

TP 6 : ARMA

DESS d'Ingénierie Mathématique, Année 2003-2004

Séries chronologiques

laurent.albera@fr.thalesgroup.com

I. INTRODUCTION

Ce TP s'effectuera sous MATLAB. La première étape dans l'étude d'une série chronologique consiste à détecter la présence de facteurs *tendance* et *saisonnier*, et le cas échéant à les estimer pour pouvoir les retrancher aux données. On se trouve alors en possession de ce que l'on nomme *fluctuations résiduelles*, en d'autres mots, la partie aléatoire de la série chronologique. Or, bien que non déterministe, cette composante peut avoir une structure particulière qu'il est alors important d'identifier. Une solution consiste à modéliser cette partie aléatoire par un processus ARMA dont il faudra estimer les différents paramètres. En effet, une des propriétés vues en cours (*Traitement des processus ARMA*, page 2) nous dit que *pour tout processus aléatoire dont la fonction d'autocorrélation tend vers 0 à l'infini, et pour tout entier $k > 0$, il est possible de trouver un processus ARMA dont la fonction d'autocorrélation est égale à celle du processus aléatoire jusqu'à l'ordre k* . Ainsi, une fois le modèle ARMA identifié, nous pouvons estimer avoir expliqué au mieux la série chronologique considérée, car alors seul le bruit blanc en entrée de l'ARMA reste inconnu. La phase explicative achevée, les données de la série chronologique pourront alors être utilisées pour, par exemple, prédire soit des données futures, soit des données passées. Ce TP consistera donc à mettre en oeuvre des techniques d'identification d'un processus ARMA.

II. HYPOTHESES ET NOTATIONS

Rappelons l'expression d'un processus ARMA(p, q) noté $\{X_i\}$ supposé centré (i.e. de moyenne nulle):

$$X_i - \phi_1 X_{i-1} - \dots - \phi_p X_{i-p} = Z_i + \theta_1 Z_{i-1} + \dots + \theta_q Z_{i-q} \quad (1)$$

où $\{Z_i\}$ est un processus aléatoire blanc centré de variance σ^2 qui sera supposé gaussien dans la suite du TP, où $\{\phi_k\}$, $\{\theta_k\}$, p et q sont les paramètres du processus ARMA. Par ailleurs, un processus ARMA est dit *causal* s'il peut être écrit sous la forme d'un MA(∞): les coefficients de ce processus MA seront notés ψ_k par la suite. Une autre expression équivalente à (1) est donnée à l'aide de l'opérateur retard B :

$$\phi(B)X_i = \theta(B)Z_i \quad (2)$$

où les polynômes ϕ et θ , de degrés respectifs p et q , ont pour coefficients les paramètres $\{\phi_i\}$ et $\{\theta_i\}$ cités dans (1) avec $\phi_0 = \theta_0 = 1$.

III. DONNÉES APPLICATIVES

Afin de tester et comparer les méthodes d'identification de processus ARMA programmées lors de ce TP, nous allons synthétiser cinq processus ARMA distincts, définis par :

$$\begin{aligned} \text{ARMA}(3,2) &: \{X_i^1, \phi(B) = -(B+3)(B+2)(B-7)/42, \theta(B) = (B-5)(B-3)/15, \sigma^2 = 1\} \\ \text{AR}(4) &: \{X_i^2, \phi(B) = (B^2-36)(B+5)(B-4)/720, \theta(B) = 1, \sigma^2 = 1\} \\ \text{AR}(1) &: \{X_i^3, \phi(B) = 1 - \frac{9}{10}B, \theta(B) = 1, \sigma^2 = 1\} \\ \text{MA}(1) &: \{X_i^4, \phi(B) = 1, \theta(B) = 1 - \frac{3}{4}B, \sigma^2 = 1\} \\ \text{MA}(2) &: \{X_i^5, \phi(B) = 1, \theta(B) = (B+3)(B+2)/6, \sigma^2 = 1\} \end{aligned} \quad (3)$$

Quelles propriétés vérifient ces cinq processus? Puis, sous MATLAB, en utilisant la fonction *filter*, générer les processus $\{X_i^j\}$ à partir de $n=1000$ échantillons ($\{Z_i^j\}$ sera supposé gaussien).

IV. FONCTIONS DE COVARIANCE ET DE CORRÉLATION

A. comparaison des fonctions exacte et empirique

En utilisant les propositions 1 et 2, comparer graphiquement les fonctions d'autocorrélation exacte et empirique associées tout d'abord au processus $\{X_i^3\}$, puis au processus $\{X_i^4\}$.

Proposition 1: Soit $\{X_i\}$ un processus AR(1) stationnaire tel que $|\phi_1| < 1$, alors $\{X_i\}$ est *causal* et les coefficients ψ_k sont donnés par $\psi_k = [\phi_1]^k$.

Proposition 2: La fonction de covariance exacte d'un processus ARMA causal nous est donnée par la formule $\gamma(h) = \sigma^2 \sum_{k=0}^{+\infty} \psi_k \psi_{k+|h|}$.

B. estimation de l'ordre d'un MA

Tout d'abord rappelons les deux propositions énoncées en cours (*Traitement des processus ARMA*, pages 11-14) visant à caractériser la fonction d'autocorrélation empirique d'un bruit blanc ainsi que celle d'un ARMA(p, q):

Proposition 3: La fonction d'autocorrélation empirique $\rho_e(h)$ d'un bruit blanc, pour $h > 0$, est comprise entre $-1.96n^{-1/2}$ et $1.96n^{-1/2}$ avec une probabilité de 95 pour cent.

Proposition 4: La fonction d'autocorrélation empirique $\rho_e(h)$ d'un ARMA(p, q), pour $h > 0$, est comprise entre $-1.96n^{-1/2}(1 + 2 \sum_{h=1}^{p+q} \rho_e(h)^2)$ et $1.96n^{-1/2}(1 + 2 \sum_{h=1}^{p+q} \rho_e(h)^2)$ avec une probabilité de 95 pour cent.

Sur une première figure, représenter la fonction d'autocorrélation empirique du bruit blanc $\{Z_i^5\}$ ainsi que l'intervalle de confiance à 95% correspondant. Puis, sur une seconde figure, représenter la fonction d'autocorrélation empirique du processus MA(2) $\{X_i^5\}$, ainsi que les différents intervalles de confiance $\{\pm 1.96n^{-1/2}(1 + 2 \sum_{h=1}^M \rho_e(h)^2), 1 \leq M \leq 5\}$. Comme il a été dit en cours, la démarche précédente permet d'estimer l'ordre M , qui dans le cas d'un processus MA(q), correspond au paramètre q .

V. ALGORITHME DE DURBIN LEVINSON

L'algorithme de Durbin Levinson vise à identifier les paramètres d'un processus AR(p), c'est-à-dire l'ordre p du modèle et les coefficients ϕ_k :

- 1) poser $(\phi_{1,1})_e = \rho_e(1)$ et $(v_1)_e = \gamma_e(0)[1 - \rho_e(1)^2]$;
- 2) calculer, pour $m = 2$, le scalaire $(\phi_{m,m})_e = [\gamma_e(m) - \sum_{k=1}^{m-1} (\phi_{m-1,k})_e \gamma_e(m-k)] / (v_{m-1})_e$;
- 3) calculer, pour $m = 2$, le vecteur $\begin{bmatrix} (\phi_{m,1})_e & \cdots & (\phi_{m,m-1})_e \end{bmatrix} = \begin{bmatrix} (\phi_{m-1,1})_e & \cdots & (\phi_{m-1,m-1})_e \end{bmatrix} - (\phi_{m,m})_e \begin{bmatrix} (\phi_{m-1,m-1})_e & \cdots & (\phi_{m-1,1})_e \end{bmatrix}$ de taille $(1 \times m-1)$;
- 4) $(v_m)_e = (v_{m-1})_e(1 - (\phi_{m,m})_e^2)$;
- 5) réitérer à partir de l'étape 2 en incrémentant la valeur de m ;
- 6) s'arrêter lorsque $m = M$.

Estimation de l'ordre p : les coefficients $\alpha_e(h) = (\phi_{h,h})_e$ étant ceux de la fonction d'autocorrélation partielle empirique du processus considéré, une estimée p_e de p peut être obtenue en prenant la plus petite valeur de k telle que $|(\phi_{h,h})_e| < 1.96n^{-1/2}$ pour $h > k$.

Estimation des coefficients du filtre AR: ces derniers sont donnés par la suite de valeurs $(\phi_k)_e = (\phi_{p_e,k})_e$ pour $1 \leq k \leq p_e$, $(\phi_0)_e$ valant 1 par hypothèse.

A titre d'application, estimer les paramètres du processus AR $\{X_i^2\}$, puis comparer-les aux valeurs exactes données dans l'équation (3).

VI. ALGORITHME DE YULE WALKER

L'algorithme de Yule Walker est une méthode d'identification de processus AR dite par bloc et qui suppose d'avoir préalablement identifié l'ordre p du modèle. Les coefficients ϕ_k et la variance σ^2 du bruit nous sont donnés par les équations suivantes :

$$\Phi_e = [(\mathbf{R}_p)_e]^{-1} (\boldsymbol{\rho}_p)_e \quad (4)$$

et

$$\sigma_e^2 = \gamma_e(0) \left[1 - (\boldsymbol{\rho}_p)_e^\top [(\mathbf{R}_p)_e]^{-1} (\boldsymbol{\rho}_p)_e \right] \quad (5)$$

où $\Phi_e = \left[(\phi_1)_e \quad \cdots \quad (\phi_p)_e \right]^\top$, $(\boldsymbol{\rho}_p)_e = \left[(\rho_1)_e \quad \cdots \quad (\rho_p)_e \right]^\top$ et $(\mathbf{R}_p)_e$ est la matrice dont la (i,j) -ième composante est donnée par $[(\rho_{i-j})_e]_{i,j=1}^p$.

En utilisant l'estimée p_e de l'ordre p du processus $\{X_i^2\}$ obtenu précédemment par l'algorithme de Durbin Levinson, estimer les coefficients ϕ_k de $\{X_i^2\}$ et la variance σ^2 du bruit. Puis comparer-les aux valeurs exactes données dans l'équation (3).