

Séries chronologiques

Agnès LAGNOUX

lagnoux@univ-tlse2.fr

ISMAG1 - MIS243Y, 2010-2011

Introduction

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 nous intéressons dans ce cours uniquement à la partie aléatoire de la série temporelle que nous noterons désormais (X_t) au lieu de ϵ_t .

Introduction

Dans ce cours, nous faisons des hypothèses plus générales sur le résidu (X_t) de la série temporelle : nous supposons qu'il sera

- 1 soit stationnaire ;
- 2 soit à accroissements stationnaires ;
- 3 soit non stationnaires.

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 (X_t).

PLAN DU COURS

- 1 Généralités et rappels sur les processus stochastiques
- 2 Les processus linéaires, AR et MA
- 3 Les processus mixtes ARMA
- 4 Les processus non stationnaires ARIMA et SARIMA
- 5 Les processus ARCH et GARCH

Processus stochastiques d'ordre 2

On dit qu'une variable aléatoire est de **carré intégrable** ou d'**ordre 2** si et seulement si $E(X_t^2) < \infty \forall t$.

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 **discrèt**. Si Θ est un intervalle de \mathbb{R} , le processus est dit à temps **continu**.

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 et l'observation à T instants d'un processus stochastique.**

Processus stochastiques d'ordre 2

- ① On appelle **bruit blanc faible** (bb faible) tout processus $\epsilon = (\epsilon_t)_{t \in \mathbb{Z}}$ tel que

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

- ② On appelle **bruit blanc fort** tout bb faible tel que les variables ϵ_t sont indépendantes.
- ③ On appelle **bruit blanc gaussien** tout bruit blanc fort ϵ tel que $\forall t, \epsilon_t \sim \mathcal{N}(0, \sigma^2)$.
- ④ 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.

Espérance conditionnelle et régression affine

L'**espérance conditionnelle** de $Y \in L^2$ sachant X est la variable aléatoire, notée $E(Y|X)$, égale à la projection de Y sur L^2_X où

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

Remarques

- 1 La projection existe car L^2_X est fermé.
- 2 Par définition, on a

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

- 3 Il existe une application r de \mathbb{R} dans \mathbb{R} telle que $E(Y|X) = r(X)$. La fct $r(X)$ est appelée **régression** de Y en X .

Espérance conditionnelle et régression affine

Soit $(X_i)_{i \in I}$ une famille de va de carré intégrable et $L_{((X_i)_{i \in I})}$ le plus petit sous-espace vectoriel engendré par les CL des X_i . On appelle **régression affine** de $Y \in L^2$ sur $(X_i, i \in I)$ la projection orthogonale de Y sur $L_{((X_i)_{i \in I})}$ notée Y^* .

Remarques

- 1 L'orthogonalité et la convergence de L^2 permettent d'introduire cette notion de projection orthogonale.
- 2 Y^* est la meilleure approximation de Y au sens de L^2 par un élément de $L_{((X_i)_{i \in I})}$ puisque

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

- 3 De plus, par définition de $L_{((X_i)_{i \in I})}$, on a $Y = Y^* + R = a_0 + \sum_{j \in J} a_j X_j + R$, avec $R \perp X_i, i \in I$.

Espérance conditionnelle et régression affine

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 :

$$Y^* = E(Y|X)$$

Opérateurs avance et retard

On appelle **opérateur retard** l'opérateur

$$B : (X_t)_{t \in \mathbb{Z}} \mapsto (BX_t)_{t \in \mathbb{Z}} := (X_{t-1})_{t \in \mathbb{Z}}.$$

- ① B est linéaire et inversible d'inverse $B^{-1} = F$ défini par

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

F est appelé **opérateur avance**.

- ② Par récurrence, on définit

$$B^k X_t = X_{t-k}, \quad k \in \mathbb{N} \quad \text{avec} \quad B^0 = I.$$

- ③ Plus généralement, on définit les polynômes en B par

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

- ④ Par extension, pour $n \rightarrow \infty$, on définit $\Psi(B) = \sum_{i=0}^{+\infty} \psi_i B^i$; cet opérateur est défini sous réserve que la série de terme général ψ_i est absolument convergente (i.e. $\sum |\psi_i| < \infty$).

Opérateurs avance et retard

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 :

$$I(X_t) = X_t \quad \forall t.$$

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

Propriétés

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus de carré intégrable.

- 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 .

Opérateurs avance et retard

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

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

- L'opérateur $\Phi(B) := 1 - \phi_1 B - \dots - \phi_p B^p$ est inversible si et seulement si les racines du polynôme Φ sont de module différent de 1.

Processus stationnaires du second ordre

Un processus $X = (X_t)_{t \in \mathbb{Z}}$ est **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 $t \in \mathbf{Z} : \mathbb{E}(X_t^2) < \infty;$

(iii) $\text{Cov}(X_s, X_{s+t}) = \text{Cov}(X_{s-1}, X_{s-1+t}) = \dots = \text{Cov}(X_0, X_t),$

$\forall t, s \in \mathbb{Z}.$

Exemples :

- 1 un bruit blanc est un processus stationnaire.
- 2 la marche aléatoire sur \mathbb{R} n'est pas stationnaire.

Fonctions d'autocovariance et d'autocorrélation

On appelle **fonction d'autocovariance** (resp. **d'autocorrélation**) la fonction γ (resp. ρ) définie de \mathbf{Z} dans \mathbb{R} par

$$\forall h, t \in \mathbf{Z}, \quad \gamma(h) := \text{Cov}(X_t, X_{t+h}) \quad (\text{resp. } \rho(h) := \frac{\gamma(h)}{\gamma(0)}).$$

Son graphe est appelé **variogramme** (resp. **corrélogramme**).

Propriétés La fonction d'autocovariance d'un processus stationnaire vérifie

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

La fonction d'autocorrélation est la version normalisée de la fonction d'autocovariance et vérifie donc le même genre de propriétés.

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 ; plus $|\rho(h)|$ est proche de 1 et plus ce lien est fort.

Exemple : Le processus AR(1) défini par

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

où $|\phi| < 1$ et η_t est un bb de variance σ^2 .

Fonction d'autocorrélation partielle

Soit $(X_t)_{t \in \mathbf{Z}}$ 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

$$\begin{aligned} X_t &= X_{t,h}^* + R_{t,h} = \lambda_{0,h} + \sum_{s=1}^h \lambda_{s,h} X_{t-s} + R_{t,h} \\ &= \lambda_{0,h} + \lambda_{1,h} X_{t-1} + \lambda_{2,h} X_{t-2} + \dots + \lambda_{h,h} X_{t-h} + R_{t,h} \end{aligned}$$

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

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

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

Fonction d'autocorrélation partielle

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

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

et

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

En d'autres termes,

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

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

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

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

Fonction d'autocorrélation partielle : 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 : $\alpha = \text{Cor}(X, Y)$.

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

Lien entre autocorrélation et 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

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

En multipliant par X_{t-1} et X_{t-2} , puis en prenant l'espérance, on obtient

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

On en déduit

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

Lien entre autocorrélation et autocorrélation partielle

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

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

Estimation de la moyenne

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

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

Propriétés

- 1 \bar{X} est un estimateur sans biais de μ : $E(\bar{X}) = \mu$.
- 2 $\text{Var}(\bar{X}) = \frac{1}{T} \sum_{|h| < T} \left(1 - \frac{|h|}{T}\right) \gamma(h)$.
- 3 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

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

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

Estimation de la fonction d'autocovariance

Rappel : 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 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**.

Propriétés

- 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 de la fonction 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**.

Propriétés

- 1 $\hat{\rho}(h)$ est de type positif.
- 2 $\forall h \in \mathbf{Z}, -1 \leq \hat{\rho}(h) \leq 1$.
- 3 Sous certaines conditions, on démontre un TCL 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 $\hat{\rho}(h)$.

Processus AR et MA

Agnès LAGNOUX

lagnoux@univ-tlse2.fr

ISMAG1 - MIS243Y, 2010-2011

PLAN

- 1 Introduction
- 2 Les processus linéaires
- 3 Les processus AR
- 4 Les processus MA

Introduction

Nous étudions maintenant la partie aléatoire d'une série temporelle.

Les processus linéaires constituent le modèle le plus simple pour décrire cette composante : ces processus sont des combinaisons linéaires de b.b. dont l'intérêt tient au comportement de leur fonction d'autocovariance, qui tend vers 0 lorsque l'on compare deux variables éloignées dans le temps :

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

Intuitivement, cela signifie que la "mémoire" du processus est contrôlée : la liaison du second ordre est plus importante pour deux variables proches dans le temps que pour des variables éloignées.

Ce modèle est donc à mémoire appelée "courte" (même si elle est "infinie" pour un processus autorégressif).

Les processus linéaires

$(X_t)_{t \in \mathbf{Z}}$ 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 $(\epsilon_t)_{t \in \mathbf{Z}}$ est un bb fort (resp. faible), de variance σ^2 et où b_i est telle que $\sum_{i=-\infty}^{+\infty} b_i^2 < +\infty$.

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

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

Les processus linéaires : exemples

- ① Soient $X = (X_t)_{t \in \mathbb{Z}}$ un processus stationnaire et $(a_i)_{i \in \mathbb{Z}}$ abstr. sommable. Alors le processus

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

est stationnaire.

- ② Soient $(\epsilon_t)_{t \in \mathbb{Z}}$ un bb et $(a_i)_{i \in \mathbb{Z}}$ abstr. sommable. Alors le processus

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

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

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

et plus généralement

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

Les processus linéaires : définitions

- ① Un processus $(X_t)_{t \in \mathbf{Z}}$ 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 $(\psi_i, i \in \mathbf{Z})$ et un processus $(Y_t)_{t \in \mathbf{Z}}$ tels que

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

- ② Un processus $(X_t)_{t \in \mathbf{Z}}$ 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 $(\psi_i, i \in \mathbf{Z})$ et un processus $(Y_t)_{t \in \mathbf{Z}}$ tels que

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

Les processus linéaires : propriétés

Soit $(X_t, t \in \mathbf{Z})$ le processus stationnaire solution de l'équation suivante :

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

où $(Y_t, t \in \mathbf{Z})$ est un processus stationnaire et avec $p \in \mathbb{N}^*$,

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

Alors

- 1 Si Φ n'a pas de racine de module égal à 1, alors il existe une **représentation inversible** du processus $(X_t, t \in \mathbf{Z})$.
- 2 Si de plus toutes les racines de Φ sont de module supérieur à 1, alors il existe une **représentation causale** du processus $(X_t, t \in \mathbf{Z})$.

Les processus MA(q) : motivation

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.

Les processus MA(q) : définition

On appelle **processus moyenne mobile d'ordre q** [Moving Average] 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 .

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.

Les processus MA(q) : premières propriétés

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.
- la somme d'un MA(q_1) et d'un MA(q_2) (non corrélés) est un MA($\leq \max(q_1, q_2)$).

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

Les processus MA(q) : des exemples

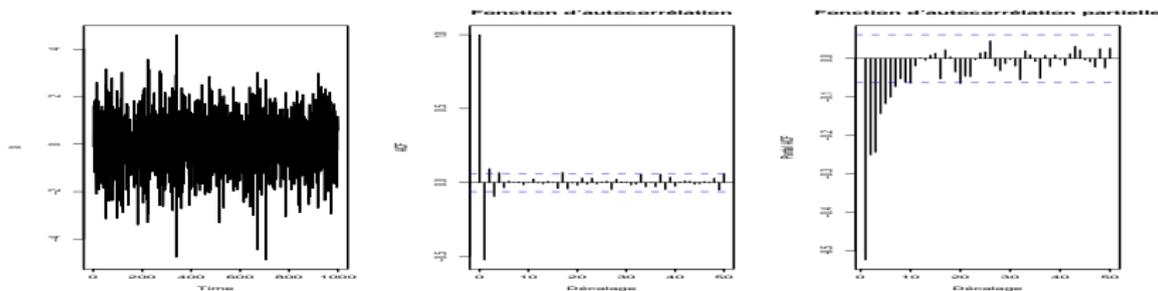


Figure: Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus MA(1) : $X_t = \eta_t + 0.8\eta_{t-1}$

Les processus MA(q) : des exemples

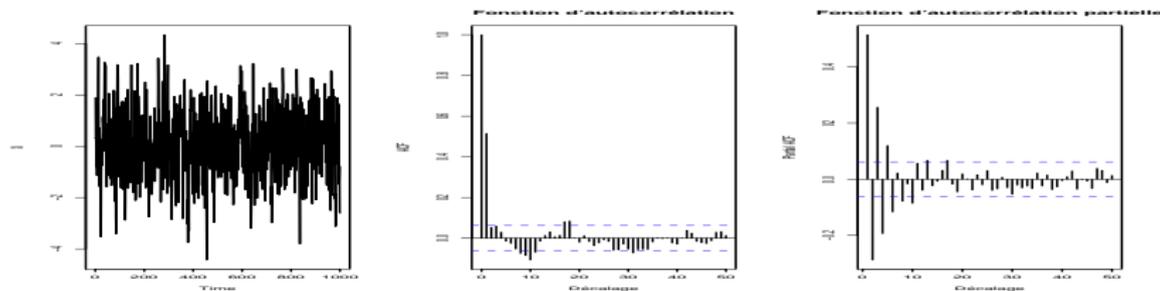


Figure: Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus MA(1) : $X_t = \eta_t - 0.8\eta_{t-1}$

Les processus MA(q) : des exemples

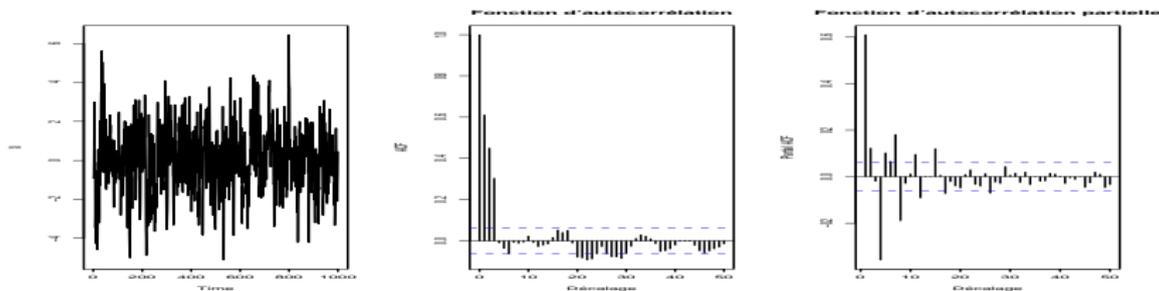


Figure: Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus MA(3) : $X_t = \eta_t - 0.8\eta_{t-1} - 0.4\eta_{t-2} - 0.9\eta_{t-3}$

Les processus MA(q) : forme autorégressive infinie

(i) Si Θ n'a pas de racine de module égal à 1, $\exists! \pi_j$ tq

$$\eta_t = \sum_{j \in \mathbb{Z}} \pi_j X_{t-j}, \quad \text{avec} \quad \sum_{i=-\infty}^{+\infty} |\pi_i| < \infty.$$

Dans ce cas, η est **inversible**.

(ii) Si Θ a des racines de module > 1 , $\exists! \pi_j$ tq

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

Les coefficients π_j convergent rapidement vers 0 lorsque $j \rightarrow \infty$.

Dans ce cas, η est **causal** et c'est l'**innovation** de X en t .

(iii) Si de plus Θ n'a pas de racine multiple, on a

$$\forall j \geq 0, \quad \pi_j = \sum_{1 \leq k \leq q} m_k \mu_k^j, \quad \text{avec} \quad m_k = \frac{1}{\prod_{l \neq k} \left(1 - \frac{\mu_l}{\mu_k}\right)}$$

et μ_k sont les co-racines de Θ .

Les processus MA(q) : bb d'innovation

Le bb d'innovation $(\epsilon_t)_t$ du MA(q) n'est pas nécessairement le processus $(\eta_t)_t$ c-à-d η_t ne représente pas l'information ajoutée au passé pour obtenir la valeur présente de X_t . Mais nous avons le résultat suivant

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 de module > 1 par leur inversion ($\mu \mapsto \frac{1}{\mu}$).

A partir de maintenant, nous ne considérerons que des processus MA donnés sous leur représentation canonique.

Les processus MA(q) : fonctions caractéristiques

- ① La variance de X_t est donnée par

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

- ② La fonction d'autocovariance est donnée par

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

Mais $\mathbb{E}(X_t) = 0$ donc tout MA est un processus stationnaire.

- 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 La fonction d'autocorrélation partielle est telle que $\exists r \in]0, 1[$,

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

Les processus $MA(q)$: identification

Insistons sur le fait que si $h > q$ la fonction d'autocorrélation d'un processus $MA(q)$ s'annule.

En pratique : 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.

Processus stationnaires et $MA(q)$

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 $MA(q)$.

Les processus MA(q) : prévision

Soit X un MA(q) tel que

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

On a donc

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

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

...

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

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

Il suffit ensuite d'exprimer explicitement les $(\eta_u)_{u \leq t}$ en fonction des $(X_u)_{u \leq t}$ grâce à la forme autorégressive infinie.

Les processus MA(q) : prévision

Erreur de prédiction

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

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

Intervalles de confiance

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.

Cependant, les MA(q) ne sont pas bien adaptés à la prévision :

- 1 $j > q \Rightarrow \hat{X}_t(j) = 0$;
- 2 les prévisions nécessitent **toutes** les valeurs passées de X .

Les processus AR(p) : motivation

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).

Les processus AR(p) : définition

On appelle **processus autorégressif d'ordre p** [AutoRegressive of order p] tout processus $(X_t)_{t \in \mathbb{Z}}$ **stationnaire** tel que

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

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 .

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

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

Les processus $AR(p)$: remarques

Contrairement aux processus MA, leur définition pose quelques problèmes :

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

Les processus AR(p) : des exemples

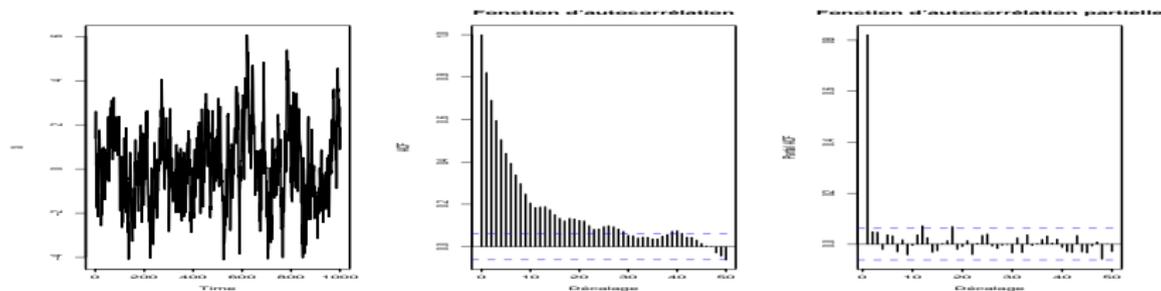


Figure: Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus AR(1) : $X_t - 0.8X_{t-1} = \eta_t$

Les processus AR(p) : des exemples

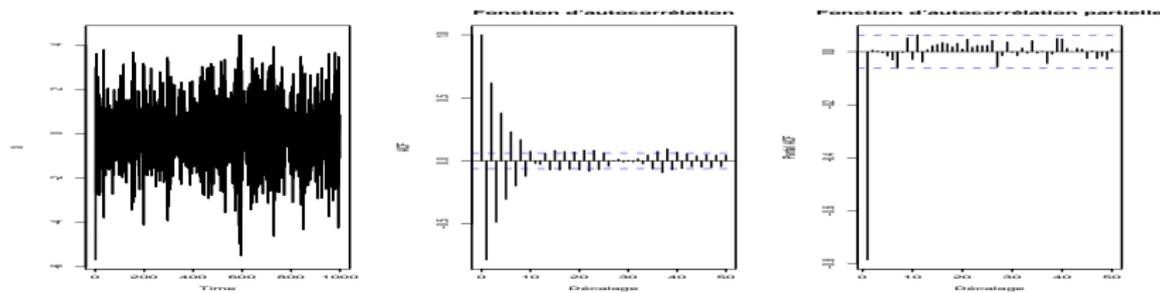


Figure: Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus AR(1) : $X_t + 0.8X_{t-1} = \eta_t$

Les processus AR(p) : des exemples

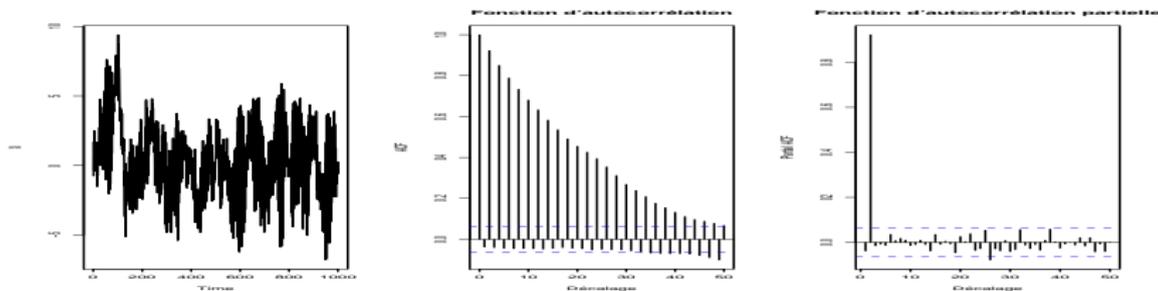


Figure: Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus AR(2) : $X_t - 0.9X_{t-2} = \eta_t$

Les processus AR(p) : des exemples

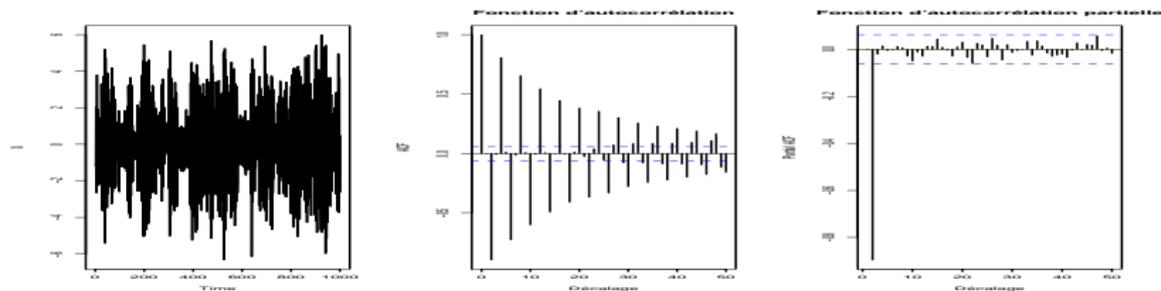


Figure: Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus AR(2) : $X_t + 0.9X_{t-2} = \eta_t$

Les processus AR(p) : des exemples

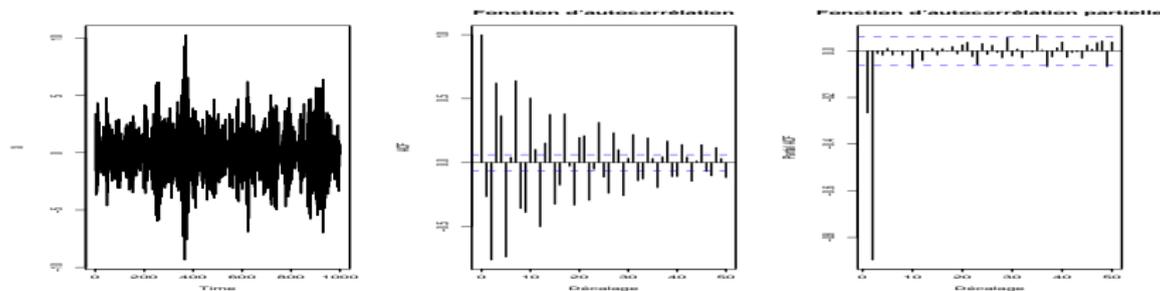


Figure: Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus AR(2) : $X_t + 0.5X_{t-1} + 0.9X_{t-2} = \eta_t$

Les processus AR(p) : un exemple d'AR(2)

Nous allons étudier le processus AR(2) suivant :

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

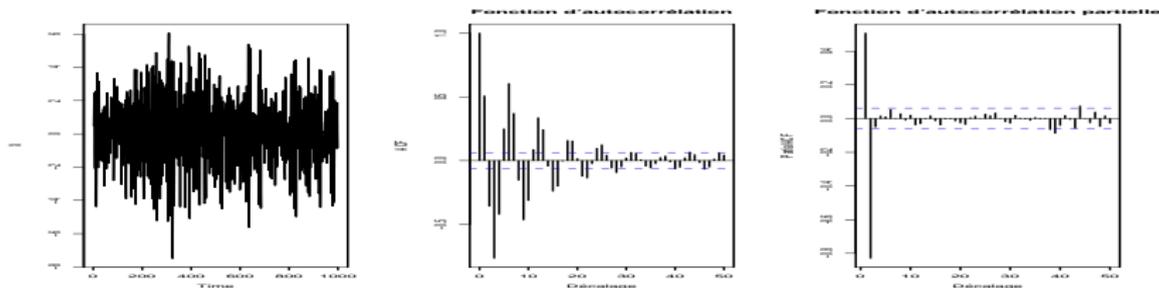


Figure: Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus AR(2) : $X_t - 0.9X_{t-1} + 0.8X_{t-2} = \eta_t$

Les processus $AR(\rho)$: un exemple d' $AR(2)$

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 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.

Les processus AR(p) : forme moyenne mobile infinie

(i) Si Φ n'a pas de racine de module égal à 1, $\exists! \psi_j$ tq

$$\forall t \in \mathbb{Z}, \quad X_t = (\Phi(B))^{-1} \eta_t = \sum_{j \in \mathbb{Z}} \psi_j \eta_{t-j}, \quad \text{avec} \quad \sum_{i=-\infty}^{+\infty} |\psi_i| < \infty.$$

Dans ce cas, X est **inversible**.

Les coefficients ψ_j convergent rapidement vers 0 lorsque $j \rightarrow \infty$.

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 Φ a des racines de module > 1 , $\exists! \psi_j$ tq

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

Dans ce cas, X est **causal** et η est l'**innovation** de X en t .

Les processus $AR(p)$: bb d'innovation

Le bb d'innovation $(\epsilon_t)_t$ d'un $AR(p)$ n'est pas nécessairement le processus $(\eta_t)_t$ c-à-d η_t ne représente pas l'information ajoutée au passé pour obtenir la valeur présente de X_t . Mais nous avons le résultat suivant

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 $MA(q)$.

A partir de maintenant, nous ne considérerons que des processus AR donnés sous leur représentation canonique.

Les processus AR(p) : fonctions caractéristiques

- 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 $|h| = p$.

Les processus AR(p) : identification

A partir de maintenant, nous ne considèrerons que des processus autorégressifs causaux.

Insistons sur le fait que si $h > p$ la fonction d'autocorrélation partielle d'un processus AR(p) s'annule.

En pratique : si à partir de données X_1, \dots, X_T , la fonction d'autocorrélation partielle empirique n'est pas significativement différente de zéro au-delà d'un certain nombre p_0 , on sera alors guidé pour choisir d'ajuster un modèle AR(p_0) aux données.

Les processus AR(p) : prévision

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

A l'horizon 1 : on a

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

que l'on prédit par

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

A l'horizon 2 : on a

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

que l'on prédit par

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

et ainsi de suite.

Les processus AR(p) : prévision

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

$$\hat{X}_t(h) = \sum_{1 \leq k \leq p} \phi_k \hat{X}_t(h-k) \quad \text{avec} \quad \hat{X}_t(-j) = X_{t-j} \quad \text{pour} \quad j \geq 0$$

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 $\hat{X}_t(h)$ par récurrence comme ci-dessus.

Les processus AR(p) : prévision

Erreur de prédiction un pas en avant

Dans le cas des processus autorégressifs causaux, le bruit blanc coïncide avec l'erreur de prédiction un pas en avant :

$$X_{t+1} - \hat{X}_t(1) = \eta_{t+1}.$$

Intervalles de confiance

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.

Ainsi les processus autorégressifs sont bien adaptés à la prédiction.

Les processus AR(p) : prévision

Pour un AR(p), les prévisions $\hat{X}_t(h)$ 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 $\gamma^\eta(0)$.

Les processus AR(p) : équations de Yule-Walker

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

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 $YW_j, j > p$.

Pour des raisons pratiques, on veut pouvoir calculer ϕ_k et $\gamma^n(0)$ lorsque les autocovariances de X sont connues.

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 $\gamma^n(0)$.

Les processus AR(ρ) : équations de Yule-Walker sur un exemple

Soit le processus AR(2) de la forme :

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

que l'on suppose **centré et causal**. En multipliant l'équation (1) 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 :

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

Les processus $AR(p)$: équations de Yule-Walker sur un exemple

Quant à la variance, partant de l'équation (1), 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 des fct d'autocovariance et d'autocorrélation, on déduit des estimateurs convergents de ϕ_1 , ϕ_2 et σ^2 .

Les processus AR(p) : équations de Yule-Walker

Plus généralement, on a les équations suivantes :

$$(YW_0) \quad \gamma^X(0) - \phi_1\gamma^X(-1) - \dots - \phi_p\gamma^X(-p) = \gamma^\eta(0)$$

$$(YW_1) \quad \gamma^X(1) - \phi_1\gamma^X(0) - \dots - \phi_p\gamma^X(1-p) = 0$$

...

$$(YW_j) \quad \gamma^X(j) - \phi_1\gamma^X(j-1) - \dots - \phi_p\gamma^X(j-p) = 0$$

Les équations de Yule-Walker peuvent ensuite être utilisées pour l'estimation des paramètres d'un processus autorégressif.

Les processus AR(p) : équations de Yule-Walker

Dans la pratique,

- 1 on commence par deviner une valeur raisonnable pour p ;
- 2 on calcule les valeurs empiriques des autocovariances $\gamma^X(0), \dots, \gamma^X(p)$;
- 3 on en déduit les valeurs de ϕ_1, \dots, ϕ_p et $\gamma^n(0)$;
- 4 on cherche ensuite à valider le modèle en comparant avec les empiriques les valeurs des $\gamma^X(j)$ pour $j > p$ déduites des autres YW $_j$.
- 5 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.

Processus mixtes ARMA

Agnès LAGNOUX

lagnoux@univ-tlse2.fr

ISMAG1 - MIS243Y, 2010-2011

Introduction

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 AR, la fct d'autocorrélation partielle possède un point de rupture après un certain nombre d'écarts ; ce dernier détermine l'ordre du polynôme AR.
- Pour un MA, la fct d'autocorrélation possède un point de rupture après un certain nombre d'écarts ; 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.

Introduction

Les processus ARMA sont des processus linéaires et jouent 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.

Les processus ARMA(p, q) : définition

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

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

où $\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q \in \mathbb{R}$ et $(\eta_t)_{t \in \mathbb{Z}}$ bb de var σ^2 .

Si $\Phi(B) = I - \phi_1 B - \dots - \phi_p B^p$ et $\Theta(B) = I - \theta_1 B - \dots - \theta_q B^q$, on a

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

- Cons :**
- a) un ARMA($p, 0$) est un AR pur ;
 - b) un ARMA($0, q$) est un MA pur ;
 - c) les processus ayant simultanément des représentations MA pure et AR pure sont les ARMA($0, 0$) i.e. des bb.

Les processus ARMA(p, q) : premières propriétés

La somme d'1 ARMA(p_1, q_1) et d'1 ARMA(p_2, q_2) (non corrélés) est 1 ARMA($\leq \max(p_1, p_2), \leq \max(q_1, q_2)$).

Ainsi

- la somme de 2 bb indépendants (=ARMA(0,0)) est 1 bb.
- la somme d'1 bb et d'1 MA(q) est 1 MA(q).
- la somme d'1 bb et d'1 AR(p) est 1 ARMA($p, \leq p$).
- la somme d'1 AR(p_1) et d'1 AR(p_2) est 1 ARMA($p_1 + p_2, \leq p_1 + p_2$).
- la somme d'1 AR(p) et d'1 MA(q) est 1 ARMA($p, \leq p + q$).
- la somme d'1 MA(q_1) et d'1 MA(q_2) est 1 MA($\leq q_1 + q_2$).

Les processus ARMA(p, q) : un exemple

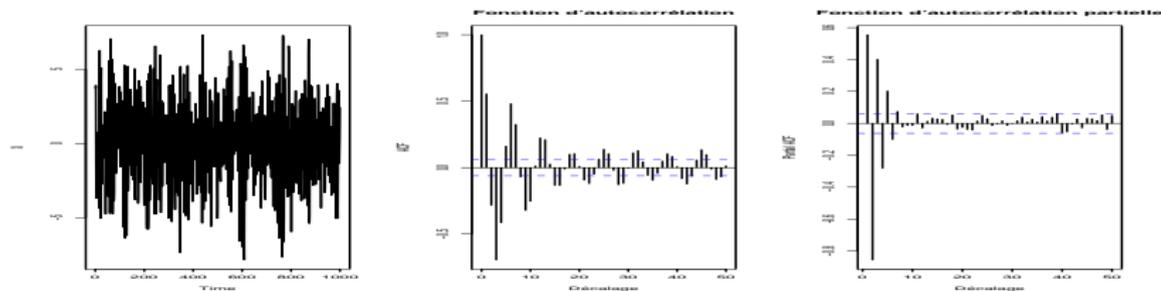


Figure: Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus ARMA(2,1) : $X_t - 0.9X_{t-1} + 0.8X_{t-2} = \eta_t - 0.7\eta_{t-1}$

Les processus ARMA(p, q) : forme moyenne mobile infinie

(i) Si les racines de Φ sont de module $\neq 1$, $\Phi(B)$ est **inversible** et X est donnée par

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

(ii) Si les racines de Φ sont de module > 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 $i > p$.

Les processus ARMA(p, q) : forme moyenne mobile infinie

Ces équations peuvent se résoudre successivement pour ψ_0, ψ_1, \dots

Ainsi

$$\psi_0 = 1$$

$$\psi_1 = -\theta_1 + \psi_0\phi_1 = -\theta_1 + \phi_1$$

$$\psi_2 = -\theta_2 + \psi_0\phi_2 + \psi_1\phi_1 = -\theta_2 + \phi_2 - \theta_1\phi_1 + \theta_1^2 \dots$$

Exemple : On considère X défini par

$$X_t - X_{t-1} + \frac{1}{4}X_{t-2} = \eta_t + \eta_{t+1}, \quad \text{avec}(\eta_t) \text{ bb de } \text{var}\sigma^2.$$

On a donc $\phi_1 = 1, \phi_2 = -\frac{1}{4}, \theta_1 = -1, p = 2$ 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

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

Les processus ARMA(p, q) : forme autorégressive infinie

(i) Si les racines de Θ sont de module $\neq 1$, $\Theta(B)$ est **inversible** et η est donné par

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

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

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

Si les racines de Φ et Θ sont de module > 1 , les 2 représentations sont vérifiées et η_t est l'**innovation** de X_t à la date t .

Les processus ARMA(p, q) : bb d'innovation

Le bb d'innovation $(\epsilon_t)_t$ d'un ARMA(p, q) n'est pas nécessairement le processus $(\eta_t)_t$ c-à-d η_t ne représente pas l'information ajoutée au passé pour obtenir la valeur présente de X_t . Mais nous avons le résultat suivant

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 que des processus ARMA donnés sous leur représentation canonique.

Les processus ARMA(p, q) : fonctions caractéristiques

La fonction d'autocovariance d'un processus ARMA(p, q) vérifie

pour $h \geq \max(p, q + 1)$

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

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

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

La suite des covariances satisfait donc 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 $\rho(h) \forall h$.

Les processus ARMA(p, q) : fonctions caractéristiques

En pratique

- On trouve tout d'abord $\gamma(0), \dots, \gamma(p)$ à partir de ces équations avec $h = 0, 1, \dots, p$.
- 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 (suite) : X défini par $X_t - X_{t-1} + \frac{1}{4}X_{t-2} = \eta_t + \eta_{t+1}$.

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 $\gamma(h) = \gamma(h-1) - \frac{1}{4}\gamma(h-2)$ pour $h \geq 3$.

Les processus ARMA(p, q) : prévision

Supposons que l'on dispose d'un grand nombre d'observations consécutives de X jusqu'à l'instant T . Une fois le modèle choisi et ses paramètres estimés, il va être possible de faire de la prévision.

Pour estimer les prévisions optimales $\hat{X}_T(h)$, il faut disposer

- des coefficients ψ_j qui lient X à son bruit blanc d'innovation 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 $\eta_u, u \leq T$.

Les processus ARMA(p, q) : prévision à partir de AR(∞)

(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 , $\sum_{j \geq 0} \alpha_j(h) z^j = \left[\frac{1}{z^h} \frac{\Theta(z)}{\Phi(z)} \right]_1 \frac{\Phi(z)}{\Theta(z)}$; où $[]_1$ signifie que l'on ne retient que les termes en z^k pour $k \geq 0$.

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

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

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

I.e. on procède comme pour 1 AR(p) dès qu'on a la représentation "AR(∞)" qui lie le bb à X .

Les processus ARMA(p, q) : prévision à partir de AR(∞)

Pour la prévision à l'horizon 1 : $\widehat{X}_T(1) = -\sum_{j=1}^{\infty} \pi_j X_{T+1-j}$.

Pour la prévision à l'horizon 2 :

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

Et ainsi de suite pour tout $h \in \mathbb{N}^*$.

Cependant, ces prévisions font intervenir des **valeurs non observées**, à savoir X_t pour $t \leq 0$. On approxime en tronquant la série :

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

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

Les processus ARMA(p, q) : prévision à partir de MA(∞)

(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

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

Les erreurs de prédiction $X_{T+h} - \widehat{X}_T(h)$ ne sont donc pas non-corrélées.

Les processus ARMA(p, q) : prévision

Erreur de prédiction

- 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.

Intervalles de confiance

Si on suppose que le bruit d'innovation est gaussien, les variables aléatoires X_t sont aussi gaussiennes, tout comme l'erreur de prédiction. Ainsi, dans le cas d'un bb fort, 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 .

Les processus ARMA(p, q) : mise à jour

Le problème est le suivant :

- 1 on a prévu $(X_{T+k})_{\{1 \leq k \leq h\}}$ à partir de $(X_{T-i})_{\{i \geq 0\}}$;
- 2 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} en écrivant le développement en MM

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

Le prédicteur est donné par

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

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 nouvelle donnée.

Les processus ARMA(p, q) : approche Box-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.

Les processus ARMA(p, q) : approche Box-Jenkins

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 la démarche Box-Jenkins pour l'ajustement de données à l'aide d'un modèle ARMA sont les suivantes :

- 1 Estimation de la tendance
- 2 Estimation des fonctions d'autocorrélation et d'autocorrélation partielle
- 3 Vérifications des hypothèses
- 4 Détermination de p_{\max} et q_{\max}
- 5 Estimation des paramètres
- 6 Amélioration de l'estimation des paramètres
- 7 Sélection de modèles

Les processus ARMA(p, q) : approche Box-Jenkins

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

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

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

2) Estimation des fonctions d'autocorrélation et d'autocorrélation partielle : on estime la fonction d'autocorrélation ρ par $\hat{\rho}_T$ et d'autocorrélation partielle τ par $\hat{\tau}_T$.

Il est également intéressant d'étudier les autocorrélations inverses : si $(X_t)_{t \in \mathbf{Z}}$ est un processus ARMA(p, q) 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**.

Les processus ARMA(p, q) : approche Box-Jenkins

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

4) **Détermination de p_{\max} et q_{\max}** : si l'étape précédente est validée, on cherche 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 $0 \leq q \leq q_{\max}$.

Les processus ARMA(p, q) : approche Box-Jenkins

- 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 bb Gaussien (grâce à un test), 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 grâce à des critères tels que AIC (Akaike Information Criterion), le SBC/BIC (Schwarz Bayesian Criterion ou Bayesian Information Criterion) et le critère de Hannan.

Les processus ARMA(p, q) : approche Box-Jenkins

Résumé de la méthodologie de Box et Jenkins sous la forme d'un graphique :

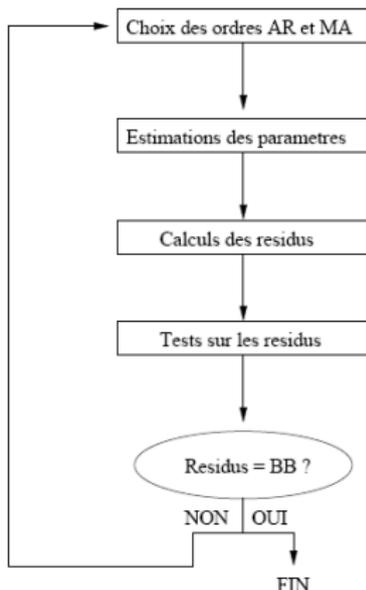


Figure: Méthodologie de Box et Jenkins

Les processus AR/MA/ARMA : bilan

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

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

Processus ARIMA et SARIMA

Agnès LAGNOUX

lagnoux@univ-tlse2.fr

ISMAG1 - MIS243Y, 2010-2011

PLAN

- 1 Introduction
- 2 Les processus ARIMA
- 3 Les processus SARIMA
- 4 Identification des modèles

Introduction

- Les séries précédemment étudiées étaient supposées stationnaires.
- Néanmoins 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 satisferait une représentation ARMA.

Dans ce chapitre, nous allons construire en toute généralité des prédicteurs pour des processus non stationnaires car comportant une tendance (ARIMA) ou une saisonnalité (SARIMA).

Un exemple introductif

Soit $(X_t)_{1 \leq t \leq T}$ avec saisonnalité de période 12. Afin de supprimer cette saisonnalité, 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 $(Y_t)_t$ puis on revient à la série $(X_t)_t$

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

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} .

Les processus ARIMA : définition

On intéresse aux processus (X_t) satisfaisant

$$\phi(B)\Delta^d X_t = \Theta(B)\eta_t, \quad (1)$$

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 (1) s'écrit aussi

$$\Phi(B)X_t = \Theta(B)\eta_t, \quad (2)$$

avec $\Phi(B) = \phi(B)(I - B)^d$. La relation (2) 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. Pour compléter la définition, il faut introduire une initialisation.

Les processus ARIMA : deux exemples

- 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 $t = 0$. Si une valeur initiale X_{-1} est donnée alors

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

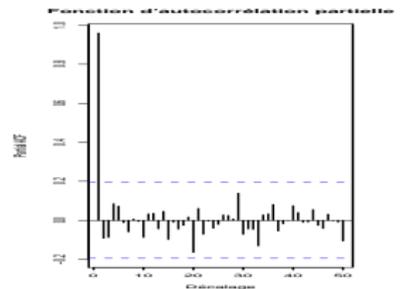
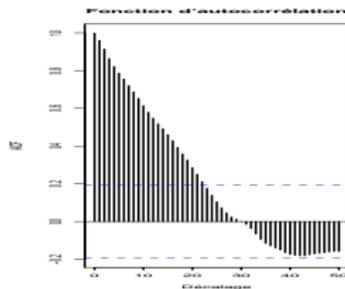
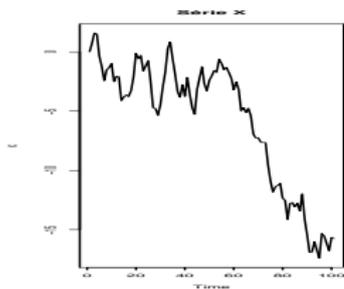
Il est naturel d'autre part de supposer que X_{-1} est non corrélée avec les valeurs futures du bruit. Le processus (X_t) est alors **défini sans ambiguïté**.

- 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} .

Un exemple de série ayant ce type de comportement : graphe de trajectoire, corrélogramme et corrélogramme partiel du processus ARIMA(1, 1, 1) de coefficient AR=0.1 et MA=0.3



Les processus ARIMA : forme moyenne mobile

Soit $(X_t)_{t \geq 0}$ un processus ARIMA(p, d, q) et $Z_{-1} = (X_{-1}, \dots, X_{-p-d}, \eta_{-1}, \dots, \eta_{-q})'$ le vecteur de ses valeurs initiales.

X_t peut s'écrire sous la forme moyenne mobile suivante :

$$X_t = \sum_{j=0}^t h_j \eta_{t-j} + h^*(t) Z_{-1}, \quad (3)$$

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 .

Les processus ARIMA : forme autorégressive

X_t peut s'écrire sous la forme autorégressive suivante :

$$\eta_t = X_t + \sum_{j=1}^t \pi_j X_{t-j} - g(t)Z_{-1}, \quad (4)$$

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.

Les processus ARIMA : prévision

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 **prévision 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,

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

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

Exemple : prédiction un pas en avant.

Les processus ARIMA : prévision et forme $MA(\infty)$

La représentation moyenne mobile 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

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

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

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

Formule de mise à jour

Les processus ARIMA : prévision et forme AR(∞)

La formule autorégressive 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

$$\hat{X}_t(k) = - \sum_{j=1}^{t+k} \pi_j \hat{X}_{t+k-j} \quad \text{avec} \quad \hat{X}_{t+k-j} = \begin{cases} \hat{X}_t(k-j) & \text{si } k > j \\ X_{t+k-j} & \text{si } k \leq j \end{cases}$$

Les processus ARIMA : prévision et 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

$$\hat{X}_t(k) = - \sum_{j=1}^{p+d} \Phi_j \hat{X}_{t+k-j} + \sum_{j=1}^q \theta_j \hat{\eta}_{t+k-j}, \quad (5)$$

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

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

Les processus ARIMA : utilisation jointe de ces formules

On veut des prédictions jusqu'à l'horizon K à partir de T , on doit donc calculer $X_T(k)$ pour $k = 1, \dots, K$.

À la date $T + 1$, on doit (a) modifier les prévisions de $X_{T+1} \dots X_{T+K}$ i.e. pour calculer $\hat{X}_{T+1}(k)$, $k = 1, \dots, K - 1$.

(b) déterminer une nouvelle prévision, celle de X_{T+K+1} .

- 1 On calcule les $\hat{X}_T(k)$, $k = 1, \dots, K$ à la première date T avec la prédiction AR; ce calcul fait, on n'aura plus à se servir des valeurs des X_s , $s \leq T$.
- 2 À la date $T + 1$, on connaît la valeur de X_{T+1} ; on peut donc calculer $\hat{X}_{T+1}(k)$, $k = 1, \dots, K - 1$ avec la prédiction MM.
- 3 On calcule $\hat{X}_{T+1}(K)$ à partir de la prédiction ARIMA à condition d'avoir pris $K > q$ (et donc $\hat{\eta}_{t+K-i} = 0$, $i = 1, \dots, q$) et $K > p + d$.

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.

Les processus SARIMA : motivation

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

Les processus SARIMA : motivation

Le modèle complet s'écrit :

$$\text{Annuel} \quad X_t = \sum_{k=1}^P f_k X_{t-12k} + V_t + \sum_{j=1}^Q g_j V_{t-12j}$$

$$\text{Mensuel} \quad \Phi(B)V_t = \Theta(B)\eta_t \quad \text{où } (\eta_t) \text{ est un bruit blanc.}$$

$$\text{Annuel} \quad F(B^{12})X_t = G(B^{12})V_t$$

$$\text{Mensuel} \quad \Phi(B)V_t = \Theta(B)\eta_t$$

La combinaison des séries annuelle et mensuelle livre un processus $\text{ARMA}(p, q) \times (P, Q)_{12}$. Si on y inclut également les tendances ($d > 0$) et les composantes déterministes périodiques ($D > 0$), on a le modèle $\text{SARIMA}(p, d, q) \times (P, D, Q)_s$.

Les processus SARIMA : définition

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

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

est un processus ARMA stationnaire de la forme

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

où Φ (resp. Θ) est le pol. générateur d'un AR(p) (resp. 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 pol. générateur d'un AR(P) (resp. MA(Q)) :

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

Les processus SARIMA : deux exemples

- ① $X_t = \mu + S_t + \eta_t$, 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).

- ② $s = 12$ (données mensuelles), $p = 1 = P$, $q = 0 = Q$,
 $d = D = 1$: soit l'opérateur

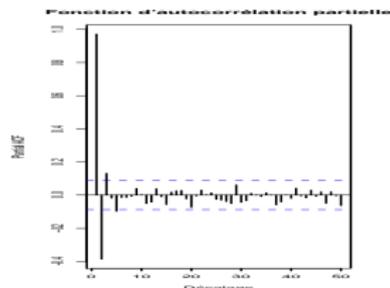
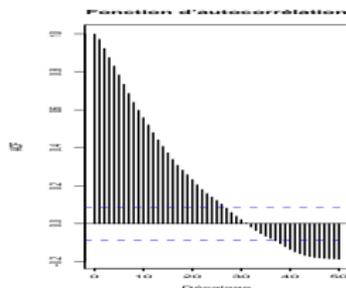
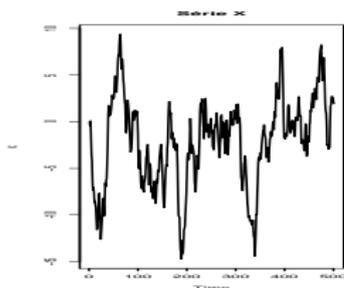
$$\Phi(z)F(z^s) := (1 - \phi z)(1 - fz^{12}) = (1 - \phi z - \phi fz^{12} + \phi fz^{13}).$$

Alors

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

est un AR(13)=AR($p + Ps$) ou encore $Y_{t-1} - Y_{t-12}$ est un AR(1).

Un exemple de série ayant ce type de comportement : Graphe de trajectoire, corrélogramme et corrélogramme partiel du processus SARIMA(1, 1, 1) × (0, 0, 0)₁₂



Les processus SARIMA : prévision

Si les données X_1, \dots, X_T suivent un modèle 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, $(d, D, s) = (1, 1, 12)$,

$$Y_t = (X_t - X_{t-12}) - (X_{t-1} - X_{t-13}) \quad (6)$$

On traite le problème de prédiction du processus Y_t du type ARMA($p + Ps, q + Qs$) comme vu précédemment 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 $t \leq T$.

Exemples

Identification et ajustement des modèles ARIMA/SARIMA

(0) Appliquer (si nécessaire) une transformation de type "Box-Cox" pour stabiliser la variance des données. ex : $\log(X_t)$, $\exp X_t$.

(1) Choisir d , D et s 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).

(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 de vraisemblance les paramètres.

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

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

Processus ARCH et GARCH

Agnès LAGNOUX

lagnoux@univ-tlse2.fr

ISMAG1 - MIS243Y, 2010-2011

PLAN DU COURS

- 1 Introduction
- 2 Les processus ARCH
- 3 Les processus GARCH
- 4 Estimation des paramètres
- 5 Prédiction

Introduction

- Les séries précédemment étudiées étaient supposées stationnaires ou du moins **homoscédastique** (variance constante).
- Si besoin, tendances et saisonnalités étaient supprimées pour obtenir une série résiduelle stationnaire.
- Néanmoins, toutes les séries résiduelles obtenues de la sorte ne sont pas nécessairement stationnaires.
- C'est le cas par exemple de la série des évolutions journalières de la bourse des valeurs de New-York (NYSE) du 19 octobre 1984 au 31 décembre 1991.

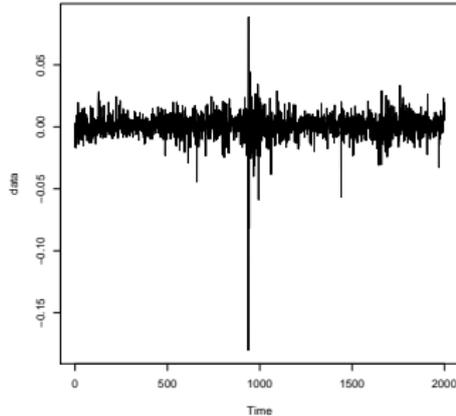
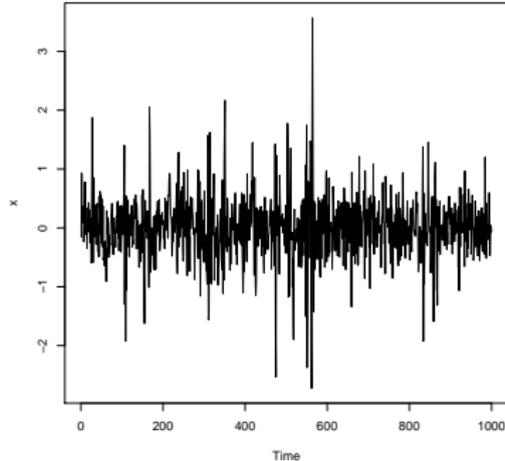


Figure: Evolution journalière de la bourse des valeurs de NY (1984-91)

- la moyenne semble constante alors que la variance change au cours du temps (on qualifie ce comportement d'hétéroscédastique) ;
- les moments de grande variabilité semblent regroupés.

Un autre exemple de série ayant ce type de comportement :



Ainsi les modèles de type ARIMA qui supposent un comportement **homoscédastique** (variance constante), ne sont pas adaptés à ce type de série.

Dans la théorie des processus ARMA, la variance d'une série est (entre autres) déterminée par la variance du processus des innovations.

Nous souhaitons à présent tenir compte d'un éventuel changement au cours du temps de la variance, provenant de l'évolution passée du processus.

C'est ce que permettent les modèles ARCH, dont l'idée est de déterminer la distribution de ϵ_t conditionnellement à toutes les valeurs passées X_{t-1}, X_{t-2}, \dots .

Nous présentons dans la suite de ce chapitre des modèles adaptés à ce type de série : les processus ARCH introduits par Engle vers 1982, ainsi que leur généralisation, les processus GARCH.

Les processus ARCH : définition

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

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

Les processus ARCH : interprétation

- La variance conditionnelle σ_t^2 est linéairement reliée au carré des p dernières observations et modélise la variation de ces observations autour de leur moyenne zéro.
- 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.
⇒ modèle ARCH permet d'expliquer ce regroupement en clusters de la volatilité.
- L'ordre p du modèle détermine la durée de persistance du choc.

Les processus ARCH : propriétés

Si X_t est un processus ARCH(p), alors

$$\mathbb{E}(X_t) = 0$$

$$\text{Var}(X_t) = \frac{\alpha_0}{1 - \sum_{i=1}^p \alpha_i} \quad \text{si} \quad \sum_{i=1}^p \alpha_i < 1$$

$$\text{Cov}(X_t, X_{t+h}) = 0 \quad \forall h > 0$$

$$\mathbb{E}(X_t | \mathcal{F}_{t-1}) = 0$$

$$\text{Var}(X_t | \mathcal{F}_{t-1}) = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_p X_{t-p}^2$$

$$\text{Cov}(X_t, X_{t+h} | \mathcal{F}_{t-1}) = 0 \quad \forall h > 0$$

Les processus ARCH : propriétés

(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.

Les processus GARCH : définition

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 $j = 1, \dots, q$.

Les processus GARCH : interprétation

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.

Un $\text{GARCH}(p, 0)$ est un $\text{ARCH}(p)$.

Les processus GARCH : propriétés

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

$$\mathbb{E}(X_t) = 0$$

$$\mathbb{E}(X_t | \mathcal{F}_{t-1}) = 0$$

$$\text{Cov}(X_t, X_{t+h}) = 0 \quad \forall h > 0$$

$$\text{Cov}(X_t, X_{t+h} | \mathcal{F}_{t-1}) = 0$$

Les processus GARCH : identification

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?).

Estimation des paramètres pour un ARCH

- 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 param. $\alpha = (\alpha_0, \dots, \alpha_p)'$ dans 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 par maximum de vraisemblance

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

Estimation des paramètres pour un GARCH

- 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 $i \leq 0$.

Estimation des paramètres pour un GARCH

- On remplace donc la fonction de vraisemblance par

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

- La fonction de vraisemblance totale est

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

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

Prédiction

Puisque $\mathbb{E}(\epsilon_{t+h}|\mathcal{F}_t) = 0$, le prédicteur optimal $\hat{X}_t(h)$ est donné par

$$\hat{X}_t(h) = \mathbb{E}(X_{t+h}|\mathcal{F}_t) = 0,$$

Les modèles GARCH n'ont donc aucun intérêt en tant que tels pour la prédiction mais plutôt pour construire des modèles plus complexes.

Intérêt des modèles GARCH : les modèles ARMA-GARCH

Weiss (1984) eut l'idée d'appliquer la théorie des modèles GARCH

- 1 aux erreurs d'une régression linéaire

$$Y_t = aX_t + \epsilon_t$$

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

- 2 à un processus d'innovation d'un modèle ARMA

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

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

Modèles ARMA-GARCH et prédiction

Partons du processus ARMA(r, s)

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

où $\Phi(B) = 1 - \phi_1 B - \dots - \phi_r B^r$ et $\Theta(B) = 1 - \theta_1 B - \dots - \phi_s B^s$ et ϵ_t est un processus GARCH(p, q).

Supposons que l'on observe cette série jusqu'au temps t .

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 $\mathbb{E}(\epsilon_{t+h} | \mathcal{F}_t) = 0$.

Modèles ARMA-GARCH et prédiction

La prévision est formellement identique à celle développée dans les chapitres consacrés aux processus linéaires. La présence d'un modèle (G)ARCH au niveau du processus d'innovations ne modifie pas la procédure de prévision.

Par contre, l'erreur de prévision est donnée par

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

et la précision de la prévision est mesurée par la variance

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

qui dépend de t point de référence à partir duquel la prévision est effectuée.