathcal{N}(0, \sigma^2 t)$. Si $\sigma = 1$, le processus de Wiener est dit standard.

Le théorème ci-dessous donne un ensemble de conditions suffisantes pour qu'un processus gaussien soit un processus de Wiener.

Théorème 4.2. *Un processus gaussien $(W_t)_{t \geq 0}$ est un processus de Wiener si :*

$$(1) \forall t \quad \mathbb{E}(W_t) = 0, V(W_t) = \sigma^2 t;$$

$$(2) \forall s, t \quad R(s, t) = \sigma^2 \min(s, t).$$

L'exercice ci-dessous montre qu'un processus de Wiener peut être approché asymptotiquement par une marche aléatoire symétrique.

Exercice Approximation du processus brownien à une dimension par la marche aléatoire symétrique

Soit un processus de Bernoulli $(X_n)_{n \in \mathbb{N}}$ où les X_n sont indépendantes de même loi : $\forall n \in \mathbb{N}, \mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = 0,5$. Pour tout $t \geq 0$, définissons $Z_t = \Delta x (X_1 + \dots + X_{[\frac{t}{\Delta t}]})$, où Δx et Δt sont des accroissements très petits de x et de t ; $[\frac{t}{\Delta t}]$ représente la partie entière de $\frac{t}{\Delta t}$.

(1) Démontrer que le processus $(Z_t)_t$ est un PAIS.

(2) Démontrer que Z_t converge en loi vers la v.a. normale $\mathcal{N}(0, \sigma^2 t)$ et conclure. [Indication : faire tendre Δx et Δt vers 0 de manière que $\frac{\Delta x}{\sqrt{\Delta t}} = \sigma$].

D'un point de vue physique, $\sigma = \sqrt{2D}$ où D est le coefficient de diffusion de la particule brownienne dans son milieu.

On généralise le processus de Wiener aux espaces \mathbb{R}^2 et \mathbb{R}^3 . Par exemple, lorsque l'espace d'états est \mathbb{R}^2 , $(W_t)_t$ est le modèle d'évolution d'une particule évoluant à la surface d'un liquide, et soumise aux impacts incessants de ses voisines, qui comme on le sait sont fort nombreuses et dissipées. Les composantes sur chacun des axes d'un repère orthonormé d'un processus de Wiener de dimension quelconque dans tout repère sont aussi des processus de Wiener de dimension 1.

Théorème 4.3. (Propriétés du processus de Wiener)

(1) Le processus de Wiener est gaussien et ergodique par rapport à la moyenne.

(2) Étant donné s, t tels que $t > s$, la densité conditionnelle de W_t sachant $W_s = x$, est définie par :

$$f_{W_t/(W_s=x)}(y) = \frac{1}{\sigma\sqrt{2\pi(t-s)}} \exp\left(-\frac{(y-x)^2}{2\sigma^2(t-s)}\right);$$

$$\mathbb{E}(W_t|W_s = x) = x; \quad V(W_t|W_s = x) = \sigma^2(t-s).$$

À partir de toute abscisse d'élongation x , atteinte au temps s , l'élongation du processus est de loi gaussienne. L'égalité $\mathbb{E}(W_t|W_s = x) = x$ signifie que l'élongation moyenne de W_t , autour d'un point d'abscisse quelconque x atteint en un instant s précédent est nulle; il faut imaginer cela pour toutes les positions x atteintes par le processus. Si l'on pose $Z_t = W_{s+t}$, $(Z_t)_{t \geq 0}$ est le processus de Wiener démarré au temps 0 en W_s égal à x . L'égalité $V(W_t|W_s = x) = \sigma^2(t-s)$ signifie que la dispersion moyenne quadratique autour de ce même point est fonction linéaire du temps $(t-s)$. Pour tout t , W_t se trouve à l'intérieur ou au voisinage de la parabole d'équation $x^2 = (1,96)^2\sigma^2 t$, et la probabilité $\mathbb{P}(|W_t| \geq 1,96\sigma\sqrt{t})$ est approximativement égale à 0,95.

(3) $(W_t)_t$ est continu (p.s.) et nulle part dérivable (p.s.).

(Les deux propriétés précédentes font de ce processus un candidat rêvé pour la galerie des notions mathématiques longtemps considérées comme « monstrueuses »; c'est ainsi que certains mathématiciens de la fin du XIX^e siècle, et non les moindres, concevaient les fonctions continues et non dérivables, dont des exemples fameux ont été exhibés par Weierstrass et Lebesgue.)

Chapitre 5

Intégrale d'Itô et Processus de diffusion

5.1 Processus de diffusion

Dans ce chapitre, on s'intéresse aux propriétés markoviennes des solutions des E.D.S. On se limitera au cas homogène ce qui simplifiera les notations mais ce qui est relativement regrettable. Cependant si l'on a bien compris le cas homogène, il est assez facile de se reporter à la littérature traitant du cas général.

5.1.1 Espace mesurable.

Soit (E, \mathcal{E}) un espace mesurable tel que, pour tout $x \in E$, $\{x\} \in \mathcal{E}$.

Définition 5.1. On appelle processus de Markov (homogène) à valeurs (E, \mathcal{E}) un terme $X =$

$(\Omega, \mathcal{F}_t, \mathcal{F}, X_t, (\mathbb{P}_x)_{x \in E})$ où

(i) pour tout $x \in E$, \mathbb{P}_x est une probabilité sur (Ω, \mathcal{F}) ,

(ii) pour tout $A \in \mathcal{F}$, $x \mapsto \mathbb{P}_x(A)$ est \mathcal{E} -mesurable,

(iii) pour tout $t \geq 0$, X_t est une v.a. \mathcal{F}_t -mesurable à valeurs (E, \mathcal{E}) ,

(iv) pour tout $x \in E$, $\mathbb{P}_x(X_0 = x) = 1$,

(v) pour tous $t, h \geq 0$, tout $x \in E$, tout $\Gamma \in \mathcal{E}$,

$$\mathbb{P}_x(X_{t+h} \in \Gamma \mid \mathcal{F}_t) = \mathbb{P}_{X_t}(X_h \in \Gamma), \quad \mathbb{P}_x p.s. \tag{5.1}$$

Comme d'habitude, notant \mathbb{E}_x l'espérance pour \mathbb{P}_x , (v) implique que, pour toute $f \in \mathcal{E}^+$, $\mathbb{E}_x(f(X_{t+h}) \mid \mathcal{F}_t) = \mathbb{E}_{X_t}(f(X_h))$. On pose alors, pour $f \in \mathcal{E}^+$,

$$P_t f(x) = \mathbb{E}_x(f(X_t)). \tag{5.2}$$

P_t vérifie $P_t 1 = 1$ et, pour $f, g \in \mathcal{E}^+$ et $\alpha \geq 0$, $P_t(\alpha f) = \alpha P_t f$, $P_t(f + g) = P_t f + P_t g$. C'est un opérateur linéaire de $b\mathcal{E}$ dans $b\mathcal{E}$. Evidemment $P_t f(x) = \int P_t(x, dy) f(y)$ où $P_t(x, \Gamma) = \mathbb{P}_x(X_t \in \Gamma)$. La famille $(P_t(x, dy), t \geq 0)$ s'appelle la **fonction de transition** de X . Pour μ probabilité sur (E, \mathcal{E}) , on définit $\mathbb{P}\mu(A) = \int \mathbb{P}_x(A) d\mu(x)$ et l'on a encore, notant \mathbb{E}_x l'espérance pour \mathbb{P}_x ,

$$\mathbb{E}_\mu(f(X_{t+h}) \mid \mathcal{F}_t) = \mathbb{E}_{X_t}(f(X_h)), \quad \mathbb{P}_\mu p.s. \tag{5.3}$$

Proposition 5.1. Soit X un processus de Markov à valeurs (E, \mathcal{E}) de fonction de transition P_t . On a, pour tout $x \in E$, tous $0 \leq t_1 < \dots < t_n$, toute $f \in (\mathcal{E}^{\otimes n})^+$,

$$\mathbb{E}_x(f(X_{t_1}, \dots, X_{t_n})) = \int P_{t_1}(x, dx_1) \dots P_{t_n - t_{n-1}}(x_{n-1}, dx_n) f(x_1, x_2, \dots, x_n). \tag{5.4}$$

Utilisant la propriété de semi-groupe :

$$P_{s+t} = P_s P_t = P_t P_s. \tag{5.5}$$

5.1.2 Semi-groupe de Feller.

Rappelons que l'on note $C_0 = C_0(\mathbb{R}^d)$ l'espace des fonctions continues sur \mathbb{R}^d tendant vers 0 à l'infini qu'on munit de la norme $\|f\| = \sup_x |f(x)|$.

Définition 5.2. On appelle semi-groupe de Feller sur \mathbb{R}^d une famille $(P_t, t \geq 0)$ d'opérateurs linéaires positifs de C_0 dans C_0 vérifiant :

- (i) $P_0 = I$ et, pour tout $t \geq 0, P_t 1 = 1$,
- (ii) pour tous $s, t \geq 0, P_{s+t} = P_s P_t = P_t P_s$,
- (iii) pour toute $f \in C_0, \|P_t f - f\| \rightarrow 0$ lorsque $t \downarrow 0$.

On a donc $|P_t f(x)| \leq \|f\| P_t 1(x) = \|f\|$ d'où $\|P_t f\| \leq \|f\|$.

Définition 5.3. Soit $(P_t, t \geq 0)$ un semi-groupe de Feller sur \mathbb{R}^d . On dit que $f \in \mathcal{D}_A$ et que $Af = g$ s'il existe $g \in C_0$ telle que $\|\frac{1}{t}(P_t f - f) - g\| \rightarrow 0$ lorsque $t \downarrow 0$.

L'opérateur (\mathcal{D}_A, A) s'appelle **le générateur infinitésimal** de P_t . Donnons quelques propriétés élémentaires.

Proposition 5.2. Si $f \in \mathcal{D}_A, \frac{d}{dt} P_t f$ existe et vaut $P_t Af$ et l'on a $P_t f(x) - f(x) = \int_0^t P_s Af(x) ds$.

5.1.3 Processus de diffusion.

Soient $a_{i,j}(x), b_i(x), 1 \leq i, j \leq d$, des fonctions mesurables localement bornées sur \mathbb{R}^d . On suppose que, pour tout $x \in \mathbb{R}^d$,

$$a_{i,j}(x) = a_{j,i}(x), \sum_{i,j=1}^d (x)\theta_i\theta_j \geq 0 \text{ quel que soit } \theta \in \mathbb{R}^d.$$

La matrice $a(x) = (a_{i,j}(x), 1 \leq i, j \leq d)$ est donc symétrique semi-définie positive. On pose

$$L = \frac{1}{2} \sum_{i,j=1}^d a_{i,j}(x) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^d b_i(x) \frac{\partial}{\partial x_i}. \quad (5.6)$$

Définition 5.4. On appelle processus de diffusion de générateur L donné par (5.6) un processus de Markov $X = (\Omega, \mathcal{F}_t, \mathcal{F}, X_t, (\mathbb{P}_x)_{x \in E})$ à valeurs \mathbb{R}^d et à trajectoires continues tel que, pour tout $x \in \mathbb{R}^d, t \geq 0$ et $f \in C_k^2$,

$$\mathbb{E}_x(f(X_t)) - f(x) = \mathbb{E}_x\left(\int_0^t Lf(X_s) ds\right). \quad (5.7)$$

Chapitre 6

Estimation paramétrique des processus stationnaires

6.1 L'approche de Box et Jenkins

Les différents paramètres d'un processus *ARMA* sont les suivants :

- les coefficients du polynôme Φ ;
- les coefficients du polynôme Θ ;
- la variance σ^2 du bruit blanc.

En fait, implicitement, il y a avant tout l'ordre (p, q) du processus *ARMA* à déterminer.

Nous abordons maintenant la méthode de Box-Jenkins qui est une des méthodes de traitement des processus *ARMA* la plus couramment utilisée (en particulier sous *R*, *SAS*) en raison de sa simplicité, de l'économie de temps qu'elle permet et de la fiabilité de ses résultats. On décrit dans cette section une démarche générale pour l'ajustement de données à l'aide d'un modèle *ARMA*. Dans la suite, on suppose que l'on dispose de T observations X_1, \dots, X_T d'une série chronologique stationnaire. Les étapes de cette démarche sont les suivantes :

1. Estimation de la tendance : on commence par estimer la moyenne

$$\hat{Z}_t = \frac{1}{T} \sum_{i=1}^T X_i,$$

puis on centre la série d'observations le cas échéant.

2. Estimation des fonctions d'autocorrélation et d'autocorrélation partielle : on estime la fonction d'autocorrélation ρ par $\hat{\rho}_T$ et d'autocorrélation partielle τ par $\hat{\tau}_T$. Il est également intéressant d'étudier les autocorrélations inverses : si $(X_t)_{t \in \mathbb{Z}}$ est un processus $ARMA(p, q)$ solution de l'équation $\Phi(B)X_t = \Theta(B)\eta_t$ et si Θ est inversible, alors l'équation $\Phi(B)X_t = \Theta(B)\eta_t$ définit un processus dual $ARMA(q, p)$ dont on peut étudier la fonction d'autocorrélation, appelée **fonction d'autocorrélation inverse**.

3. Vérifications des hypothèses : si $\hat{\rho}_T(h)$ et $\hat{\tau}_T(h)$ ne tendent pas vers 0 quand h croît, alors on rejette l'hypothèse d'une modélisation par un processus $ARMA$. Dans ce cas, on peut essayer de transformer les données à l'aide d'outils classiques : différenciation, différenciation saisonnière, transformation linéaire, exponentielle, transformation de Box-Cox (incluant la transformation logarithmique)... en effectuant un minimum de transformations.

4. Détermination de p_{max} et q_{max} : si l'étape précédente s'est bien passée, on cherche alors à modéliser les données par un $MA(q_{max})$ puis un $AR(p_{max})$ avec p_{max} et q_{max} maximum en s'aidant des autocorrélations et des autocorrélations partielles. On considèrera par la suite l'ensemble des modèles $ARMA(p, q)$ avec $0 \leq p \leq p_{max}$ et $0 \leq q \leq q_{max}$.

5. Estimation des paramètres : pour chaque processus $ARMA$ d'ordre (p, q) choisi, on estime les différents paramètres du modèle : les coefficients ϕ_1, \dots, ϕ_p (partie AR), les coefficients $\theta_1, \dots, \theta_q$ (partie MA) et la variance σ^2 du processus d'innovation.

6. Amélioration de l'estimation des paramètres : si on peut émettre l'hypothèse de bruit blanc gaussien (pour cela, on peut utiliser un test adéquat), on est alors dans le cas d'un

processus gaussien et les estimateurs peuvent être affinés par la méthode du maximum de vraisemblance (algorithme de Box et Jenkins).

7. Sélection de modèles : on compare les différents modèles. Pour cela, on dispose de plusieurs critères tels que *AIC* (Akaike Information Criterion), le *SBC/BIC* (Schwarz Bayesian Criterion ou Bayesian Information Criterion) et le critère de Hannan.

6.1.1 Pré-sélection de modèles

Si les données semblent provenir d'un processus stationnaire, on cherche alors à modéliser les données par un $MA(q_{max})$ puis un $AR(p_{max})$ avec p_{max} et q_{max} maximum.

Détermination de q_{max} Si $(X_t)_{t \in \mathbb{Z}}$ est un processus $MA(q)$, on montre que pour $h > q$ fixé,

$$\sqrt{T} \hat{\rho}_T(h) \underset{d, T \rightarrow \infty}{\longrightarrow} \mathcal{N} \left(0, \sum_{|k| \leq q} \hat{\rho}_T(k)^2 \right).$$

On peut donc en déduire un intervalle de confiance pour $\hat{\rho}_T(h)$ et $h > q$:

$$\mathbb{P} \left(|\hat{\rho}_T(h)| \leq z_{\alpha/2} \sqrt{\frac{1}{T} \sum_{|k| \leq q} \hat{\rho}_T(k)^2} \right) = 1 - \alpha,$$

avec $\alpha \in]0, 1[$ et $z_{\alpha/2}$ le quantile d'ordre $\alpha/2$ de la loi normale centrée réduite. En pratique, si "trop" de valeurs de $\hat{\rho}_T$ sont en dehors de l'intervalle de confiance pour $h > q$, alors on peut rejeter l'hypothèse d'une modélisation $MA(q)$. On choisit donc q_{max} le plus grand possible tel qu'un processus $MA(q_{max})$ puisse être ajusté aux données.

Détermination de p_{max} Si $(X_t)_{t \in \mathbb{Z}}$ est un processus $AR(p)$, on montre que pour $h > p$ fixé,

$$\sqrt{T} \hat{\tau}_T(h) \underset{d, T \rightarrow \infty}{\longrightarrow} \mathcal{N}(0, 1).$$

On peut donc en déduire un intervalle de confiance pour $\hat{\tau}_T(h)$ et $h > p$:

$$\mathbb{P} \left(|\hat{\tau}_T(h)| \leq \frac{z_{\alpha/2}}{\sqrt{T}} \right) = 1 - \alpha,$$

avec $\alpha \in]0, 1[$ et $z_{\alpha/2}$ le quantile d'ordre $\alpha/2$ de la loi normale centrée réduite. En pratique, si "trop" de valeurs de $\hat{\tau}_T$ sont en dehors de l'intervalle de confiance pour $h > p$, alors on peut rejeter l'hypothèse d'une modélisation $AR(p)$. On choisit donc p_{max} le plus grand possible tel qu'un processus $AR(p_{max})$ puisse être ajusté aux données.

6.1.2 Estimation des paramètres

Les paramètres d'un processus $ARMA$ peuvent être estimés à l'aide de la fonction d'autocorrélation. Or, on dispose d'estimateurs convergents pour la fonction d'autocorrélation. On peut donc en déduire des estimateurs convergents pour les paramètres d'un processus $ARMA$.

Nous allons procéder comme pour les processus autorégressifs à partir des équations de Yule-Walker puisqu'il existe également des équations de Yule-Walker qui sont adaptées aux processus $ARMA$. En utilisant l'équation (2.6) dans l'espérance de $X_t X_{t-h}$, on obtient l'équation suivante :

$$\sum_{i=0}^p \phi_i \mathbb{E}(X_{t-i} X_{t-h}) = \sum_{j=0}^q \theta_j \mathbb{E}(\eta_{t-j} X_{t-h}).$$

Or, pour tout $h > 1$, $\mathbb{E}(\eta_{t-j} X_{t-h}) = 0$. D'où, la fonction d'autocorrélation ρ vérifie l'équation récurrente ci-dessous :

$$\forall h \geq q + 1, \quad \rho(h) + \phi_1 \rho(h - 1) + \dots + \phi_p \rho(h - p) = 0.$$

D'où, pour $q + 1 \leq h \leq q + p$, on obtient un système de p équations linéaires en ϕ_1, \dots, ϕ_p (on rappelle que $\phi_0 = 0$) qui est le suivant :

$$\begin{pmatrix} \rho(q+1) \\ \rho(q+2) \\ \vdots \\ \rho(q+p) \end{pmatrix} = \begin{pmatrix} \rho(q) & \rho(q-1) & \dots & \rho(q-p+1) \\ \rho(q+1) & \rho(q) & \dots & \rho(q-p+2) \\ \vdots & \vdots & \dots & \vdots \\ \rho(q+p-1) & \rho(q+p-2) & \dots & \rho(q) \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix}$$

Remarque 6.1. Si le processus est un AR pur (i.e. si $q = 0$) alors on retrouve les équations de Yule-Walker déjà vues en utilisant la symétrie de la fonction d'autocorrélation.

Ainsi on peut en exprimer ϕ_1, \dots, ϕ_p à l'aide de la fonction d'autocorrélation et en déduire des estimateurs de ces paramètres en remplaçant la fonction ρ par son estimation $\hat{\rho}$. On obtient donc une estimation des paramètres de la partie autorégressive d'un processus ARMA.

Il reste donc à estimer les paramètres $\theta_1, \dots, \theta_q$ et σ^2 . Désormais on suppose que les coefficients ϕ_1, \dots, ϕ_p sont connus ou estimés. On pose :

$$Y_t = \sum_{j=0}^p \phi_j X_{t-j}.$$

La covariance entre Y_t et Y_{t+h} vaut :

$$\rho_Y(h) = \mathbb{E}(Y_t Y_{t+h}) = \sum_{j=0}^p \sum_{i=0}^p \phi_i \phi_j X_{t-j} X_{t+h-i} = \sum_{j=0}^p \sum_{i=0}^p \phi_i \phi_j \rho_{h+j-i}.$$

Elle ne dépend donc que de paramètres connus ou pouvant être estimés. Par ailleurs, en partant du membre de droite de l'équation (2.6), on obtient que :

$$\rho_Y(h) = \mathbb{E} \left[\left(\sum_{j=0}^q \theta_j \eta_{t-j} \right) \left(\sum_{i=0}^q \theta_i \eta_{t+j-i} \right) \right].$$

Comme $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc faible de variance σ^2 , on obtient que :

$$\forall 0 \leq h \leq q, \quad \rho_Y(h) = \sigma^2 \sum_{j=h}^q \theta_j \theta_{j-h},$$

et, pour tout $h > q$, $\rho_Y(h) = 0$. Alors, en posant $x_h = \sigma \theta_h$ pour tout $0 \leq h \leq q$, le système devient :

$$\forall 0 \leq h \leq q, \quad \rho_Y(h) = \sigma^2 \sum_{j=h}^q x_j x_{j-h},$$

On a donc un système d'équations quadratiques. Dès que $q \geq 2$, il n'existe pas de formules explicites pour les solutions d'un tel système. On sait qu'il existe un nombre fini de solutions distinctes : dans le cadre de ce cours, on ne s'intéressera uniquement aux solutions menant à un polynôme ayant toutes ses racines de module supérieur à 1. Box et Jenkins préconisent l'utilisation de la méthode de Newton pour résoudre numériquement le système.

6.1.3 Sélection de modèles

Une fois choisi l'ensemble des processus *ARMA* permettant de modéliser les données, il y a donc $(p_{max} + 1)(q_{max} + 1)$ modèles possibles et il ne faut en retenir qu'un seul. Pour cela, on dispose de critères de comparaison de modèles que l'on souhaite être le plus petit possible. Les critères classiques sont les suivants :

- Critère d'Akaike (*AIC*) :

$$AIC(p, q) = \log \hat{\sigma}^2 + \frac{2(p+q)}{T}.$$

- Critère de Schwarz (SBC ou BIC) :

$$SBC(p, q) = \log \hat{\sigma}^2 + \frac{2(p+q) \log T}{T}.$$

- Critère de Hannan :

$$\varphi(p, q) = \log \hat{\sigma}^2 + \frac{c(p+q) \log \log T}{T},$$

avec $c > 2$.

Pour ces trois critères, $\hat{\sigma}^2$ est l'estimation de la variance σ^2 du bruit blanc. Ces critères sont basés sur le principe de la pénalisation. En effet, on peut montrer que si on passe d'un modèle *ARMA*(p, q) à un modèle *ARMA*($p+1, q$) ou à un modèle *ARMA*($p, q+1$), alors la variance estimée diminue. Ces critères corrigent donc ce phénomène et on cherche un

compromis entre une faible variance et un faible nombre de paramètres à estimer, respectant ainsi le principe de parcimonie (retenir le modèle le moins complexe pour espérer une qualité d'ajustement acceptable).

6.1.4 Test sur les résidus

Les tests, concernant le bruit blanc, ont pour but de vérifier si l'estimation des résidus est bien cohérente avec les hypothèses relatives au bruit blanc.

L'analyse de la fonction d'autocorrélation permet de vérifier si on est en présence d'un bruit blanc. Lorsque T est assez grand, les autocorrélations d'un bruit blanc sont approximativement indépendantes et de loi $\mathcal{N}(0, \frac{1}{T})$. Ainsi 95% des autocorrélations devraient se trouver dans l'intervalle

$$\left[-1.96/\sqrt{T}; 1.96/\sqrt{T} \right].$$

Plutôt que de regarder si chaque autocorrélation est dans les bornes de l'intervalle précédent, on peut aussi effectuer des tests sur la loi du processus étudié. Un des tests le plus utilisé est celui du portemanteau proposé par Box et Pierce : on cherche à tester l'indépendance des résidus. On note $\hat{\rho}_{\hat{\eta}}$ la fonction d'autocorrélation empirique des résidus estimés $\hat{\eta}$. On considère la statistique de Box-Pierce Q_{BP} définie par la somme des H premières autocorrélations empiriques au carré :

$$Q_{BP} = \sum_{h=1}^H \hat{\rho}_{\hat{\eta}}(h)^2.$$

D'après la remarque précédente sur la normalité des autocorrélations, on montre que, sous l'hypothèse nulle, si $T > H > p + q$, alors :

$$Q_{BP} \underset{d}{\xrightarrow{T \rightarrow \infty}} \chi^2(H - p - q).$$

L'indice H est généralement choisi entre 15 et 20 (en pratique, on ne considère pas des processus ARMA d'ordre très élevé).

Cette statistique de test peut être améliorée (dans le sens où le test est plus puissant) en tenant compte du fait que la variance de $\hat{\rho}_{\hat{\eta}}(h)$ vaut $(T-h)/(T(T+2))$. On obtient alors la statistique de Ljung-Box :

$$Q_{LB} = T(T+2) \sum_{h=1}^H (T-h)^{-1} \hat{\rho}_{\hat{\eta}}(h)^2,$$

qui, sous l'hypothèse nulle, converge également vers une loi du χ^2 à $H-p-q$ degrés de liberté, quand T tend vers l'infini. Si on rejette l'hypothèse nulle, alors il faut essayer de trouver un autre modèle qui pourrait mieux convenir.

Chapitre 7

Prévision

7.1 Prévision des modèles ARMA (p, q)

Supposons que l'on dispose d'un grand nombre d'observations consécutives de X jusqu'à l'instant T . Une fois le modèle choisi et ses paramètres estimés, il va être possible de faire de la prévision. Pour estimer les prévisions optimales $\hat{X}_T(h)$, il faut disposer

- des coefficients ψ_j qui lient X à son bruit blanc d'innovation $\eta : X_t = \sum_{j \geq 0} \psi_j \eta_{t-j}$,
- des coefficients π_j qui lient η à $X : \eta_t = \sum_{j \geq 0} \pi_j X_{t-j}$,
- des valeurs numériques observées des $X_u, u \leq T$,
- des valeurs numériques observées des $\eta_u, u \leq T$.

7.1.1 Formule déduite de la forme autorégressive

Le théorème suivant donne les prévisions optimales obtenues grâce à la représentation $AR(\infty)$ du processus lorsque celle-ci existe :

Théorème 7.1. (i) Les prévisions optimales s'expriment comme des combinaisons linéaires

$$\widehat{X}_T(h) = \sum_{j \geq 0} \alpha_j(h) X_{T-j},$$

avec pour chaque h

$$\sum_{j \geq 0} \alpha_j(h) z^j = \left[\frac{1}{z^h} \frac{\Theta(z)}{\Phi(z)} \right]_1 \frac{\Phi(z)}{\Theta(z)}$$

où $[\]_1$ signifie que l'on ne retient que les termes en z^k pour $k \geq 0$.

(ii) Les $\widehat{X}_T(h)$ s'obtiennent par récurrence selon

$$\widehat{X}_T(h) = - \sum_{j \geq 1} \pi_j \widehat{X}_T(h-j),$$

avec les conditions initiales pour $t \leq T$: $\widehat{X}_T(t-h) = X_t$.

Remarquons que pour $h = 1$, on a $\alpha_j = -\pi_{j+1}$ et il est inutile de calculer les $\alpha_j(h)$ pour $h \geq 2$.

Autrement dit on procède comme pour un $AR(p)$ à partir du moment où l'on dispose de la représentation " $AR(\infty)$ " qui exprime le bruit blanc d'innovation à partir de X .

La deuxième équation du théorème permet de calculer récursivement les prévisions $\widehat{X}_T(h)$ à partir des prévisions précédentes. Il faut d'abord remarquer que, pour $t \leq T$, $\widehat{X}_T(t-h) = X_t$. Ainsi, pour la prévision à l'horizon 1, on a la formule suivante :

$$\widehat{X}_T(1) = - \sum_{j=1}^{\infty} \pi_j X_{T+1-j},$$

qui ne fait intervenir que des valeurs observées de la série temporelle. Pour la prévision à l'horizon 2, on a une formule basée sur les observations et sur la prévision donnée ci-

dessus :

$$\widehat{X}_T(2) = -\pi_1 \widehat{X}_T(1) - \sum_{j=2}^{\infty} \pi_j X_{T+2-j},$$

Et ainsi de suite pour tout $h \in \mathbb{N}^*$. Cependant, ces prévisions font intervenir des valeurs non observées, à savoir X_t pour $t \leq 0$. Il faut alors effectuer une approximation en tronquant la série. On obtient alors la prévision suivante :

$$\widehat{X}_T^*(h) = - \sum_{j=1}^{T+h-1} \pi_j X_{T+h-j},$$

avec toujours $\widehat{X}_T^*(t - T) = X_t$ pour $t \leq T$.

7.1.2 Formule déduite de la forme moyenne mobile

Le théorème suivant donne les prévisions optimales obtenues grâce à la représentation $MA(\infty)$ du processus lorsque celle-ci existe :

Théorème 7.2. (i) Les $\widehat{X}_T(h)$ s'expriment comme des combinaisons linéaires des valeurs passées du bruit blanc d'innovation

$$\widehat{X}_T(h) = \sum_{j \geq h} \psi_j \eta_{T+h-j}.$$

(ii) Enfin, la variance de l'erreur de prédiction est donnée par

$$\mathbb{E}((X_{T+h} - \widehat{X}_{T+h})^2) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2.$$

Les erreurs de prédiction $X_{T+h} - \widehat{X}_T(h)$ ne sont donc pas non-corrélées. En effet, comme on a :

$$X_{T+h} - \widehat{X}_T(h) = \sum_{j=0}^{h-1} \psi_j \eta_{T+h-j}.$$

alors la covariance entre l'erreur de prévision à l'horizon k et l'erreur de prévision à l'horizon h avec $h \geq k$ vaut :

$$\mathbb{E}((X_{T+h} - \widehat{X}_T(h))(X_{T+k} - \widehat{X}_T(k))) = \sigma^2 \sum_{j=0}^{k-1} \psi_j \psi_{j+h-k}.$$

7.1.3 Remarques et propriétés

On peut montrer deux propriétés intéressantes :

Proposition 7.1. 1. L'erreur de prévision à horizon 1 pour un processus ARMA est le bruit d'innovation η_{T+1} .

2. La variance de l'erreur de prévision à horizon h pour un processus ARMA croît depuis la variance du bruit d'innovation (valeur prise pour $h = 1$) jusqu'à la variance du processus lui-même.

Intervalle de confiance pour la prévision

De la même façon que pour les MA, si on suppose que le bruit d'innovation est gaussien, les variables aléatoires X_t sont elles aussi gaussiennes, tout comme l'erreur de prédiction. Ainsi, dans le cas d'un bruit blanc fort (donc gaussien), on peut obtenir des intervalles de confiance pour la prédiction à un horizon donné :

$$\mathbb{P}(X_{T+h} \in [\widehat{X}_T(h) - z_{\alpha/2}\sigma(h); \widehat{X}_T(h) + z_{\alpha/2}\sigma(h)]) = 1 - \alpha,$$

où $z_{\alpha/2}$ est le quantile d'ordre $\alpha/2$ de la loi normale centrée réduite et où $\sigma(h)^2$ est l'erreur quadratique moyenne au pas h .

Mise à jour

Le problème est le suivant : on a, en se basant sur les observations antérieures à T , prévu $X_{T+1}, X_{T+2}, \dots, X_{T+h}$. On observe maintenant X_{T+1} . Non seulement on n'a plus à le prévoir, mais encore on connaît maintenant la valeur de l'erreur de prévision $X_{T+1} - \widehat{X}_T(1)$. De plus, on va pouvoir modifier les prévisions de X_{T+2}, \dots, X_{T+h} . Pour cela, on écrit le développement en moyenne mobile

$$X_{T+h} = \eta_{T+h} + \psi_1 \eta_{T+h-1} + \dots + \psi_{h-1} \eta_{T+1} + \psi_h \eta_T + \dots$$

Le prédicteur est donné par

$$\widehat{X}_{T+1}(h-1) = \psi_{h-1} \eta_{T+1} + \dots + \dots = \psi_{h-1} \eta_{T+1} + \widehat{X}_T(h).$$

Donc la prévision de X_{T+h} s'obtient en ajoutant $\psi_{h-1} \eta_{T+1}$ à la prévision précédemment obtenue. Or on connaît la valeur de ce terme complémentaire. En effet, η_{T+1} est, on l'a vu, l'erreur de prévision $X_{T+1} - \widehat{X}_T(1)$. On a donc le résultat suivant

Proposition 7.2. *Lorsque l'observation X_{T+1} vient s'ajouter aux précédentes, la mise à jour des prévisions se fait de la façon suivante :*

$$\widehat{X}_{T+1}(h-1) = \psi_{h-1} (X_{T+1} - \widehat{X}_T(1)) + \widehat{X}_T(h).$$

Autrement dit, on ajoute au prédicteur précédent de X_{T+h} une correction proportionnelle à l'erreur que l'on avait faite en prédisant, avant de l'avoir observée, la donnée que l'on vient de recueillir.

Bibliographie

- [1] R. Azencott, D. Dacunha. Castelle : *Séries d'observations irrégulières*, Masson 1984.
- [2] P.J. Brockwell. R.A. Davis : *Introduction to Time Series and Forecasting*, 1998.
- [3] G. Box, G. Jenkins : *Time series analysis*, Holden Day. 1976.
- [4] Yves Caumel : *Probabilités et processus stochastiques* 2011
- [5] ESIEA 4A : *Introduction aux Processus Stochastiques* 2013
- [6] Dauphine-Eurisko : *Processus Aléatoires Stationnaires et Processus ARMA*.
- [7] W.A. Fuller : *Introduction to statistical time series*, JOHN WILEY and SONS 1976.
- [8] C. Gouriéroux, A. Montfort : *Cours de séries temporelles*, Economica. 1983.
- [9] C. Chatfield : *The analysis of time series*, Chapman-Hall. 1975.
- [10] Agnès Lagnoux : *Renforcement Statistique Séries chronologiques*.
- [11] Pierre Priouret : *Introduction aux processus de diffusion* 2005
- [12] Silverman, B. W : *Density Estimation. London*, Chapman and Hall. 1986.