



Renforcement Statistique Séries chronologiques

Agnès Lagnoux
lagnoux@univ-tlse2.fr

ISMAG
MASTER 1 - MI0B246X

Table des matières

1	Généralités et rappels sur les processus stochastiques	6
1.1	Variables de carré intégrable	6
1.2	Processus stochastiques	8
1.3	Opérateurs avance et retard	8
1.3.1	Inversion de $I - \lambda B$	10
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	12
1.4.3	Estimations des fonctions caractéristiques	16
2	Les processus linéaires, MA et AR	18
2.1	Processus linéaires	18
2.2	Processus moyennes mobiles	21
2.2.1	Définition d'un $MA(q)$	21
2.2.2	Forme autorégressive infinie d'un $MA(q)$	21
2.2.3	Bruit blanc d'innovation d'un $MA(q)$	22
2.2.4	Fonctions caractéristiques et propriétés d'un $MA(q)$	23
2.2.5	Prédiction d'un $MA(q)$	24
2.2.6	Erreur et intervalle de confiance pour la prévision d'un $MA(q)$	26
2.3	Processus autorégressifs	26
2.3.1	Définition d'un $AR(p)$	26
2.3.2	Ecriture moyenne mobile infinie d'un $AR(p)$	26
2.3.3	Bruit blanc d'innovation d'un $AR(p)$	27
2.3.4	Fonctions caractéristiques d'un $AR(p)$	27
2.3.5	Prédiction d'un $AR(p)$	32
2.3.6	Equations de Yule-Walker d'un $AR(p)$	33
3	Les processus mixtes ARMA	37
3.1	Définition d'un $ARMA(p, q)$	37
3.2	Expression d'un $ARMA(p, q)$	37
3.2.1	Représentation moyenne mobile infinie	37
3.2.2	Représentation autorégressive	38
3.3	Bruit blanc d'innovation d'un $ARMA(p, q)$	38
3.4	Autocovariances et propriétés d'un $ARMA(p, q)$	39
3.5	Prévision des modèles $ARMA(p, q)$	40
3.5.1	Formule déduite de la forme autorégressive	40
3.5.2	Formule déduite de la forme moyenne mobile	41
3.5.3	Remarques et propriétés	42
3.6	L'approche de Box et Jenkins	42
3.6.1	Pré-sélection de modèles	43
3.6.2	Estimation des paramètres	44
3.6.3	Sélection de modèles	45
3.6.4	Test sur les résidus	45
3.7	Résumé	46

4	Processus non stationnaires ARIMA et SARIMA	49
4.1	Processus ARIMA	49
4.1.1	Définition	49
4.1.2	Formes moyenne mobile et autorégressive	51
4.1.3	Prévision optimale des modèles ARIMA	52
4.1.4	Exemples de processus ARIMA et prédiction	54
4.2	Processus SARIMA	56
4.2.1	Motivation et définition	56
4.2.2	Prévision des processus SARIMA	58
4.3	Identification et ajustement des modèles ARIMA/SARIMA	59
5	Processus ARCH et GARCH	60
5.1	Les processus ARCH	61
5.2	Les processus GARCH	62
5.3	Les processus ARMA-GARCH	64
5.4	Estimation des paramètres	65
5.5	Prédiction	66

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 pré-traitement 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.

Enfin, les séries précédemment étudiées étaient supposées à transformation près stationnaires. Si besoin, tendances et saisonnalités étaient supprimées pour obtenir une série résiduelle stationnaire ou du moins une série à accroissements stationnaires. Néanmoins, toutes les séries résiduelles obtenues de la sorte ne sont pas nécessairement stationnaires : il peut arriver que la variance d'un processus varie au cours du temps. C'est le cas des séries ARCH ou plus généralement GARCH. Nous verrons comment procéder dans le cadre de telles séries.

Dans chacun des cas, une fois le modèle choisi, on estime les paramètres inconnus à partir des observations. Des tests permettent ensuite de vérifier que le modèle identifié est bien adapté aux observations. Enfin, le modèle identifié peut servir à résoudre des problèmes de contrôle, de détection, d'interpolation ou de prédiction des valeurs futures de (ϵ_t).

A partir de maintenant, la série résiduelle sera notée X_t au lieu de ϵ_t .

Dans un premier temps, nous donnerons quelques généralités sur les processus stochastiques. Nous considérerons ensuite les processus stationnaires du second ordre. Après avoir défini la stationnarité et la notion d'innovation, on donnera d'une part leur analyse et d'autre part des éléments de description tels que l'estimation de la moyenne, de l'autocovariance... La seconde partie de ce cours (partie principale) traitera des séries temporelles. Tout d'abord, nous présenterons l'étude des séries stationnaires et leur modélisation (processus linéaires, AR et MA dans le second chapitre et processus ARMA dans le troisième). Puis nous traiterons les processus non stationnaires ARIMA et SARIMA et enfin, dans le

dernier chapitre nous ferons une rapide présentation des processus ARH et GARCH particulièrement utilisés en finance.

1 Généralités et rappels sur les processus stochastiques

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 $\mathbb{E}(X_t^2) < \infty$.

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

$$\begin{aligned} L^2 \times L^2 &\longrightarrow \mathbb{R} \\ (X, Y) &\longmapsto \mathbb{E}(XY) \end{aligned}$$

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

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

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

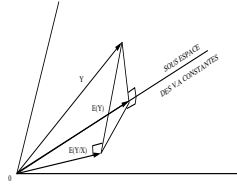
L'espérance conditionnelle ou régression

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 $(\Omega, \mathcal{B}_X, \mathbb{P})$, \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

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

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_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 Θ 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 (Ω', \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 considérerons 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 $B^2 = BoB$ 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 : $AoB(X) = BoA(X)$.

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

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

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 Φ est inversible. On appellera Ψ 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) $\mathbb{E}(X_t) = \mu, \forall t \in \mathbb{Z}$;
- (ii) X_t est de carré intégrable pour tout $t \in \mathbb{Z} : \mathbb{E}(X_t^2) < \infty$;
- (iii) $\text{Cov}(X_s, X_{s+t}) = \text{Cov}(X_{s-1}, X_{s-1+t}) = \dots = \text{Cov}(X_0, X_t), \forall t, s \in \mathbb{Z}$.

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 $\forall k \in \mathbb{N}, t_1, \dots, t_k \in \mathbb{Z}$ 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.*

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

Notons que si X est un processus strictement stationnaire, la loi de X_t ne dépend pas de t : les var 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 fautive. 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. $|\gamma(h)| \leq \gamma(0), \forall h$.

Preuve *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. $\forall h \in \mathbb{Z}, \quad \rho(-h) = \rho(h)$: 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 $\rho(h) = \phi^h$ 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 $X_t - X_{t,h}^*$ 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,

$$\begin{aligned} X_t &= \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_{h-1} X_{t-h+1} + U \\ X_{t-h} &= \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_{h-1} X_{t-h+1} + V \end{aligned}$$

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 $X_{t-h+1}, \dots, X_{t-1}$.

Preuve En effet, on a

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

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) 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 α : $\alpha = \text{Cor}(X, Y)$. 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. $\text{Var}(\bar{X}) = \frac{1}{T} \sum_{|h| < T} \left(1 - \frac{|h|}{T}\right) \gamma(h)$.

Preuve 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

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

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

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

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.

2 Les processus linéaires, MA et AR

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 :

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

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

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

Preuve 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.

Preuve *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 $H_{-\infty}(X) = \bigcap_t H_{\leq t}(X)$ 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 défini 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 supposons de plus centrés : les processus MA, les processus AR. Il suffira dans le cas où X est non centré de moyenne μ ($\mathbb{E}(X_t) = \mu \forall t$) 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, ARIMA et SARIMA.

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(∞) 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 ($\mu \mapsto \frac{1}{\mu}$), les déplaçant 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 déduit 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 $|\mu_i| \leq 1 \forall i$ puisque nous venons de voir qu'il est toujours possible de se ramener à ce cas quitte à changer de bruit blanc.

2.2.4 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

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

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

$$\text{Var}(X_t) = \gamma^X(0) = (1 + \theta_1^2 + \dots + \theta_q^2)\gamma^n(0) > \gamma^n(0) = \sigma^2, \quad \forall t$$

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

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

Preuve 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) :

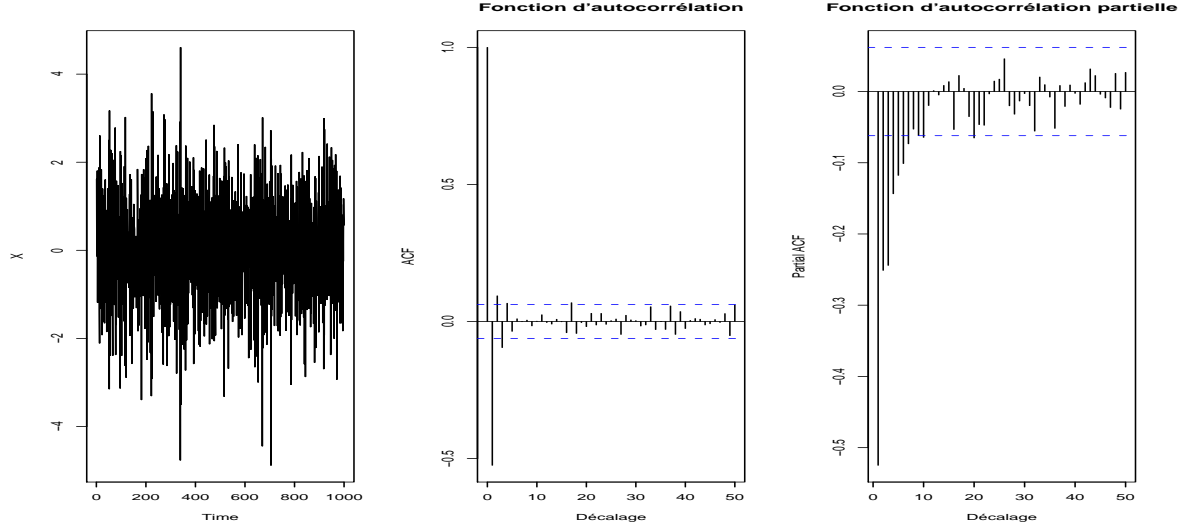


FIGURE 1 – Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus MA(1) : $X_t = \epsilon_t - 0.8\epsilon_{t-1}$

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

$$\begin{aligned} \hat{X}_t(1) &= -\theta_1\eta_t - \dots - \theta_q\eta_{t-(q-1)} \\ \hat{X}_t(2) &= -\theta_2\eta_t - \dots - \theta_q\eta_{t-(q-2)} \\ &\dots \\ \hat{X}_t(q) &= -\theta_q\eta_t \\ \hat{X}_t(j) &= 0, \quad j > q \end{aligned}$$

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 $j > q \Rightarrow \hat{X}_t(j) = 0$ et d'autre part, les prévisions nécessitent la combinaison de *toutes* les valeurs du passé de X .

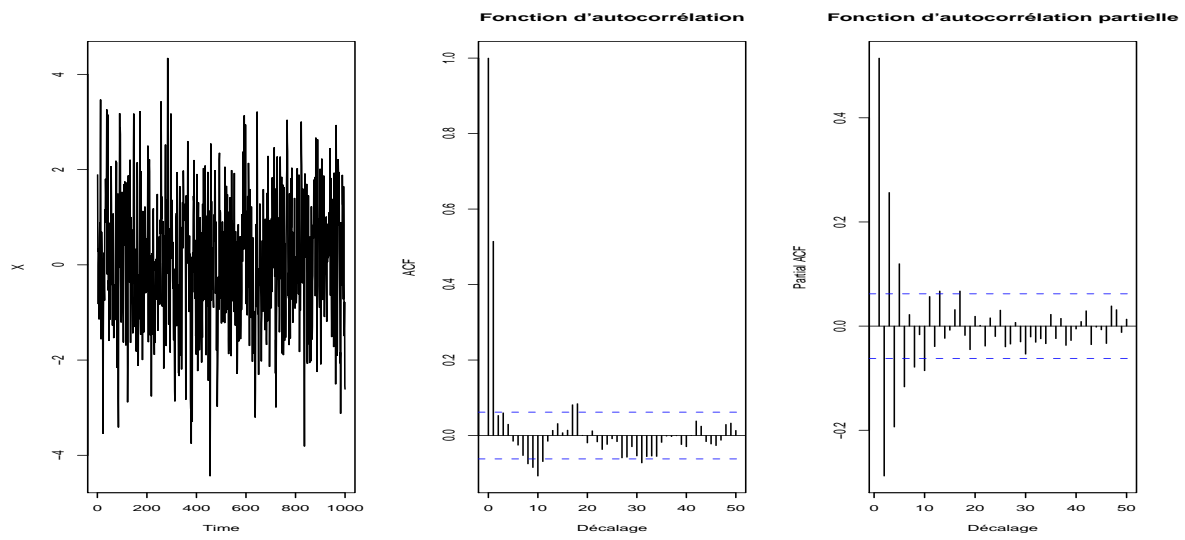


FIGURE 2 – Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus MA(1) : $X_t = \epsilon_t + 0.8\epsilon_{t-1}$

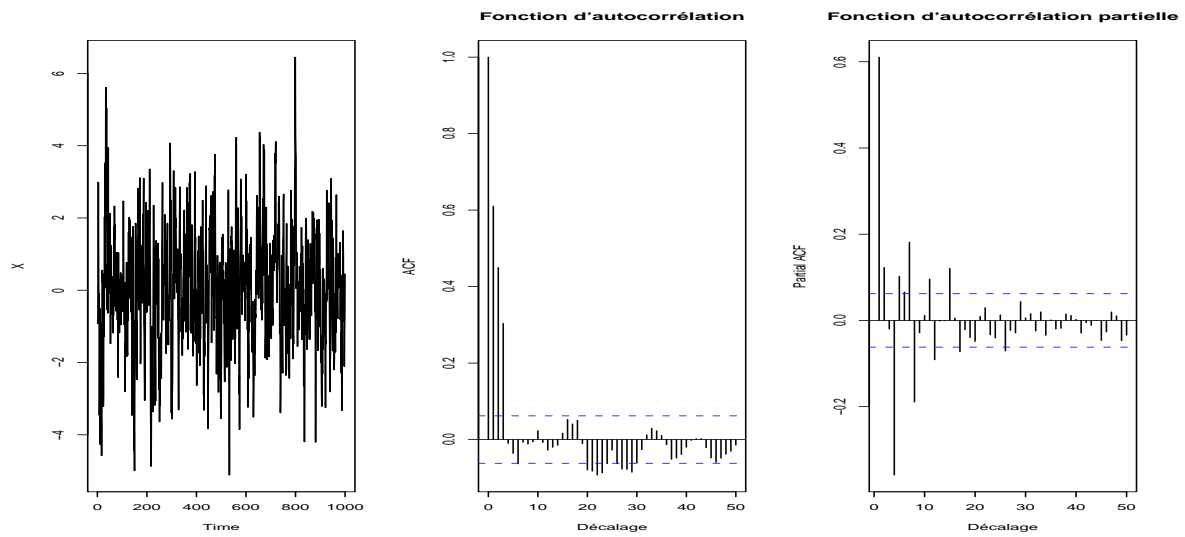


FIGURE 3 – Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus MA(3) : $X_t = \epsilon_t - 0.8\epsilon_{t-1} - 0.4\epsilon_{t-2} - 0.9\epsilon_{t-3}$

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

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

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 $\sum_{i=1}^p \phi_i X_{t-i}$ (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

$$\forall t \in \mathbb{Z}, \quad X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \eta_t \tag{2}$$

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 :

$$\Phi(B)X_t = \eta_t, \quad t \in \mathbb{Z}. \tag{3}$$

Remarque 2.2 1. En prenant l'espérance de (2) 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 Φ , 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 (3) 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 Φ 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 η_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 Si 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 considèrerons plus que des processus $AR(p)$ donnés sous leur representation canonique i.e. vérifiant l'équation

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

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

$$X_t - \phi X_{t-1} = \eta_t \quad \text{i.e.} \quad (I - \phi B)X_t = \eta_t \quad (4)$$

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

$$X_t = \sum_{j=0}^{+\infty} \phi^j \eta_{t-j}. \quad (5)$$

La fonction de transfert a donc une infinité de termes. On vérifie que X défini par la relation (5) 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 (5) satisfait l'équation (4) et est par conséquent l'unique solution stationnaire de (4).

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 inférieure 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 (5) 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 considèrera plus de $AR(1)$ avec $|\phi| > 1$.

3) Dans le cas où $|\phi| = 1$, il n'y a pas de solution stationnaire à (4) et par conséquent les $AR(1)$ n'existent pas selon la définition des processus AR .

On peut généraliser cette notion de causalité aux processus $AR(p)$ et $ARMA$. Dorénavant, nous ne considèrerons plus que des processus $AR(p)$ donnés sous leur représentation canonique

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

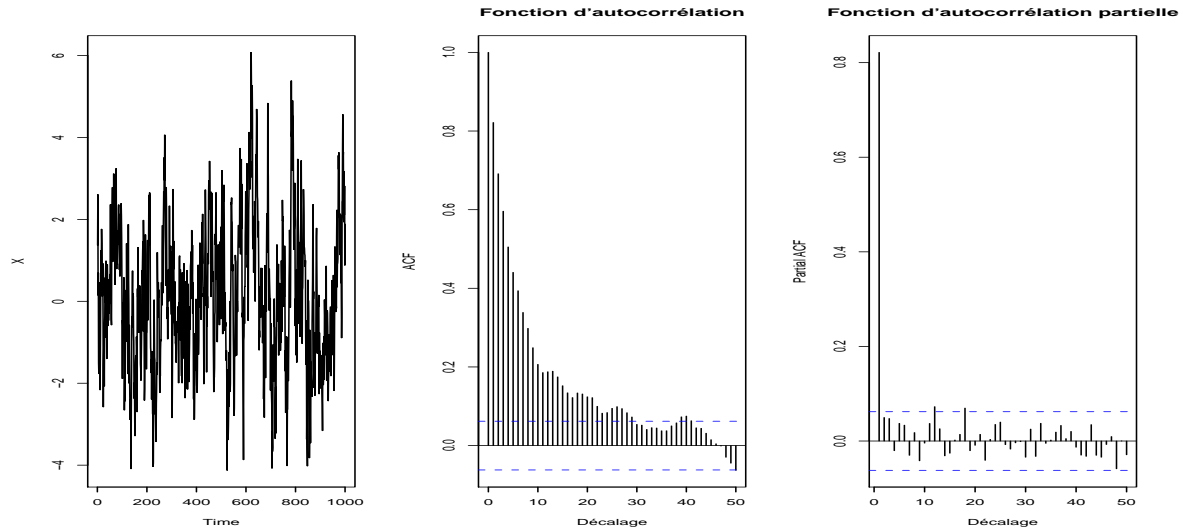


FIGURE 4 – Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus AR(1) : $X_t - 0.8X_{t-1} = \epsilon_t$

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

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 $|h| = p$.

Preuve 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

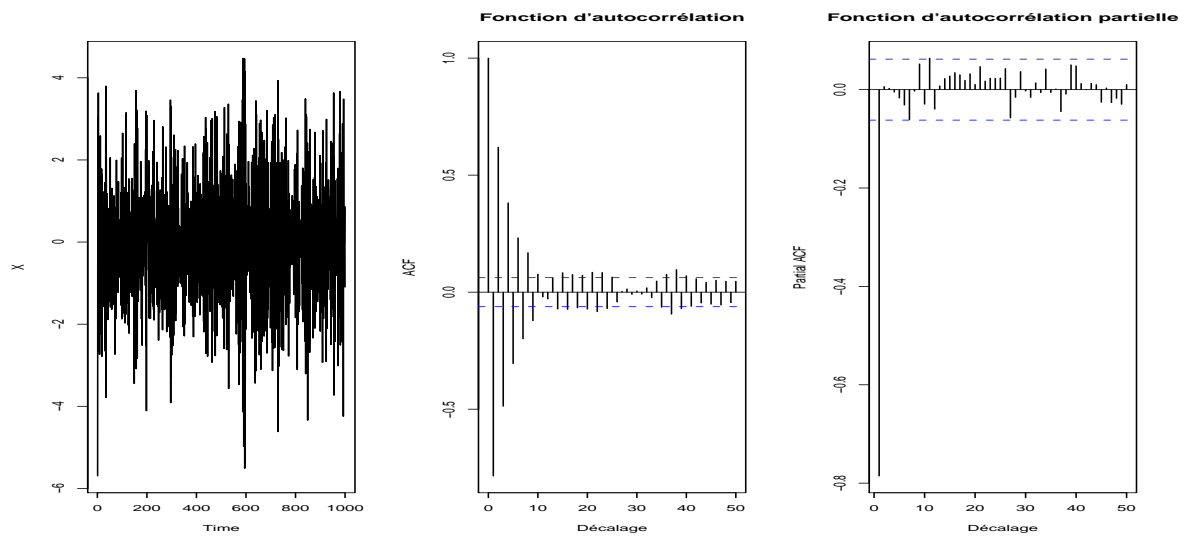


FIGURE 5 – Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus AR(1) : $X_t + 0.8X_{t-1} = \epsilon_t$

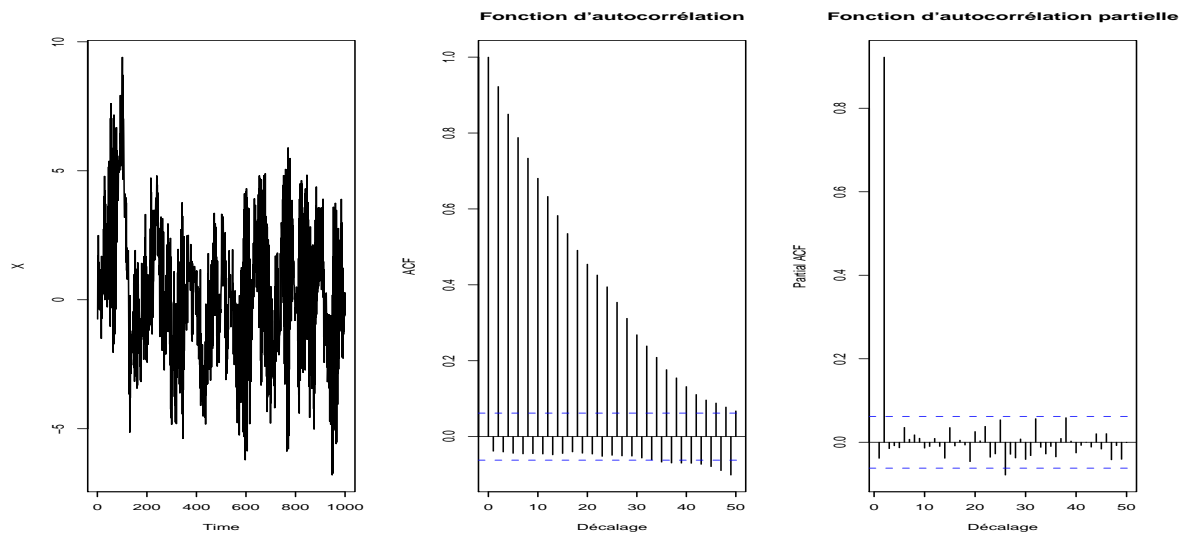


FIGURE 6 – Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus AR(2) : $X_t - 0.9X_{t-2} = \epsilon_t$

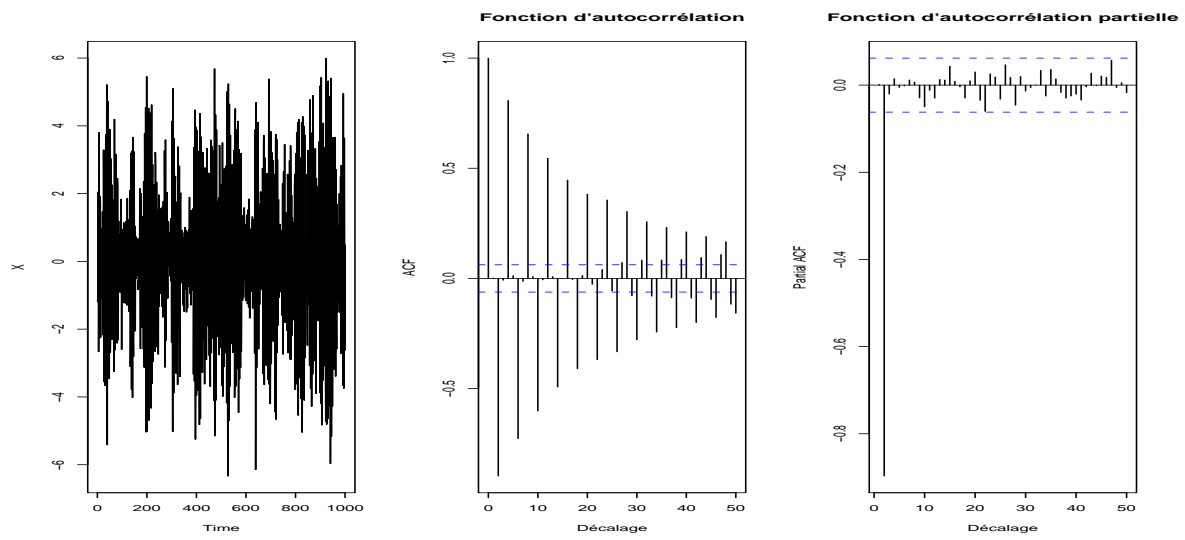


FIGURE 7 – Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus AR(2) : $X_t + 0.9X_{t-2} = \epsilon_t$

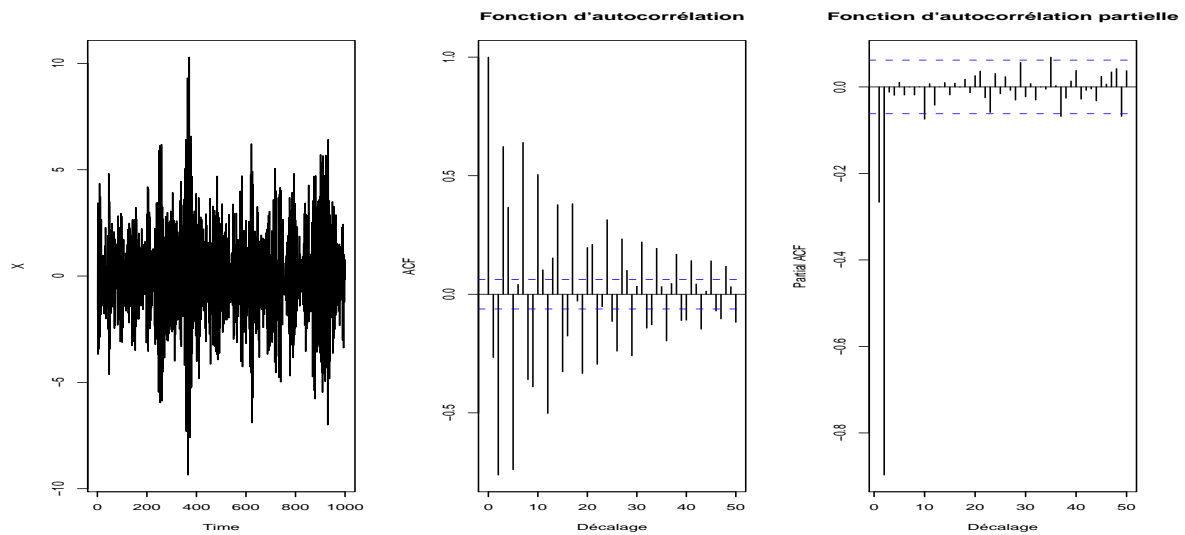


FIGURE 8 – Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus AR(2) : $X_t + 0.5X_{t-1} + 0.9X_{t-2} = \epsilon_t$

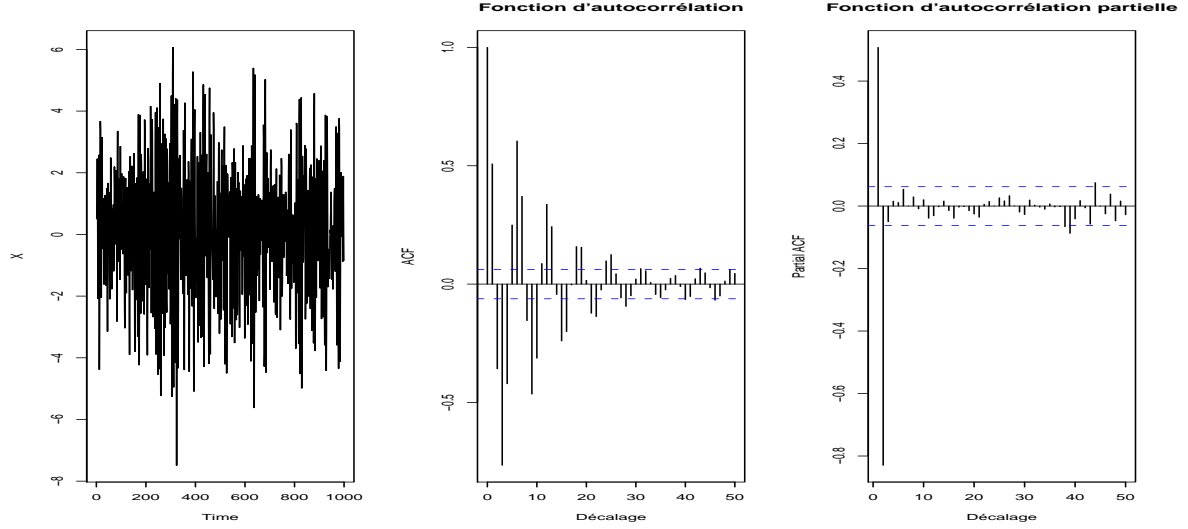


FIGURE 9 – Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus AR(2) : $X_t - 0.9X_{t-1} + 0.8X_{t-2} = \epsilon_t$

dates après et une forte valeur six dates après (le caractère faible ou fort dépend bien évidemment du signe de ρ). La présence du bruit de variance σ^2 explique évidemment que la fonction ne soit pas strictement périodique.

2.3.5 Prédiction d'un AR(p)

Les AR(p) sont bien adaptés à la prédiction. Les prévisions optimales $\hat{X}_t(h)$ 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

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

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

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

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 $\hat{X}_t(h)$ pour $h > 0$ sont des combinaisons linéaires de $X_t, X_{t-1}, \dots, X_{t-p+1}$:

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

qui s'obtiennent par récurrence selon

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

avec les conditions initiales $\hat{X}_t(-j) = X_{t-j}$ 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,*

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

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

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

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 :

$$\forall t \in \mathbb{Z}, \quad X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = \eta_t. \quad (6)$$

On le suppose de plus centré et causal comme usuellement.

En multipliant l'équation (6) 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 (6), 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 :

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

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 :

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

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

$$\hat{\sigma}^2 \xrightarrow{\mathbb{P}, T \rightarrow \infty} \sigma^2.$$

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

Théorème 2.12 *Etant donnés ϕ_1, \dots, ϕ_p et $\gamma^n(0)$, 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 $\gamma^n(0)$.*

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 $\gamma^n(0)$. 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 $\gamma^n(0)$ est petit par rapport à $\gamma^X(0)$. Cela signifie que les racines (ou co-racines) de Φ doivent être proches du cercle unité.

3 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'écart; ce dernier détermine 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.

3.1 Définition d'un ARMA(p, q)

Définition 3.1 On appelle **processus autorégressif moyenne mobile** d'ordre (p, q) tout processus $(X_t)_{t \in \mathbb{Z}}$ stationnaire tel que

$$\forall t \in \mathbb{Z}, \quad X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \eta_t - \theta_1 \eta_{t-1} - \dots - \theta_q \eta_{t-q} \quad (7)$$

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 3.1 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.

3.2 Expression d'un ARMA(p, q)

3.2.1 Représentation moyenne mobile infinie

Propriété 3.1 (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

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

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 3.1 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

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

3.2.2 Représentation autorégressive

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

Propriété 3.2 (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

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

(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 présentes 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 3.2 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 .

3.3 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 3.1 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 $\forall k, \quad |\lambda_k| < 1$ et $\forall j, \quad |\mu_j| \leq 1$.

Théorème 3.2 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

- $\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p = \prod_{i=1}^p (1 - \lambda_i z)$ où $|\lambda_i| \leq 1 \forall i$
 - $\Theta(z) = 1 - \theta_1 z - \dots - \theta_q z^q = \prod_{i=1}^q (1 - \mu_i z)$ 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.**

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

Propriété 3.3 La fonction d'autocovariance d'un processus ARMA(p, q) vérifie pour $h \geq \max(p, q + 1)$

$$\gamma(h) = \sum_{i=1}^p \phi_i \gamma(h - i) \quad (8)$$

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

$$\gamma(h) = \sum_{i=1}^p \phi_i \gamma(h - i) + \sigma^2 (\psi_0 - \psi_1 \theta_{h+1} - \dots - \psi_{q-h} \theta_q) \quad (9)$$

Exercice Démonstration du résultat précédent

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

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

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 3.2 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 3.1 La somme d'un ARMA(p_1, q_1) et d'un ARMA(p_2, q_2) (non corrélés) est un ARMA($\leq \max(p_1, p_2), \leq \max(q_1, q_2)$).

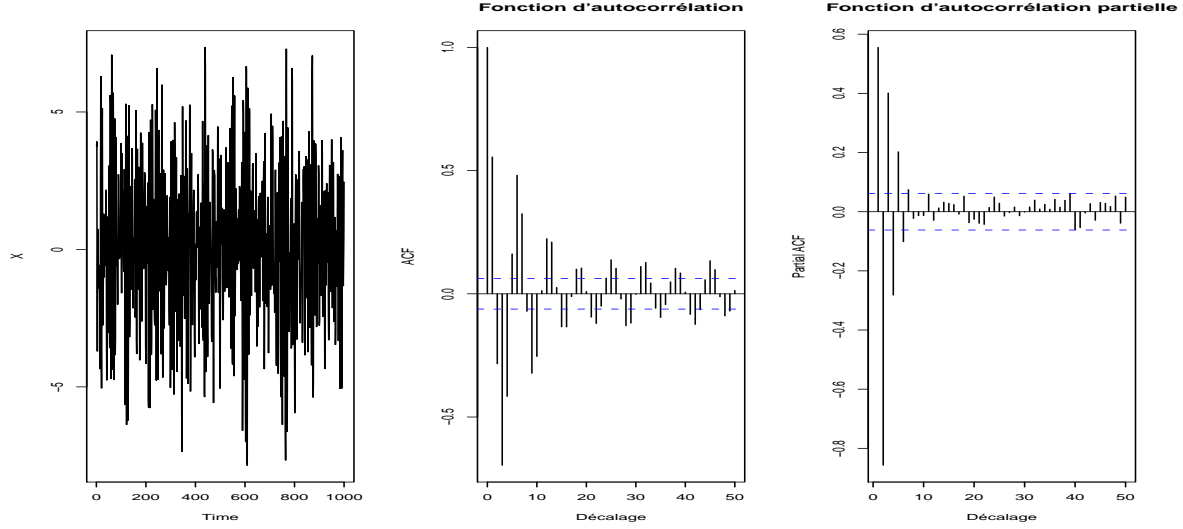


FIGURE 10 – Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus ARMA(2,1) : $X_t - 0.9X_{t-1} + 0.8X_{t-2} = \eta_t - 0.7\eta_{t-1}$

Par exemple,

- la somme de deux bruits blancs indépendants (=ARMA(0,0)) est un bruit blanc.
- la somme d'un bruit blanc et d'un MA(q) est un MA($\leq q$) (en fait un MA(q)).
- la somme d'un bruit blanc et d'un AR(p) indépendant est un ARMA($p, \leq p$).
- la somme d'un AR(p_1) et d'un AR(p_2) est un ARMA($p_1 + p_2, \leq p_1 + p_2$).
- la somme d'un AR(p) et d'un MA(q) est un ARMA($p, \leq p + q$).
- la somme d'un MA(q_1) et d'un MA(q_2) est un MA($\leq q_1 + q_2$).

Dans la figure Fig. 10, nous avons tracé un graphe de trajectoire d'ARMA ainsi que les corrélogramme et corrélogramme partiel empiriques associés.

3.5 Prédiction 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 prédiction. 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 η : $X_t = \sum_{j \geq 0} \psi_j \eta_{t-j}$,
- 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 $\eta_u, u \leq T$.

3.5.1 Formule déduite de la forme autorégressive

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

Théorème 3.3 (i) *Les prévisions optimales s'expriment comme des combinaisons linéaires*

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

avec pour chaque h

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

où []₊ 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 $t \leq T$: $\widehat{X}_T(t-T) = X_t$.

Remarquons que pour $h = 1$, on a $\alpha_j(1) = -\pi_{j+1}$ 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(∞)" 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 $t \leq T$, $\widehat{X}_T(t-T) = X_t$. 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 $t \leq T$.

3.5.2 Formule déduite de la forme moyenne mobile

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

Théorème 3.4 (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

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

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 :

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

3.5.3 Remarques et propriétés

On peut montrer deux propriétés intéressantes :

Proposition 3.2 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 :

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

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 $X_{T+1} - \hat{X}_T(1)$. 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

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

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 $X_{T+1} - \hat{X}_T(1)$. On a donc le résultat suivant

Proposition 3.3 Lorsque l'observation X_{T+1} vient s'ajouter aux précédentes, la mise à jour des prévisions se fait de la façon suivante :

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

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.

3.6 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

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

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 $\widehat{\rho}_T$ et d'autocorrélation partielle τ par $\widehat{\tau}_T$.

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 $\widehat{\rho}_T(h)$ et $\widehat{\tau}_T(h)$ 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.

3.6.1 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,

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

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

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

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 $\widehat{\rho}_T$ 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,

$$\sqrt{T} \widehat{\tau}_T(h) \xrightarrow[T \rightarrow \infty]{d} \mathcal{N}(0, 1).$$

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

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

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 $\widehat{\tau}_T$ 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.

3.6.2 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 (7) 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 :

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

Remarque 3.3 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 $\widehat{\rho}$. 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 :

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

Elle ne dépend donc que de paramètres connus ou pouvant être estimés. Par ailleurs, en partant du membre de droite de l'équation (7), 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 :

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

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.

3.6.3 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).

3.6.4 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 :

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

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.

3.7 Résumé

Pour conclure, nous présentons en Figure 11 un petit résumé de la méthodologie de Box et Jenkins sous la forme d'un graphique :

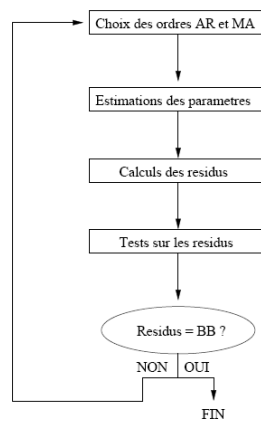


FIGURE 11 – Méthodologie de Box et Jenkins

ainsi qu'un tableau récapitulatif des propriétés d'autocovariances et d'autocorrélations des processus ARMA, AR et MA utiles pour le choix de modèles :

modèle	MA(q)	AR(p)	ARMA(p, q)
autocov.	$\forall h > q \gamma(h) = 0$	$\lim_{h \rightarrow \infty} \gamma(h) = 0$	$\lim_{h \rightarrow \infty} \gamma(h) = 0$
autocorr.	$\forall h > q \rho(h) = 0$	$\lim_{h \rightarrow \infty} \rho(h) = 0$	$\lim_{h \rightarrow \infty} \rho(h) = 0$
autocorr. partielle	$\lim_{h \rightarrow \infty} \tau(h) = 0$	$\forall h > p \tau(h) = 0$ et $\tau(p) = \phi_p$	

4 Processus non stationnaires ARIMA et SARIMA

Dans ce chapitre nous allons finalement résoudre le problème de la construction en toute généralité des prédicteurs pour des processus soit stationnaires, soit non stationnaires (à cause de la présence éventuelle d'une tendance ou d'une saisonnalité). Nous allons voir surtout comment l'approche paramétrique de Box-Jenkins, c'est-à-dire la modélisation du type ARMA à un nombre fini de paramètres, est capable de fournir de bons prédicteurs (linéaires) si le nombre d'observations d'une période assez homogène de la fin de notre série est suffisamment grand. Nous pourrions comparer cette approche avec une approche plutôt non paramétrique, le lissage exponentiel, qui est utile pour le cas d'une série de courte durée (ou le cas d'une courte période d'observation à la fin de la série à partir d'une rupture importante de la structure de la série observée).

Il est évident que pour la plupart des séries économiques, l'hypothèse de stationnarité n'est pas valide. En revanche, si l'on considère par exemple les différences premières (ou plus généralement les différences d'ordre d) de telles séries, l'hypothèse de stationnarité devient souvent vraisemblable. Il est donc naturel de considérer la classe des processus dont la différence d'un certain ordre satisfait une représentation ARMA.

Exemple 4.1 *Considérons une série temporelle $(X_t)_{1 \leq t \leq T}$ qui présente une saisonnalité de période 12. Afin de supprimer cette saisonnalité, c'est la série suivante que nous étudions :*

$$Y_t = X_t - X_{t-12} \quad \text{pour } t = 13, \dots, T.$$

L'ajustement d'un modèle ARMA et les prévisions sont réalisées sur la série $(Y_t)_t$. Il est ensuite nécessaire de revenir à la série initiale, car ce sont les prévisions de $(X_t)_t$ qui nous intéressent. De l'équation précédente, on obtient :

$$\begin{aligned} X_t &= Y_t + X_{t-12} \\ &= Y_t + Y_{t-12} + X_{t-24} \\ &\vdots \\ &= Y_t + Y_{t-12} + Y_{t-24} + \dots + Y_{r(t)+12} + X_{r(t)} \end{aligned}$$

où $r(t)$ est le reste de la division euclidienne de t par 12. Puisque l'on connaît les X_t pour $t \leq T$ ainsi que les prévisions $\hat{Y}_T(h)$ de Y_{T+h} , on peut en déduire les prévisions de X_{T+h} . Ecrire pour cet exemple la prévision à l'horizon 1 de la série $(X_t)_{1 \leq t \leq T}$.

Ce chapitre a pour but de combiner les méthodes d'ajustement des modèles ARMA et leur prévision vues aux chapitres précédents pour le traitement des composantes non stationnaires, notamment tendances (ce qui mène aux processus ARIMA) et saisonnalités (processus SARIMA). Nous allons voir comment la modélisation ARMA se généralise et comment faire de la prévision pour des processus non stationnaires (après avoir bien identifié le modèle ajusté avec les méthodes de diagnostic).

4.1 Processus ARIMA

4.1.1 Définition

Si on note $\Delta^d X_t$ la différence d'ordre d de X_t c'est-à-dire $(I - B)^d X_t$, on va s'intéresser aux processus (X_t) satisfaisant

$$\phi(B)\Delta^d X_t = \Theta(B)\eta_t, \quad (10)$$

où les racines de ϕ et Θ sont de module supérieur à 1 et où (η_t) est un bruit blanc de variance σ^2 . La relation (10) s'écrit aussi

$$\Phi(B)X_t = \Theta(B)\eta_t, \quad (11)$$

avec $\Phi(B) = \phi(B)(I - B)^d$. La relation (11) est analogue à la définition d'un ARMA($p + d, q$) à la différence importante près que le polynôme Φ admet 1 comme racine d'ordre d .

La présentation qui vient d'être faite n'est pas suffisante. En effet, le polynôme $\Phi(B)$ n'étant pas inversible, la relation (11) ne détermine pas le processus de manière unique c'est-à-dire que contrairement au cas stationnaire, on ne peut pas en supposant que la relation (11) est vraie pour tout t en déduire X_t par inversion de $\Phi(B)$ puisque la série en η_t obtenue est divergente. Pour compléter la définition, il faut introduire une initialisation. Nous allons considérer ce mécanisme de démarrage sur deux exemples simples.

Exemple 4.2 1. Soit une *marche aléatoire* c'est-à-dire un processus X satisfaisant

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

à partir d'une certaine date par exemple $t = 0$. Nous avons déjà vu au Chapitre 1 que ce processus n'est pas stationnaire. Si une valeur initiale X_{-1} (déterministe ou aléatoire) est donnée alors

$$X_t = X_{-1} + \sum_{j=0}^t \eta_j.$$

Il est naturel d'autre part de supposer que X_{-1} est non corrélée avec les valeurs futures du bruit. Le processus (X_t) est alors défini sans ambiguïté i.e. de manière unique.

2. De même si on considère un processus X satisfaisant

$$X_t - X_{t-1} = \eta_t - \theta\eta_{t-1}, \quad \forall t \geq 0,$$

ce processus est complètement défini par la donnée de valeurs initiales X_{-1} et η_{-1} .

Plus généralement, on adopte la définition suivante

Définition 4.1 Soient $p, q, d \geq 0$. Un processus (X_t) est un processus ARIMA(p, d, q) [Autoregressif moyenne mobile intégré] s'il satisfait une équation du type

$$\phi(B)(I - B)^d X_t = \Theta(B)\eta_t, \quad (12)$$

où

$$\begin{cases} \phi(B) = I + \phi_1 B + \dots + \phi_p B^p, & \phi_p \neq 0 \\ \Theta(B) = I + \theta_1 B + \dots + \theta_q B^q, & \theta_q \neq 0 \end{cases}$$

sont des polynômes dont les racines sont de module supérieur à 1, où les conditions initiales

$$(X_{-1}, X_{-2}, \dots, X_{-p-d}, \eta_{-1}, \eta_{-2}, \dots, \eta_{-q})$$

sont non corrélées avec $\eta_0, \eta_1, \dots, \eta_t$, $\eta = (\eta_t)_{t \geq -q}$ étant un bruit blanc de variance σ^2 .

Il est clair que, pour des conditions initiales et un processus donnés, le processus est parfaitement défini puisque X_0 est donné par l'équation (12) au temps $t = 0$, X_1 est donné par l'équation (12) au temps $t = 1$ et ainsi de suite.

Remarque 4.1 On inclut ainsi les composantes non stationnaires de tendance (trend) dans la modélisation des séries chronologiques d'où le nom de processus ARIMA : processus autoregressif moyenne mobile intégré.

Les processus ARIMA sont utiles pour modéliser des processus qui ont des corrélations positives et lentement décroissantes car cette propriété des autocorrélations peut être le signe d'une tendance dans la série. En pratique, les modèles ARIMA sont préférés aux modèles AR dont les racines sont proches du cercle de l'unité. Par exemple, le modèle AR(2)

$$(I - 0.815B)(I - 0.989B)X_t = \eta_t$$

possède des racines proches du cercle unité, mais un modèle ARIMA

$$(I - aB)(I - B)X_t = \eta_t$$

est plus stable pour l'analyse et la prévision.

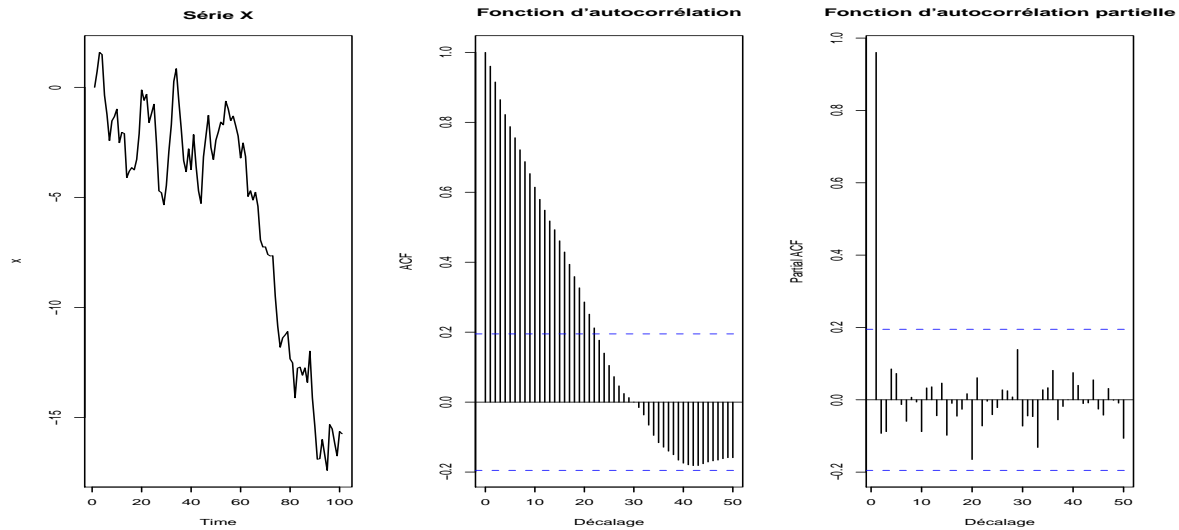


FIGURE 12 – Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus ARIMA(1, 1, 1) de coefficient AR=0.1 et MA=0.3

Exemple 4.3 – Comme premiers exemples de processus ARIMA, nous venons de voir la marche aléatoire qui est un processus ARIMA(0,1,0) et le processus défini par

$$X_t - X_{t-1} = \eta_t - \theta\eta_{t-1}, \quad \forall t \geq 0,$$

qui est un ARIMA(0,1,1).

– Le processus $X_t = a + bt + \eta_t$ est un exemple de processus ARIMA(0,1,1), car

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

est un MA(1) (non inversible) de moyenne b . Pour supprimer la moyenne, on étudie plutôt $Y_t = X_t - bt$ qui vérifie la relation

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

– Le processus AR(1) défini par $X_t - \varphi X_{t-1} = \eta_t$ est un exemple de processus ARIMA(1,0,0).

Proposition 4.1 Soit X un processus ARIMA(p, d, q). Le processus $(\Delta^d X_t)$ tend vers un processus ARMA stationnaire lorsque t tend vers l'infini.

$\Delta^d X_t$ est dit **asymptotiquement stationnaire**.

Preuve A faire en exercice.

4.1.2 Formes moyenne mobile et autorégressive

Notons $Z_{-1} = (X_{-1}, \dots, X_{-p-d}, \eta_{-1}, \dots, \eta_{-q})'$ le vecteur des valeurs initiales. On peut mettre X_t sous une forme moyenne mobile et sous une forme autorégressive faisant intervenir Z_{-1} .

Propriété 4.1 Soit $(X_t)_{t \geq 0}$ un processus ARIMA(p, d, q) et Z_{-1} le vecteur de ses valeurs initiales.

(i) X_t peut s'écrire sous la forme moyenne mobile :

$$X_t = \sum_{j=0}^t h_j \eta_{t-j} + h^*(t) Z_{-1}, \quad (13)$$

où les h_j sont les coefficients de la division selon les puissances croissantes de Θ par Φ (en particulier $h_0 = 1$) et $h^*(t)$ est un vecteur ligne de fonctions de t .

(ii) X_t peut s'écrire sous la forme autorégressive :

$$\eta_t = X_t + \sum_{j=1}^t \pi_j X_{t-j} - g(t)Z_{-1}, \quad (14)$$

où les π_j sont les coefficients de la division selon les puissances croissantes de Φ par Θ et $g(t)$ est un vecteur ligne de fonctions de t tendant vers zéro lorsque t tend vers l'infini.

Preuve A faire en exercice.

4.1.3 Prédiction optimale des modèles ARIMA

Considérons un processus ARIMA(p, q, d) $(X_t)_{t \geq 0}$ défini par

$$\Phi(B)X_t = \Theta(B)\eta_t, \quad \forall t \geq 0$$

et les conditions initiales

$$Z_{-1} = (X_{-1}, \dots, X_{-p-d}, \eta_{-1}, \dots, \eta_{-q})'$$

non corrélées avec $\eta_0, \dots, \eta_t, \dots$

Comme les processus ARIMA(p, d, q) satisfont des contraintes linéaires, il est naturel de chercher une prédiction linéaire $\hat{X}_t(k)$ définie comme une combinaison linéaire des valeurs passées et présente du bruit blanc. La prédiction optimale de X_{t+k} pour $k > 0$ faite à la date t , notée $\hat{X}_t(k)$, est par définition la régression affine de X_{t+k} sur $(Z_{-1}; X_i, i = 0, \dots, t)$ ou, ce qui est équivalent, sur $(Z_{-1}; \eta_i, i = 0, \dots, t)$.

Les trois représentations AR(∞), MA(∞) et ARMA(p, q) nous aideront dans la prédiction.

Formule déduite de la forme moyenne mobile

La représentation moyenne mobile (13) du processus permet d'écrire

$$X_{t+k} = \sum_{j=0}^{t+k} h_j \eta_{t+k-j} + h^*(t+k)Z_{-1}.$$

La régression affine de X_{t+k} sur $(Z_{-1}; \eta_i, i = 0, \dots, t)$ est

$$\hat{X}_t(k) = \sum_{j=k}^{t+k} h_j \eta_{t+k-j} + h^*(t+k)Z_{-1},$$

ou

$$\hat{X}_t(k) = \sum_{j=0}^t h_{j+k} \eta_{t-j} + h^*(t+k)Z_{-1}. \quad (15)$$

On sait que cette régression affine peut s'interpréter comme une espérance conditionnelle dans deux cas : le cas où l'ensemble des variables aléatoires est normal et le cas où l'hypothèse de non corrélation des η_i entre eux et de Z_{-1} avec les $(\eta_i, i \geq 0)$ est remplacée par une hypothèse d'indépendance.

L'erreur de prédiction $e_t(k) = X_{t+k} - \hat{X}_t(k)$ s'obtient immédiatement

$$e_t(k) = \sum_{j=0}^{k-1} h_j \eta_{t+k-j}.$$

En particulier l'erreur de prévision à l'horizon 1 est

$$e_t(1) = \eta_{t+1}.$$

Formule de mise à jour

La formule (15) n'est pas directement utilisable puisqu'elle fait intervenir les η_j qui sont non observables. Elle permet cependant d'obtenir une formule de mise à jour intéressante :

$$\hat{X}_{t+1}(k-1) - \hat{X}_t(k) = \sum_{j=k-1}^{t+k} h_j \eta_{t+k-j} - \sum_{j=k}^{t+k} h_j \eta_{t+k-j} = h_{k-1} \eta_{t+1},$$

ou

$$\hat{X}_{t+1}(k-1) - \hat{X}_t(k) = h_{k-1} [X_{t+1} - \hat{X}_t(1)]. \quad (16)$$

Formule déduite de la forme autorégressive

La formule autorégressive (14) permet également de déduire une relation utile ; on a

$$X_{t+k} = - \sum_{j=1}^{t+k} \pi_j X_{t+k-j} + g(t+k)Z_{-1} + \eta_{t+k}.$$

Le terme $g(t+k)Z_{-1}$ devient négligeable pour t assez grand et on obtient l'approximation

$$X_{t+k} \approx - \sum_{j=1}^{t+k} \pi_j X_{t+k-j} + \eta_{t+k},$$

d'où

$$\hat{X}_t(k) = - \sum_{j=1}^{t+k} \pi_j \hat{X}_{t+k-j}, \quad (17)$$

avec

$$\hat{X}_{t+k-j} = \begin{cases} \hat{X}_t(k-j) & \text{si } k > j \\ X_{t+k-j} & \text{si } k \leq j \end{cases}$$

Formule déduite de la forme ARIMA

Enfin la formule de définition d'un ARIMA fournit la relation

$$X_{t+k} = - \sum_{j=1}^{p+d} \Phi_j X_{t+k-j} + \sum_{j=0}^q \theta_j \eta_{t+k-j},$$

avec $\Phi(B) = \sum_{j=0}^{p+d} \Phi_j B^j$, $\Theta(B) = \sum_{j=0}^q \theta_j B^j$ et $\phi_0 = \theta_0 = 1$.

On en déduit

$$\hat{X}_t(k) = - \sum_{j=1}^{p+d} \Phi_j \hat{X}_{t+k-j} + \sum_{j=1}^q \theta_j \hat{\eta}_{t+k-j}, \quad (18)$$

avec

$$\hat{X}_{t+k-j} = \begin{cases} \hat{X}_t(k-j) & \text{si } k > j \\ X_{t+k-j} & \text{si } k \leq j \end{cases}$$

et

$$\hat{\eta}_{t+k-j} = \begin{cases} 0 & \text{si } k > j \\ \eta_{t+k-j} & \text{si } k \leq j \end{cases}$$

Utilisation jointe de ces formules

Supposons que l'on souhaite effectuer des prédictions jusqu'à l'horizon K . A la date T , on doit calculer

$X_T(k)$ pour $k = 1, \dots, K$. A la date $T + 1$, on dispose d'une observation nouvelle qui doit être prise en compte pour modifier les prévisions de $X_{T+1} \dots X_{T+K}$ c'est-à-dire pour calculer $\hat{X}_{T+1}(k)$, $k = 1, \dots, K - 1$. On doit également déterminer une nouvelle prévision, celle de X_{T+K+1} . Ces calculs peuvent être menés en utilisant de façon jointe les diverses formules précédentes et après remplacement des divers paramètres par des estimateurs convergents. On procède donc comme suit :

1. On calcule les $\hat{X}_T(k)$, $k = 1, \dots, K$ à la première date T considérée à l'aide de (17) ; ce calcul fait, on n'aura plus à se servir des valeurs des X_s , $s \leq T$.
2. A la date $T + 1$, on connaît la valeur de X_{T+1} ; on peut donc calculer $\hat{X}_{T+1}(k)$, $k = 1, \dots, K - 1$ à l'aide de (16).
3. Il reste à calculer $\hat{X}_{T+1}(K)$ qu'on peut facilement obtenir à partir de (18) à condition d'avoir pris la précaution de prendre $K > q$ (et donc $\hat{\eta}_{t+K-i} = 0$, $i = 1, \dots, q$) et $K > p + d$ (ce qui évite d'avoir à garder trace des observations puisqu'on a $\hat{X}_{t+K-i} = \hat{X}_t(K - i)$, $i = 1, \dots, p + d$).

Connaissant $\hat{X}_{T+1}(k)$, $k = 1, \dots, K$, on peut répéter les phases de mise à jour (2) et (3) à la date $T + 2$ et ainsi de suite.

4.1.4 Exemples de processus ARIMA et prédiction

Exemple 4.4 La prévision linéaire $\hat{X}_t(k)$ pour un processus $AR(1)$ (=processus $ARIMA(1,0,0)$) à moyenne 0 satisfait la récursion Yule Walker

$$\hat{X}_t(k) = \phi \hat{X}_t(k - 1)$$

et donc est simplement

$$\hat{X}_t(k) = \phi^k X_t.$$

Pour un processus $AR(1)$ à moyenne connue μ , elle est donnée par

$$\hat{X}_t(k) - \mu = \phi^k (X_t - \mu).$$

Exemple 4.5 On considère la marche aléatoire simple. C'est un processus $ARIMA(0,1,0)$. La prévision linéaire $\hat{X}_t(k)$ satisfait donc la récursion Yule Walker

$$\hat{X}_t(k) = \hat{X}_t(k - 1)$$

et est donc constante

$$\hat{X}_t(k) = X_t$$

(c'est un cas limite de la formule dans l'exemple précédent sur la prévision $AR(1)$).

Exemple 4.6 On considère le processus $AR(2)$

$$\phi(B)X_t = (I - \lambda_1 B)(I - \lambda_2 B)X_t = \eta_t$$

(avec λ_1, λ_2 les inverses des racines de $\phi(z) = 0$).

– Montrer que les prévisions Box-Jenkins $\hat{X}_t(k)$ au temps t satisfont la récursion

$$\hat{X}_t(k) - (\lambda_1 + \lambda_2)\hat{X}_t(k - 1) + \lambda_1\lambda_2\hat{X}_t(k - 2) = 0.$$

– Montrer qu'elles sont de la forme : pour $\lambda_1 \neq \lambda_2$

$$\hat{X}_t(k) = A\lambda_1^k + B\lambda_2^k$$

et pour $\lambda_1 = \lambda_2 = \lambda$

$$\hat{X}_t(k) = (Ak + B)\lambda^k$$

– En déduire qu'en termes des deux dernières valeurs observées X_t et X_{t-1} , elles sont données par :

$$\hat{X}_t(k) = \begin{cases} \frac{\lambda_1^{k+1} - \lambda_2^{k+1}}{\lambda_1 - \lambda_2} X_t - \frac{\lambda_1 \lambda_2}{\lambda_1 - \lambda_2} (\lambda_1^k - \lambda_2^k) X_{t-1} & \text{si } \lambda_1 \neq \lambda_2 \\ \lambda^k (k + 1) X_t - \lambda^{k+1} k X_{t-1} & \text{si } \lambda_1 = \lambda_2 = \lambda \end{cases}$$

– En déduire $\lim_{k \rightarrow \infty} \hat{X}_t(k)$ dans le cas causal.

Il est facile d'étendre cette approche pour tous les processus autorégressifs ARIMA($p, d, 0$) à ordres p et d finis, et d'obtenir des formules explicites de prévision en termes de racines de l'équation $\phi(z) = 0$.

Exemple 4.7 On considère le processus ARIMA(1,1,0). Déduire de l'exercice précédent la formule de prévision Box-Jenkins pour ce processus. Calculer la limite $\lim_{k \rightarrow \infty} \hat{X}_t(k)$ pour un processus ARIMA(1,1,0).

Exemple 4.8 Plus généralement, dans le cas des processus ARIMA($p, d, 0$), la fonction de prévision $\hat{X}_t(k)$ est la solution de la récursion $\Phi(B)\hat{X}_t(k) = 0$ qui passe par les valeurs pivots $X_t, \dots, X_{t-p-d+1}$. Déduire la formule de prévision Box-Jenkins pour un processus ARIMA(1,2,0) et calculer sa limite lorsque $k \rightarrow \infty$.

Exemple 4.9 On considère un processus $(X_t)_t$ pour lequel la série différenciée deux fois est un bruit blanc, c'est à dire que X_t est un processus ARIMA(0,2,0). Montrer que la fonction de prévision Box-Jenkins est donnée par

$$\hat{X}_t(k) = X_t + k(X_t - X_{t-1}), \quad k \geq 0;$$

donc les prévisions se trouvent sur la droite qui passe par les deux derniers points.

Exemple 4.10 Plus généralement, la prévision linéaire $\hat{X}_t(k)$ pour le processus ARIMA(0, d , 0) est donnée par le polynôme d'ordre $d - 1$ qui passe par les d derniers points.

Définition 4.2 Les dernières $p + d$ valeurs ($\hat{X}_t(q), \hat{X}_t(q - 1), \dots, \hat{X}_t(q - d - p + 1)$) qui précèdent $\hat{X}_t(q + 1)$ (donc avant le point où la récursion de Yule Waker devient valable) s'appellent les **valeurs pivots** ou encore **valeurs pivotales**.

En conclusion de ces premiers exemples, nous voyons que le "type" de la fonction de prévision $\hat{X}_t(k)$ dans le cas des processus autorégressifs (sans partie MA) est déterminée complètement par la fonction $\phi(z)$ et les valeurs pivots. Dans la suite, nous verrons que cela reste vrai dans un certain sens pour les processus ARIMA(p, d, q).

Pour la prévision linéaire $\hat{X}_t(k)$ des processus ARIMA(p, d, q), nous avons eu besoin aussi d'une estimation de $\eta_{t-1}, \eta_{t-2}, \dots, \eta_{t-q}$, i.e. du "bruit inobservable passé" du modèle, ou au moins de $\eta_{-1}, \eta_{-2}, \dots, \eta_{-q}$, qu'on approxime parfois par 0. On peut aussi recourir à la représentation autorégressive ARIMA(1, d , 0), dans quel cas on aura besoin de $X_{t-1}, X_{t-2}, \dots, X_{t-q}$ qui sont aussi inobservables. Dans tous les cas, le résultat final demandera une approximation des valeurs précédant le début d'observation; l'approximation la plus simple dans l'absence des moyennes est $\eta_{t-1}, \eta_k = X_k = 0$ pour $k < 0$.

Exemple 4.11 On considère le processus MA(1) avec $|\theta| < 1$. On vérifie facilement que $\hat{X}_t(k) = 0$, pour $k \geq 2$. Pour $k = 1$, la définition :

$$X_{t+1} = \eta_{t+1} - \theta\eta_t$$

donne :

$$\hat{X}_t(1) = -\theta\eta_t.$$

Pour se débarrasser de η_t , on peut utiliser la représentation :

$$\eta_t = (I - \theta B)^{-1}\eta_t = \sum_{i=0}^{\infty} \theta^i X_{t-i}$$

Donc,

$$\hat{X}_t(1) = -\sum_{i=0}^{\infty} \theta^{i+1} X_{t-i}.$$

Si on suppose que “l’information” est finie, i.e. $\mathcal{F}_t = \{X_t, X_{t-1}, \dots, X_1\}$, la formule de prévision obtenue par récursion est :

$$\begin{aligned}\hat{X}_t(1) &= \mathbb{E}(X_{t+1}|\mathcal{F}_t) \\ &= -\sum_{i=0}^{t-1} \theta^{i+1} X_{t-i} - \sum_{i=t}^{\infty} \theta^{i+1} X_{t-i} \\ &= -\sum_{i=0}^{t-1} \theta^{i+1} X_{t-i} - \sum_{j=0}^{\infty} \theta^{j+t+1} X_{-j} \\ &= -\sum_{i=0}^{t-1} \theta^{i+1} X_{t-i} - \theta^{t+1} \eta_0\end{aligned}$$

Comme η_0 n’est pas connu, en pratique on tronque et on utilise : $\hat{X}_t(1) = -\sum_{i=0}^{t-1} \theta^{i+1} X_{t-i}$. Si $|\theta| < 1$ et t est grand, la différence sera négligeable et l’approximation bonne.

Cet exemple montre qu’une estimation du “bruit inobservable” η_t ou au moins de η_0 est incontournable pour les modèles ARIMA avec $q \geq 1$.

Les inconnues peuvent être enlevées en utilisant la représentation inverse “ π ” du bruit en fonction de la série ou en utilisant $\hat{\eta}_t = X_t - \hat{X}_{t-1}(1)$ (les dernières se calculent récursivement). Une estimation arbitraire de η_0, η_{-1}, \dots ou de X_0, X_{-1}, \dots sera nécessaire.

Remarque 4.2 Pour $k > q$, la formule est exactement la récurrence homogène Yule-Walker $\phi(B)\hat{X}_t(k) = 0$, et donc la prévision sera donnée par la solution de cette équation qui passe par les $p+d$ points pivots $X_{t+q}, \dots, X_{t+q-p-d+1}$, ajustées en tenant compte du bruit η_t , comme indiqué en (33). Par exemple, pour les processus ARIMA(0, d, q) la fonction de prévision $\hat{X}_t(k)$ est un polynôme d’ordre $d-1$ en k .

4.2 Processus SARIMA

Si l’on veut en même temps traiter les saisonnalités de période s (sans supposer une répétition exacte, déterministe, des données), on est amenés à définir les processus SARIMA.

4.2.1 Motivation et définition

La principale motivation des processus SARIMA est de modéliser des données saisonnières, par exemple mensuelles ($s = 12$). Soit

$$Z_t^{[m]} := X_{m+12t}, \quad m = 1, \dots, 12.$$

On formule l’hypothèse selon laquelle chaque série $Z_t^{[m]}$ suit le même modèle ARMA(P, Q) :

$$Z_t^{[m]} = \sum_{k=1}^P \phi_k Z_{t-k}^{[m]} + V_t^{[m]} + \sum_{j=1}^Q \theta_j V_{t-j}^{[m]}$$

où $V_t^{[m]}$ est un bruit blanc si le mois m est fixe.

Mais si on définit (V_t) par $V_{m+12t} = V_t^{[m]}$, $m = 1, \dots, 12$, V_t n’est plus non corrélée : on a donc aussi un modèle ARMA(p, q) pour la dépendance inter-mensuelle.

Ainsi on a le diagramme suivant

$$\begin{array}{ccc} X_t & \longrightarrow & Z_t^{[m]} := X_{m+12t} \\ \downarrow & & \downarrow \\ V_t & \longleftarrow & V_t^{[m]} \end{array}$$

avec V_t tel que $V_{m+12t} := V_t^{[m]}$ et on a deux dépendances à modéliser :

- la dépendance **intra-mensuelle** des $Z_t^{[m]}$ pour tous les m . On la modélise par un ARMA(p, q) avec le même couple (p, q) pour tous les m .
- la dépendance **inter-mensuelle** de V_t . On la modélise par un ARMA(P, Q).

Le modèle complet s'écrit alors

$$\begin{aligned} \text{Dép. intra-mensuelle} \quad X_t &= \sum_{k=1}^P f_k X_{t-12k} + V_t + \sum_{j=1}^Q g_j V_{t-12j} \\ \text{càd} \quad F(B^{12})X_t &= G(B^{12})V_t \\ \text{Dép. inter-mensuelle} \quad \Phi(B)V_t &= \Theta(B)\eta_t \end{aligned}$$

où (η_t) est un bruit blanc. En rassemblant les résultats on obtient la relation

$$\Phi(B)F(B^{12})X_t = \Phi(B)G(B^{12})V_t = G(B^{12})\Phi(B)V_t = \Theta(B)G(B^{12})\eta_t.$$

Les polynômes

- Φ donne le p ;
- Θ donne le q ;
- F donne le P ;
- G donne le Q ;

La combinaison des séries intra-mensuelle et inter-mensuelle livre un processus ARMA(p, q) \times (P, Q)₁₂.

Si on y inclut également les tendances ($d > 0$) et les composantes déterministes périodiques ($D > 0$), on a le modèle SARIMA(p, d, q) \times (P, D, Q)_s. On a naturellement la définition suivante

Définition 4.3 Soient p, q, d et $s \geq 0$ et $P, Q, D \geq 0$. Un processus (X_t) est un processus SARIMA(p, d, q) \times (P, D, Q)_s [Autoregressif moyenne mobile et saisonnalité intégrées] si

$$Y_t = \Delta^d \Delta_s^D X_t = (I - B)^d (I - B^s)^D X_t, \quad (19)$$

est un processus ARMA stationnaire de la forme

$$\Phi(B)F(B^s)Y_t = \Theta(B)G(B^s)\eta_t \quad (20)$$

où Φ (resp. Θ) est le polynôme générateur d'un AR(p) (resp. d'un MA(q)) :

$$\Phi(z) = 1 - \sum_{k=1}^p \phi_k z^k \quad \text{et} \quad \Theta(z) = 1 - \sum_{k=1}^q \theta_k z^k,$$

et où, pour la saisonnalité $Y_t - Y_{t-s}$, F (resp. G) est le polynôme générateur d'un AR(P) (resp. d'un MA(Q)) :

$$F(z) = 1 - \sum_{k=1}^P f_k z^k \quad \text{et} \quad G(z) = 1 - \sum_{k=1}^Q g_k z^k.$$

Exemple 4.12 – $X_t = \mu + S_t + \eta_t$, où $S_t = S_{t+s} \quad \forall t \geq 0$, est un processus SARIMA(0, 0, 0) \times (0, 1, 1)_s. En effet,

$$Y_t = (I - B^s)X_t = \eta_t - \eta_{t-s}$$

est un ARMA(0, s) (non inversible).

- Si on étudie le modèle SARIMA(1, 1, 0) \times (1, 1, 0)₁₂, on a ainsi $s = 12$ (données mensuelles), $p = 1 = P, q = 0 = Q, d = D = 1$. Donc Y_t est défini par :

$$Y_t = (I - B)(I - B^{12})X_t$$

et on a la relation

$$Y_t = \phi Y_{t-1} + f Y_{t-12} - \phi f Y_{t-13} + \eta_t$$

car $\Phi(z)F(z^s) := (1 - \phi z)(1 - f z^{12}) = (1 - \phi z - \phi f z^{13})$. Ainsi Y_t est un AR(13) = AR($p + Ps$) ou encore $Y_{t-1} - Y_{t-12}$ est un AR(1).

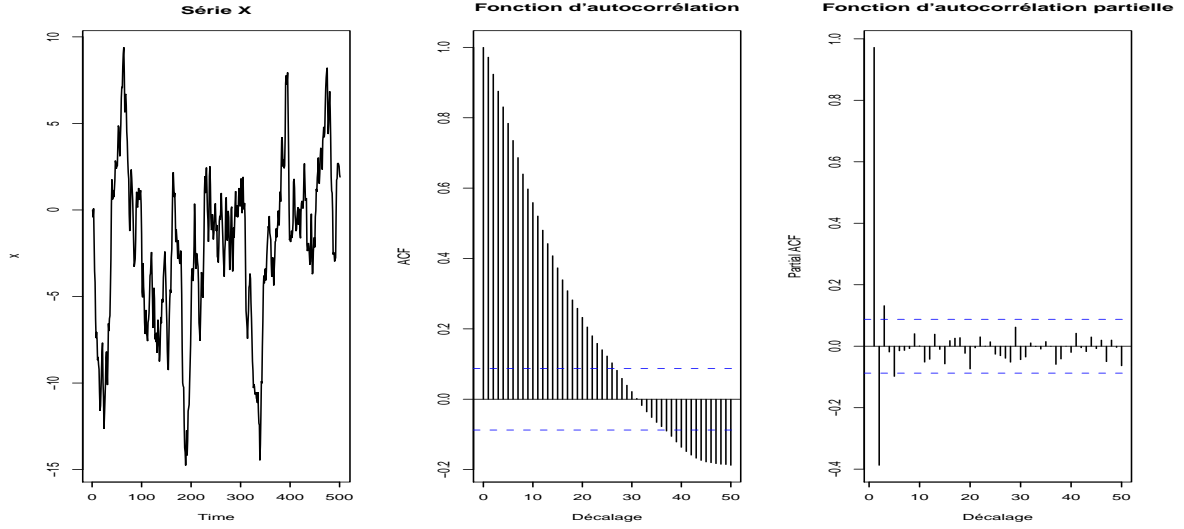


FIGURE 13 – Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus $\text{SARIMA}(1, 1, 1) \times (0, 0, 0)_{12}$

4.2.2 Prédiction des processus SARIMA

Si les données X_1, \dots, X_T suivent un modèle $\text{SARIMA}(p, d, q) \times (P, D, Q)_s$, on définit

$$Y_t = (I - B)^d (I - B^s)^D X_t$$

Si par exemple, $(d, D, s) = (1, 1, 12)$,

$$Y_t = (X_t - X_{t-12}) - (X_{t-1} - X_{t-13}) \quad (21)$$

On traite le problème de prédiction du processus Y_t du type $\text{ARMA}(p + Ps, q + Qs)$ comme vu dans la Section 3.5, et on rentre après au niveau du processus X_t du type SARIMA en exprimant $\hat{X}_T(1)$ en fonction de $\hat{Y}_T(1)$ et les valeurs observées X_t pour $t \leq T$. Nous expliquons ce principe brièvement dans le cadre de deux exemples :

Exemple 4.13 (a) En cas d'un processus purement autorégressif $\text{AR}(p + Ps)$ de Y_t , par exemple pour $p = 1, P = 1, s = 12$, c'est-à-dire pour le processus $\text{AR}(13)$,

$$Y_t = \phi Y_{t-1} + f Y_{t-12} - \phi f Y_{t-13} + \eta_t,$$

la prédiction linéaire de cette partie autorégressive est donnée par

$$\hat{Y}_T(1) = \phi Y_T + f Y_{T-11} - \phi f Y_{T-12}$$

Donc, la forme de l'équation (21) implique la prédiction d'un processus X_t du type $\text{SARIMA}(1, 1, 0) \times (1, 1, 0)_{12}$ par

$$\hat{X}_T(1) = \hat{Y}_T(1) + X_{T-11} + X_T + X_{T-12}.$$

(b) Si on veut inclure le traitement d'une partie $\text{MA}(q + Qs)$ de Y_t , on passe par le prédicteur d'un processus ARMA en utilisant les résidus. Par exemple, pour la partie Y_t stationnaire d'un processus $\text{SARIMA}(1, 1, 1) \times (1, 1, 1)_{12}$, les résidus sont donnés par :

$$\eta_t := Y_t - \phi Y_{t-1} - f Y_{t-12} + \phi f Y_{t-13} - \theta \eta_{t-1} - g \eta_{t-12} + \theta g \eta_{t-13}$$

Le passage au niveau des observations originales X_t se fait comme dans le cas (a), c'est-à-dire par simple inversion des opérateurs de différences d'ordre 1 et 12.

Remarque 4.3 Cette technique de prévision est valable si T est suffisamment grand et l'horizon h n'est pas trop grand. Dans le cas contraire, on utilise des méthodes non paramétriques, comme le lissage exponentiel.

Remarque 4.4 Le lissage exponentiel est une méthode alternative à la prédiction SARIMA qui est utile, par exemple, si T est petit (dans ce cas, la qualité du modèle ajusté peut être médiocre) ou si il y a des changements de la structure de la série (des ruptures).

4.3 Identification et ajustement des modèles ARIMA/SARIMA

L'identification et l'ajustement des modèles ARIMA/SARIMA aux données X_1, \dots, X_T se fait selon le même principe que pour les processus ARMA et selon le plan suivant :

(0) Appliquer (si nécessaire) une transformation de type "Box-Cox" pour stabiliser la variance des données (parfois la variance augmente en fonction du niveau de la série. Les deux transformations les plus importantes de type Box-Cox sont : $\log(X_t)$, $\exp X_t$.

(1) Choisir d , D et s (souvent il suffit de prendre $0 \leq d \leq 2$, $0 \leq D \leq 1$) de sorte que

$$(I - B)^d(I - B^s)^D X_t = Y_t$$

soit stationnaire.

(2) Calculer les autocorrélations $\hat{\rho}$ et les autocorrélations partielles $\hat{\tau}$ empiriques et examiner les $\hat{\rho}(ks)$ et $\hat{\tau}(ks)$ pour $k \geq 0$, pour trouver (P, Q) de sorte que $\hat{\rho}(ks)$ et $\hat{\tau}(ks)$ correspondent à un modèle ARMA(P, Q) (et décroissent exponentiellement).

(3) Choisir (p, q) tels que $\hat{\rho}(1), \dots, \hat{\rho}(s-1)$ et $\hat{\tau}(1), \dots, \hat{\tau}(s-1)$ soient compatibles avec un modèle ARMA(p, q).

(4) Estimer par maximum du vraisemblance les paramètres $\phi_k, \theta_k, f_k, g_k$ et σ_η^2 .

(5) Les étapes (2)-(4) peuvent se faire selon plusieurs alternatives : calculer $AIC(p, q)$, $AIC(P, Q)$ (ou $AIC(p+Ps, q+Qs)$) et prendre le minimum. Une autre stratégie serait, au lieu de poursuivre les étapes (2)-(5), de calculer le critère AIC pour les modèles SARIMA($p, 0, q$) \times ($P, 0, Q$) $_s$ pour les Y_t .

(6) Contrôler la qualité du modèle par analyse des résidus :

$$\eta_t(p, q) = X_t - \sum_{k=1}^p \hat{\phi}_k(p, q) X_{t-k} - \sum_{k=1}^q \hat{\phi}_k(p, q) \eta_{t-k}.$$

Par exemple, inspection des fonctions d'autocorrélation et d'autocorrélation partielle empiriques des résidus, ou application des tests d'hypothèse sur bruit blanc.

5 Processus ARCH et GARCH

Les séries précédemment étudiées étaient supposées stationnaires ou non-stationnaires mais étaient dans tous les cas “homoscédastique” c’est-à-dire avaient une variance constante dans le temps. Si besoin, tendances et saisonnalités étaient supprimées pour obtenir une série résiduelle stationnaire ou du moins une série à accroissements stationnaires. Néanmoins, toutes les séries résiduelles obtenues de la sorte ne sont pas nécessairement stationnaires : il peut arriver que la variance d’un processus varie au cours du temps. C’est le cas par exemple de la série représentée par la figure 14, qui contient les évolutions journalières de la bourse des valeurs de New-York (NYSE) du 19 octobre 1984 au 31 décembre 1991.

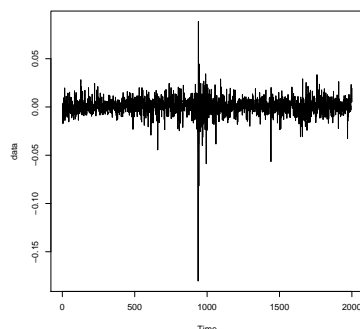


FIGURE 14 – Evolution journalière de la bourse des valeurs de New-York (1984-1991)

Comme on peut le voir, la moyenne semble constante alors que la variance change au cours du temps (on qualifie ce comportement d’“hétéroscédastique”). De plus, les moments de grande variabilité semblent regroupés. Les modèles de type ARIMA qui supposent un comportement “homoscédastique” (variance constante), ne sont pas adaptés à ce type de série. Nous présentons dans cette section des modèles adaptés à ce type de série : les processus ARCH (AutoRegressive Conditionally Heteroscedastic) introduits par Engle vers 1982, ainsi que leur généralisation, les processus GARCH.

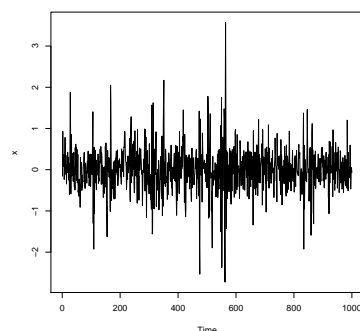


FIGURE 15 – Simulation d’un ARCH(2)

Ces modèles permettent de prendre en compte une variation de la variance au cours du temps (et, plus largement, des moments d’ordres supérieurs ou égaux à deux). Dans la théorie des processus ARMA, la variance d’une série est (entre autres) déterminée par la variance du processus des innovations. Pour simplifier, considérons le bruit blanc gaussien

$$X_t = \eta_t$$

où $\eta_t \sim \mathcal{N}(0, \sigma^2)$. La variance de ce processus est la variance du processus des innovations (η_t), c'est-à-dire σ^2 . Nous souhaitons à présent tenir compte d'un éventuel changement au cours du temps de la variance, provenant de l'évolution passée du processus (variance instantanée). C'est ce que permettent les modèles ARCH, dont l'idée est de déterminer la distribution de η_t conditionnellement à toutes les valeurs passées X_{t-1}, X_{t-2}, \dots .

5.1 Les processus ARCH

Définition

Définition 5.1 *Un processus ARCH(p) [AutoRegressive Conditional Heteroskedasticity] est défini par*

$$X_t = \epsilon_t$$

avec

$$\epsilon_t | X_{t-1}, X_{t-2}, \dots \sim \mathcal{N}(0, \sigma_t^2)$$

et

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_p X_{t-p}^2$$

Comme sa définition l'indique, la variance conditionnelle d'un processus ARCH au temps t dépend des valeurs passées ($X_{t-i}, i > 0$) et n'est donc pas constante (puisque les valeurs de X_{t-i} changent). Cette dépendance au passé est précisée par la dernière équation, qui relie linéairement σ_t^2 au carré des p dernières observations et modélise la variation de ces observations autour de leur moyenne zéro. Cette relation peut être expliquée intuitivement comme suit : si les observations aux temps $t-1, \dots, t-p$ sont élevées (en valeur absolue), alors la variance σ_t^2 aura tendance à être également élevée. De la même façon, de petites observations X_{t-1}, \dots, X_{t-p} en valeur absolue seront suivies (en tendance) par une petite variance. Nous avons observé dans l'introduction que les périodes de haute volatilité suivent toujours un choc important dans la série et s'aggrègent après celui-ci. Nous voyons maintenant que la dernière relation intervenant dans la définition du modèle ARCH permet d'expliquer ce regroupement en clusters de la volatilité. De plus, l'ordre p du modèle détermine la durée de persistance du choc.

Propriétés

Si X_t est un processus ARCH(p), alors

$$\begin{aligned} \mathbb{E}(X_t) &= 0 \\ \mathbb{E}(X_t | \mathcal{F}_{t-1}) &= 0 \\ \text{Var}(X_t) &= \frac{\alpha_0}{1 - \sum_{i=1}^p \alpha_i} \quad \text{si} \quad \sum_{i=1}^p \alpha_i < 1 \\ \text{Var}(X_t | \mathcal{F}_{t-1}) &= \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_p X_{t-p}^2 \\ \text{Cov}(X_t, X_{t+h}) &= 0 \quad \forall h > 0 \\ \text{Cov}(X_t, X_{t+h} | \mathcal{F}_{t-1}) &= 0 \quad \forall h > 0 \end{aligned}$$

Remarque 5.1 *Un processus ARCH est conditionnellement hétéroscédastique (car la variance conditionnelle varie dans le temps) mais inconditionnellement homoscedastique (la variance non conditionnelle est invariante dans le temps)!*

Proposition 5.1 (i) *Condition suffisante de stationnarité : $\sum_{i=1}^p \alpha_i < 1$.*

(ii) *La distribution d'un processus ARCH a*

- *un skewness nul (moment centré d'ordre 3) : la distribution est donc symétrique,*
- *un kurtosis (moment centré d'ordre 4 standardisé) supérieur à 3 : la distribution est donc plus aplatie qu'une gaussienne.*

Preuve (i) D'une part, la première équation montre que la moyenne de ces processus est constante. Ensuite, la cinquième assure que la covariance entre les éléments de la série ne dépend que du temps séparant ces éléments. En conséquence, le processus est faiblement stationnaire si sa variance existe (c'est-à-dire, si elle est finie). Par la troisième équation, on déduit donc qu'une condition suffisante de stationnarité est

$$\sum_{i=1}^p \alpha_i < 1.$$

(ii) Si, on s'intéresse maintenant aux moments d'ordre supérieur des processus ARCH afin d'analyser d'autres caractéristiques de ces modèles. Puisque, les innovations ϵ_t sont conditionnellement distribuées suivant une loi normale, le troisième moment conditionnel $\mathbb{E}(\epsilon_t^3 | \mathcal{F}_{t-1}) = 0$, ce qui implique que le coefficient de skewness de la distribution de X_t est nul. En conséquence, la distribution de X_t , puisqu'elle est continue, est également symétrique. L'analyse du quatrième moment standardisé (le kurtosis, ou coefficient d'aplatissement de Fisher) donne une indication sur l'aplatissement de la distribution de X_t . Il est défini, pour un processus à moyenne zéro, par

$$\gamma_2 = \frac{\mathbb{E}(X_t^4)}{\mathbb{E}(X_t^2)^2}$$

et est égal à 3 dans le cas d'une distribution normale. Un coefficient > 3 indique une distribution plus aplatie que celle de la loi normale et inversement s'il est < 3 . Il est difficile d'écrire une expression générale concise du kurtosis d'un processus ARCH(p). Néanmoins, on peut le calculer explicitement si X_t est un processus ARCH(1) : il est donné dans ce cas par

$$\gamma_2 = \frac{\mathbb{E}(X_t^4)}{\mathbb{E}(X_t^2)^2} = 3 \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2}, \quad \text{si } 3\alpha_1^2 < 1,$$

(si $3\alpha_1^2 \geq 1$ alors le quatrième moment du processus n'existe pas). Puisque $0 < \alpha_1 < 1/\sqrt{3} < 1$, on en déduit que $\gamma_2 > 3$ et donc que la distribution de X_t est plus aplatie que celle de la normale. Ce résultat se généralise pour tout p . \square

Cette dernière propriété est intéressante pour détecter un ARCH en pratique.

5.2 Les processus GARCH

Définition

Définition 5.2 Un processus GARCH(p, q) [Generalized ARCH] est défini par :

$$X_t = \epsilon_t$$

avec

$$\epsilon_t | X_{t-1}, X_{t-2}, \dots \sim \mathcal{N}(0, \sigma_t^2)$$

et

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_p X_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_q \sigma_{t-q}^2$$

avec $\alpha_0 > 0$, $\alpha_i \geq 0$ pour $i = 1, \dots, p$ et $\beta_j \geq 0$ pour $j = 1, \dots, q$.

Un processus GARCH peut être vu comme un processus ARCH d'ordre infini. Ainsi, la généralisation des processus ARCH aux processus GARCH est similaire à la généralisation des processus autorégressifs aux processus ARMA. Un processus GARCH peut ainsi représenter formellement de façon plus parcimonieuse un processus ARCH comprenant un nombre élevé de paramètres.

Remarque 5.2 Un GARCH($p, 0$) est un ARCH(p).

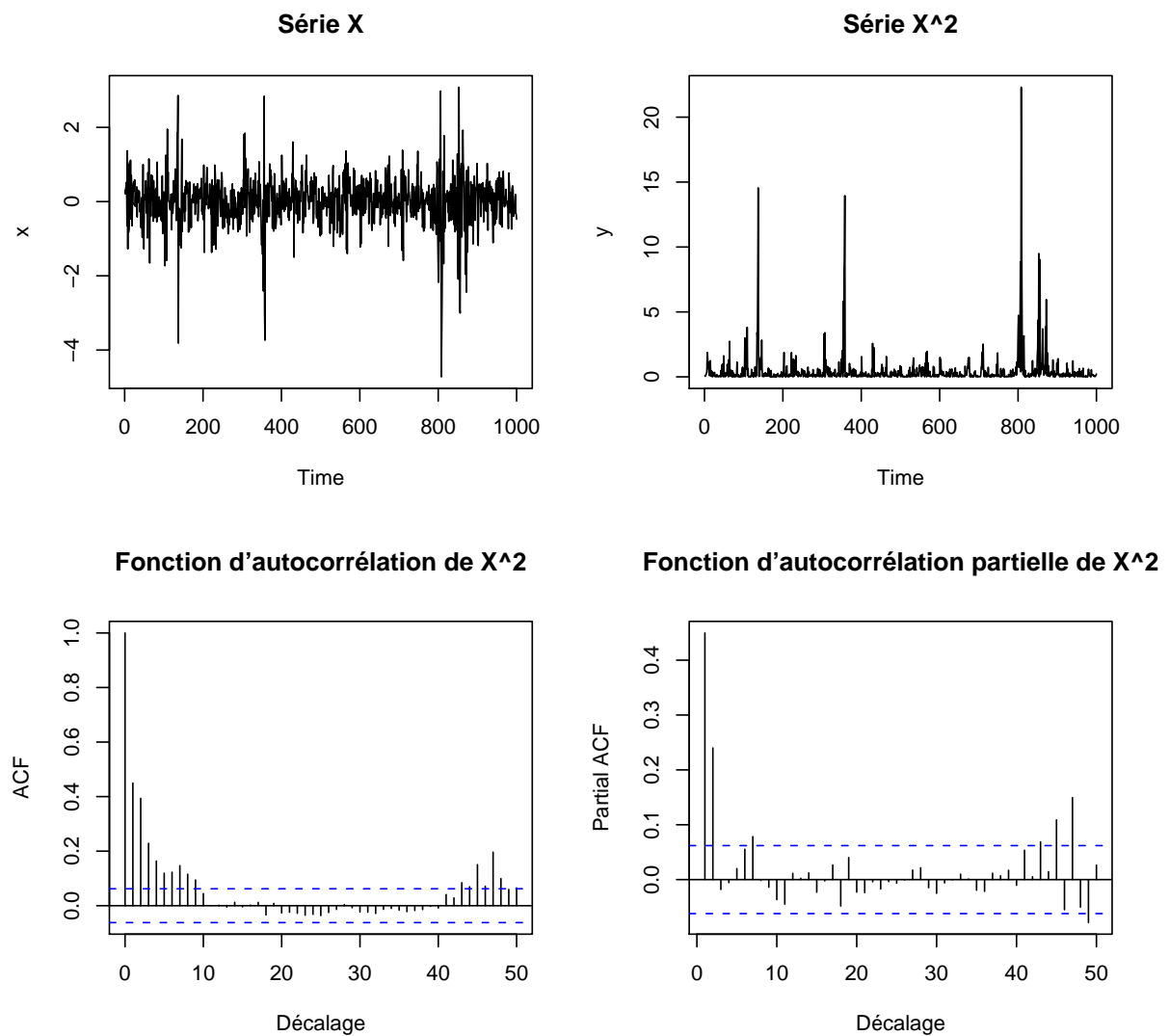


FIGURE 16 – Simulation d'un ARCH(2)

Propriétés

Proposition 5.2 *Si X_t est un processus GARCH(p, q), alors*

$$\begin{aligned}\mathbb{E}(X_t) &= 0 \\ \mathbb{E}(X_t | \mathcal{F}_{t-1}) &= 0 \\ \text{Cov}(X_t, X_{t+h}) &= \sigma_h = 0 \quad \forall h > 0 \\ \text{Cov}(X_t, X_{t+h} | \mathcal{F}_{t-1}) &= 0\end{aligned}$$

Propriété 5.1 *Soit X_t un processus GARCH(p, q), et soit $m = \max(p, q)$. Le processus X_t^2 admet une représentation ARMA(m, q).*

Ainsi, pour identifier un GARCH(p, q), on identifiera tout d'abord le processus ARMA(m, q) qui modélise X_t^2 . Pour identifier p dans le cas où $m = q$ ($p = q$), il faut effectuer des tests de significativité

des coefficients a_q, \dots, a_1 du processus ARMA(m, q) (sont-ils significativement non nuls?).

5.3 Les processus ARMA-GARCH

Weiss (1984) eut l'idée d'appliquer la théorie des modèles GARCH aux erreurs d'une régression linéaire ou à un processus d'innovation d'un modèle ARMA. Dans le premier cas, le modèle consiste à considérer un modèle de régression linéaire avec erreurs GARCH :

$$Y_t = aX_t + \epsilon_t$$

avec $\epsilon_t \sim \text{GARCH}(p, q)$. Ce modèle est appelé modèle de régression avec erreurs GARCH.

Dans le deuxième cas, le modèle consiste en un processus ARMA avec un processus d'innovations GARCH :

$$\phi(B)X_t = \Theta(B)\epsilon_t$$

avec $\epsilon_t \sim \text{GARCH}(p, q)$. Ce modèle est appelé modèle ARMA-GARCH.

Identification

Nous savons que l'outil de base permettant d'identifier l'ordre d'un modèle ARMA est la fonction d'autocorrélation ou d'autocorrélation partielle des observations X_t . Cet outil est en réalité toujours valable dans le cas d'un modèle ARMA avec processus d'innovation GARCH. En effet, par les propriétés des processus GARCH étudiées précédemment, les innovations du modèle sont de moyenne nulle et non corrélées. On en déduit facilement que le comportement des fonctions d'autocorrélation et d'autocorrélation partielle sont identiques aux modèles ARMA étudiés précédemment. L'identification de la partie ARMA du modèle s'effectue donc de la même manière que dans le cas des processus linéaires étudiés dans les chapitres précédents. Cependant, les innovations GARCH imposent une modification du théorème central limite pour la fonction d'autocovariance $\text{gamma}^X(h)$. Weiss (1984) a montré que les innovations ARCH font croître l'erreur standardisée de l'estimateur de la fonction d'autocorrélation. En conséquence, les bandes de confiance pour le test de significativité de la fonction empirique doivent également être adaptées.

Supposons maintenant qu'un modèle ARMA ait été identifié pour X_t et que ses paramètres sont estimés. Nous pouvons donc écrire formellement :

$$\hat{\phi}(B)X_t = \Theta(B)\epsilon_t$$

et calculer les résidus. Comme nous l'avons étudié dans les chapitres précédents, le test portmanteau permet à ce niveau de conclure si des corrélations linéaires persistent dans les résidus et. Par la propriété de non corrélation des processus GARCH, nous savons que ce test ne suffit pas pour conclure qu'on ne peut ajuster un tel modèle aux résidus. Or, la présence d'un modèle GARCH dans les résidus peut être la cause d'une mauvaise identification des paramètres si on n'en tient pas compte pour modéliser le processus d'innovations. En conséquence, le développement de méthodes pour tester si une composante GARCH est présente ou non dans le processus d'innovations est très important.

Une méthode pour identifier une telle dépendance non linéaire des résidus consiste à utiliser à nouveau le test Portmanteau sur le carré des résidus e_t^2 . La procédure pour tester s'il réside des corrélations sur e_t^2 est alors la suivante :

- Calcul de la moyenne de e_t^2

$$\hat{\nu}_T = \frac{1}{T} \sum_{t=1}^T e_t^2$$

- Calcul de la fonction d'autocorrélation empirique de e_t^2

$$\hat{\rho}_T^{e^2}(h) = \frac{\sum_{t=1}^{T-h} (e_{t+h}^2 - \hat{\nu}_T)(e_t^2 - \hat{\nu}_T)}{\sum_{t=1}^{T-h} (e_{t+h}^2 - \hat{\nu}_T)^2}$$

– Calcul de la statistique du test Portmanteau

$$Q_T(e^2) = \sum_{i=1}^K \left(\hat{\rho}_T^{e^2}(h) \right)^2$$

Sous l'hypothèse nulle selon laquelle les e_t^2 sont non corrélés, cette statistique est distribuée asymptotiquement selon une loi χ_K^2 . En cas de rejet de l'hypothèse nulle, il faudra donc envisager d'ajuster un modèle (G)ARCH au processus des innovations. Pour identifier quels ordres peuvent être ajustés à ce processus, rappelons que si ϵ_t est un processus GARCH(p, q), alors le ϵ_t^2 est un processus ARMA(m, q), où $m = \max(p, q)$. Ceci implique que l'analyse des autocorrélations du carré des résidus e_t^2 peut être mise en oeuvre afin de trouver m et q , et d'en déduire p .

5.4 Estimation des paramètres

Estimation des paramètres ARCH

L'estimation des paramètres de modèles ARCH se base sur la maximisation de la fonction de vraisemblance. Par hypothèse, X_t est conditionnellement gaussien. La vraisemblance associée à X_t conditionnellement au passé \mathcal{F}_{t-1} est donc

$$\mathcal{L}(x_t | \mathcal{F}_{t-1}; \alpha) = \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left(-\frac{x_t^2}{2\sigma_t^2}\right)$$

et dépend du vecteur de paramètres $\alpha = (\alpha_0, \dots, \alpha_p)'$ à travers σ_t . La fonction de vraisemblance de $(x_1, \dots, x_T)'$ conditionnelle à \mathcal{F}_0 est par conséquent

$$\mathcal{L}_T(x_1, \dots, x_T; \alpha) = \prod_{t=1}^T \mathcal{L}(x_t | \mathcal{F}_{t-1}; \alpha)$$

L'estimateur est alors défini comme le vecteur $\hat{\alpha}_T = (\alpha_{0,T}, \dots, \alpha_{p,T})'$ qui maximise le logarithme de cette fonction de vraisemblance :

$$\hat{\alpha}_T = \operatorname{argmax} \log \mathcal{L}_T(x_1, \dots, x_T; \alpha)$$

Estimation des paramètres GARCH

Remarquons que l'estimation par maximum de vraisemblance d'un modèle ARMA est rendue plus difficile que celle d'un processus autorégressif pur, puisque le processus d'innovations n'est pas directement observé. Le même phénomène survient lorsqu'on tente de maximiser la vraisemblance d'un processus GARCH. En effet, la vraisemblance associée à X_t conditionnellement au passé \mathcal{F}_{t-1} s'écrit

$$\mathcal{L}(x_t | \mathcal{F}_{t-1}; \alpha) = \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left(-\frac{x_t^2}{2\sigma_t^2}\right)$$

mais cette fois, la variance σ_t^2 dépend des valeurs passées de la variance conditionnelle $\sigma_{t-1}^2, \dots, \sigma_{t-q}^2$. Ces valeurs n'étant pas observées en pratique, la maximisation directe de la vraisemblance est rendue impossible.

En pratique, on estime successivement les valeurs de $\sigma_1^2, \dots, \sigma_{t-1}^2$ avant de calculer la vraisemblance. Ainsi, pour un vecteur $\alpha^0 = (\alpha_0^0, \dots, \alpha_p^0, \beta_1^0, \dots, \beta_q^0)'$ fixé de paramètres, on calcule récursivement

$$\hat{\sigma}_s^2 = \alpha_0^0 + \sum_{i=1}^p \alpha_i^0 X_{s-i}^2 + \sum_{j=1}^q \beta_j^0 \hat{\sigma}_{s-j}^2$$

avec la convention $X_i = 0$ et σ_i^2 si $i \leq 0$. On remplace donc la fonction de vraisemblance par

$$\hat{\mathcal{L}}(x_t | \mathcal{F}_{t-1}; \alpha^0) = \frac{1}{\sqrt{2\pi}\hat{\sigma}_t} \exp\left(-\frac{x_t^2}{2\hat{\sigma}_t^2}\right)$$

et la fonction de vraisemblance totale est

$$\widehat{\mathcal{L}}_T(x_1, \dots, x_T; \alpha^0) = \prod_{t=1}^T \widehat{\mathcal{L}}(x_t | \mathcal{F}_{t-1}; \alpha^0)$$

Cette fonction de vraisemblance peut être calculée pour différentes valeurs du vecteur α^0 et sa maximisation livre l'estimateur de maximum de vraisemblance.

5.5 Prédiction

Une application importante de la théorie des modèles ARCH consiste à évaluer la précision de prévision des valeurs futures d'une série chronologique. Dans le cas d'un processus ARMA, nous pouvons voir que la variance des prévisions dépend de l'horizon de prédition et de la variance inconditionnelle de la série. En particulier, cette variance est indépendante du comportement local de la volatilité du processus à l'instant où on s'apprête à calculer les intervalles de prévision. Par contre, en ajustant un modèle GARCH, nous allons voir comment il est possible d'utiliser cette volatilité locale pour mesurer les intervalles de prévision. Partons du processus ARMA(r, s)

$$\phi(B)X_t = \Theta(B)\epsilon_t \quad (22)$$

où $\Phi(B) = 1 - \phi_1 B - \dots - \phi_r B^r$ et $\Theta(B) = 1 - \theta_1 B - \dots - \theta_s B^s$ et ϵ_t est un processus GARCH(p, q). Supposons que l'on observe cette série jusqu'au temps t . Toute prévision $\hat{X}_t(h)$ à l'instant $t + h$ reproduisant la structure de la série est de la forme

$$X_{t+h} = \sum_{i=1}^r \phi_r X_{t+h-i} + \sum_{j=1}^s \theta_j \epsilon_{t+h-j} + \epsilon_{t+h}.$$

Le prédicteur optimal $\hat{X}_t(h)$ est donné par

$$\hat{X}_t(h) = \mathbb{E}(X_{t+h} | \mathcal{F}_t) = \sum_{i=1}^r \phi_r \mathbb{E}(X_{t+h-i} | \mathcal{F}_t) + \sum_{j=1}^s \theta_j \mathbb{E}(\epsilon_{t+h-j} | \mathcal{F}_t)$$

puisque $\mathbb{E}(\epsilon_{t+h} | \mathcal{F}_t) = 0$. À cause des diverses propriétés que nous avons décrites sur la moyenne conditionnelle des processus GARCH, on peut effectuer dans cette dernière relation quelques simplifications. L'essentiel est que, jusqu'à présent, la prévision est formellement identique à celle développée dans les chapitres consacrés aux processus linéaires. En d'autres termes, à ce niveau, la présence d'un modèle (G)ARCH au niveau du processus d'innovations ne modifie pas la procédure de prévision.

Calculons à présent l'erreur de prévision dans le but de déterminer les intervalles de prévision. En considérant que les paramètres ϕ_i sont compatibles avec l'hypothèse d'inversibilité du processus, le modèle (22) peut se réécrire sous la forme d'un processus MA(∞), et on peut donc écrire :

$$X_{t+h} = \sum_{i=0}^{\infty} \gamma_i \epsilon_{t+h-i} \quad (23)$$

où γ_i est le coefficient du terme z_i dans le développement de $\phi^{-1}(z)\Theta(z)$. En utilisant cette représentation pour calculer le prédicteur optimal, on peut écrire :

$$\hat{X}_t(h) = \sum_{i=h}^{\infty} \gamma_i \epsilon_{t+h-i}$$

et on en déduit l'erreur de prévision :

$$e_t(h) = X_{t+h} - \hat{X}_t(h) = \sum_{i=0}^{h-1} \gamma_i \epsilon_{t+h-i} \quad (24)$$

La précision de la prévision peut à présent être mesurée par la variance de $e_t(h)$ conditionnellement à l'information \mathcal{F}_t disponible à l'instant t :

$$\text{Var}(e_t(h)|\mathcal{F}_t) = \sum_{i=0}^{h-1} \gamma_i^2 \mathbb{E}(\epsilon_{t+h-i}^2 | \mathcal{F}_t).$$

Nous pouvons à présent voir la grande différence entre la prévision avec ou sans effet ARCH dans le processus d'innovation : si un effet ARCH est présent, alors $\mathbb{E}(\epsilon_{t+h-i}^2 | \mathcal{F}_t)$ dépend en général de t , donc du point de référence à partir duquel la prévision est effectuée. À l'inverse, dans le cas d'un modèle homoscédastique dans lequel $\mathbb{E}(\epsilon_{t+h-i}^2 | \mathcal{F}_t) = \sigma^2$, la variance de la prévision des erreurs ne dépend pas de l'ensemble d'informations contenues dans \mathcal{F}_t .

Remarque 5.3 Pour calculer l'erreur de prévision (24) en pratique, nous pouvons utiliser une procédure itérative basée sur l'observation suivante : en utilisant la forme ARMA($\max(p, q), q$) d'un processus GARCH(p, q), on peut écrire :

$$\epsilon_{t+h}^2 = \alpha_0 + \sum_{i=1}^{\max(p, q)} (\alpha_i + \beta_i) \epsilon_{t+h-i}^2 - \sum_{j=1}^q \beta_j \eta_{t+h-j} + \eta_{t+h}.$$

L'espérance conditionnelle par rapport à \mathcal{F}_t de cette relation permet d'obtenir $\mathbb{E}(\epsilon_{t+h}^2 | \mathcal{F}_t)$ en fonction de $\mathbb{E}(\epsilon_{t+h-i}^2 | \mathcal{F}_t)$ pour $i > 0$.

Exemple 5.1 À titre d'exemple, considérons le cas du processus suivant

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

où $|\phi| < 1$ et ϵ_t est un processus GARCH(1,1). Les prévisions optimales du processus sont données par la récursion sur h de

$$\hat{X}_t(h) = \phi \hat{X}_t(h-1)$$

où la valeur initiale de la récursion est $\hat{X}_t(1) = \phi X_t$. La représentation MA(∞) du processus AR(1) s'écrit sous la forme (23) avec $\gamma_i = \phi^i$. Dès lors, la variance de l'erreur de prévision s'écrit pour tout $h \geq 1$:

$$\text{Var}(\epsilon_t(h)) = \sum_{i=0}^{h-1} \phi^{2i} \mathbb{E}(\epsilon_{t+h-1}^2 | \mathcal{F}_t)$$

où l'espérance conditionnelle peut être calculée récursivement par la relation

$$\mathbb{E}(\epsilon_{t+h}^2 | \mathcal{F}_t) = \alpha_0 + (\alpha_1 + \beta_1) \mathbb{E}(\epsilon_{t+h-1}^2 | \mathcal{F}_t), \quad h > 1$$

avec comme espérance conditionnelle initiale $\mathbb{E}(\epsilon_{t+1}^2 | \mathcal{F}_t) = \alpha_0 + \alpha_1 \epsilon_t^2 + \beta_1 \sigma_t^2$.

En particulier, si X_t est un GARCH(1,1) avec $\alpha_1 + \beta_1 < 1$, alors cette erreur de prédiction

$$\mathbb{E}(\epsilon_{t+h}^2 | \mathcal{F}_t) = \alpha_0 + (\alpha_1 + \beta_1) \mathbb{E}(\epsilon_{t+h-1}^2 | \mathcal{F}_t) \xrightarrow{h \rightarrow \infty} \frac{\alpha_0}{1 - \alpha_1 - \beta_1}.$$

Cette erreur reste donc asymptotiquement finie.

Par contre, si $\alpha_1 + \beta_1 = 1$, l'erreur de prédiction croît linéairement avec h :

$$\mathbb{E}(\epsilon_{t+h}^2 | \mathcal{F}_t) = \alpha_0 + \mathbb{E}(\epsilon_{t+h-1}^2 | \mathcal{F}_t) = \dots = h\alpha_0 + \epsilon_t^2.$$